

Active Inference is a Subtype of Variational Inference

Wouter W. L. Nuijten^{1,2} Mykola Lukashchuk^{1*}

¹Department of Electrical Engineering, Eindhoven University of Technology, the Netherlands

²Lazy Dynamics, Utrecht, the Netherlands
w.w.l.nuijten@tue.nl

Abstract

Automated decision-making under uncertainty requires balancing exploitation and exploration. Classical methods treat these separately using heuristics, while Active Inference unifies them through Expected Free Energy (EFE) minimization. However, EFE minimization is computationally expensive, limiting scalability. We build on recent theory recasting EFE minimization as variational inference, formally unifying it with Planning-as-Inference and showing the epistemic drive as a unique entropic contribution. Our main contribution is a novel message-passing scheme for this unified objective, enabling scalable Active Inference in factored-state MDPs and overcoming high-dimensional planning intractability.

1 Introduction

Automated decision-making under uncertainty is a central, long-standing challenge across control theory and artificial intelligence. When the system dynamics are well-known and deterministic, classical methods like Optimal Control [Bellman, 1954] and Model Predictive Control (MPC) [Cutler and Ramaker, 1979] provide principled frameworks for determining optimal actions. These approaches, focused primarily on minimizing a predefined cost function, have been elegantly unified under the Planning as Inference (PAI) paradigm [Toussaint, 2009, Attias, 2003], showing that control can be cast as a variational inference problem on a factor graph [Todorov, 2008, Levine, 2018].

However, the real-world challenge lies in environments where dynamics are stochastic or partially unknown. Popular methods that operate under unknown dynamics are rooted in Reinforcement Learning [Sutton and Barto, 1998], which optimizes long-term utility through value function or policy estimation. These methods attempt to inject epistemic behavior by treating exploration as a distinct form of reward, as seen in Max-entropy Reinforcement Learning [Haarnoja et al., 2018, 2017].

Active Inference proposes an alternative approach to planning under uncertainty [Friston et al., 2015, Parr and Friston, 2019, Da Costa et al., 2020]. This framework provides a neurobiological explanation of intelligent behavior and posits that the optimal policy that balances exploitative and explorative behavior emerges when minimizing a quantity known as the Expected Free Energy (EFE) [Friston et al., 2010, Da Costa et al., 2020]. However, the EFE is an objective that is defined over sequences of actions and does therefore not define a variational objective over beliefs that we can optimize.

Recently, an attempt has been made to redefine EFE-based planning as a standard Variational Free Energy [De Vries et al., 2025] by adjusting the generative model by introducing epistemic priors.

In this paper, we will take a closer look at the objective defined by De Vries et al. [2025] and frame it as a form of entropic inference, as defined by Lázaro-Gredilla et al. [2024]. Afterwards, we will

*Equal contribution between authors.

derive a message passing scheme that corresponds to the found formulation of Active Inference and which can be locally minimized on a Factor Graph.

The main contributions of this paper are twofold:

- We formally reframe Active Inference’s EFE minimization as a form of entropy-corrected variational inference, explicitly demonstrating that the epistemic drive corresponds to a unique entropic contribution within the variational objective.
- We derive a message-passing scheme for this unified objective. Crucially, this scheme introduces region-extended Bethe coordinates and an r -channel reparameterization coordinate, which together turn a degenerate conditional entropy into a local cross-entropy and render the overall objective computationally feasible for local optimization on a factor graph.

The rest of the paper is structured as follows: in [Section 2](#) we recover Active Inference as a form of variational inference similar to how planning is recovered in [Lázaro-Gredilla et al. \[2024\]](#). This illustrates that the epistemic drive introduced by Active Inference can be materialized as an entropic contribution to the variational objective. In [Section 3](#) we will derive a message passing scheme to minimize the Active Inference objective, providing a method to implement scalable Active Inference.

For a definition of the terminology used in the rest of the paper, we refer the reader to [Appendix A](#).

2 Active Inference as Entropy Corrected Inference

In this section, we will rewrite the Variational Free Energy of an adjusted generative model as a form of entropy corrected inference, comparing it to other formulations of planning-as-inference and posing Active Inference as a separate method on the variational inference landscape.

We will consider the following standard biased generative model

$$p(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) \propto p(\theta)p(x_0) \prod_{t=1}^T p(y_t|x_t, \theta)p(x_t|x_{t-1}, u_t, \theta)p(u_t)\hat{p}_x(x_t)\hat{p}_y(y_t). \quad (1)$$

Here, \mathbf{y} are observations, \mathbf{x} are latent states, θ are hidden parameters, and \mathbf{u} is a sequence of control signals or actions. $\hat{p}_x(x_t)$ and $\hat{p}_y(y_t)$ represent goal priors on future states and observations, which can be proportional to a prespecified reward but do not necessarily need to be.

The variational objective defined by [De Vries et al. \[2025\]](#) manipulates the Variational Free Energy (VFE) of the model in (1) through the inclusion of epistemic priors. In this section, we demonstrate that the framework of [De Vries et al. \[2025\]](#) extends beyond Active Inference by reformulating their objective within the broader landscape of entropic inference introduced by [Lázaro-Gredilla et al. \[2024\]](#).

In this entropic inference framework, all inference types minimize a common VFE (A.1) while differing only in their entropy corrections. Following this principle, [Theorem 1](#) shows that the VFE of the epistemic-prior-augmented generative model from (A.3) can be equivalently expressed as the VFE of the original generative model plus specific entropy correction terms, thereby positioning Active Inference within the unified variational inference landscape of [Table 1](#).

Theorem 1. *The variational objective presented in [De Vries et al. \[2025\]](#) (presented in [subsection A.2](#)) can be rearranged in the following way:*

$$F_{\tilde{p}}[q] = F_p[q] + \sum_{t=1}^T \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_t, x_{t-1}, u_t)] + \mathbb{H}[q(y_t, x_t, \theta)] - \mathbb{H}[q(x_t, \theta)] \quad (2)$$

where $F_p[q]$ is the Variational Free Energy associated with the generative model.

Proof. Given in [Appendix B](#). □

We are now in the position to compare Active Inference with other forms of entropic inference. Interestingly, by [Lemma 4](#), an adjusted generative model with only $\hat{p}(u_t)$ as entropic prior recovers a form of inference surprisingly similar to [Lázaro-Gredilla et al. \[2024\]](#). However, where the

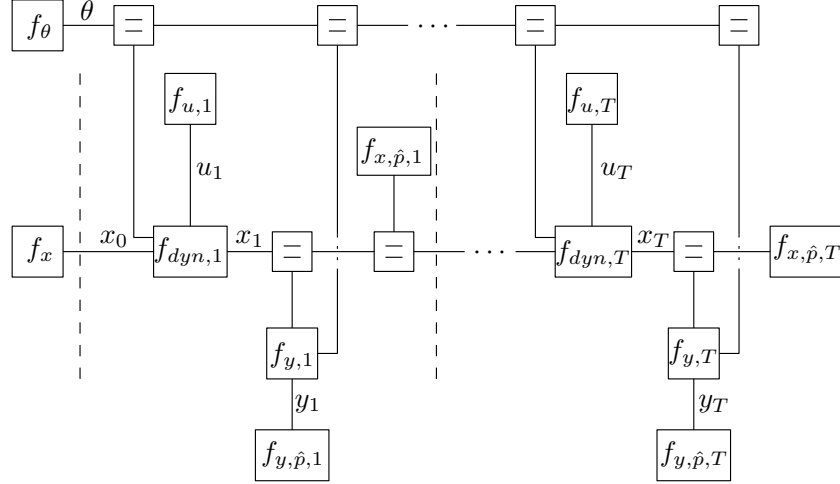


Figure 1: Factor graph representation of the generative model (1). Nodes (boxes) represent factors from the generative model: f_θ is the prior on parameters, f_{x_0} is the initial state prior, $f_{dyn,t}$ represents the dynamics $p(x_t|x_{t-1}, \theta, u_t)$, $f_{y,t}$ represents observations $p(y_t|x_t, \theta)$, and $f_{u,t}$, $f_{x,\hat{p},t}$, $f_{y,\hat{p},t}$ represent action priors and goal priors respectively. Edges (lines) represent random variables: θ (parameters), x_t (states), y_t (observations), and u_t (actions). In the Bethe approximation, each *node* a maintains a local belief $q_a(s_a)$ over its scope (the variables connected to it), while each *edge* i maintains a singleton belief $q_i(s_i)$. These local beliefs must satisfy consistency constraints (5). This factorization enables *local* optimization scheme (message passing): rather than optimizing a single global distribution $q(y, x, \theta, u)$, we optimize a collection of local beliefs that communicate through messages.

planning-as-inference objective defines a degenerate optimization procedure, this objective admits an optimization scheme. This point will be elaborated on in [Section 3](#). We will refer to this type of inference as Maximum Ambiguity (MaxAmb) planning, and with the inclusion of $\tilde{p}(x_t)$ and $\tilde{p}(y_t, x_t)$, recovers Active Inference. An overview of the different types of entropic inference is given in [Table 1](#).

Table 1: Positioning Active Inference within the variational inference landscape. Following [Lázaro-Gredilla et al. \[2024\]](#), various inference methods can be expressed as energy minimization with different entropy corrections. Active Inference emerges as a natural extension that incorporates both planning and epistemic (ambiguity-reducing) terms. Note a slight difference from the exposition presented in [\[Lázaro-Gredilla et al., 2024, Table 1\]](#): there, the entropy correction is presented for the so-called energy term, but these two frameworks are trivially equivalent in the cases presented below. However, we find this table clearer when written as an entropic correction for VFE, because it becomes much easier to determine the degenerate schemes (this point will be elaborated in detail in [section 3](#)).

Type of inference	Entropy correction (relative to VFE)
Marginal	0
MAP	$\mathbb{H}[q]$
Planning	$\sum_{t=1}^T \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_{t-1})]$ (Appendix C)
MaxAmb planning	$\sum_{t=1}^T \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_t, x_{t-1}, u_t)]$
Active Inference	$\sum_{t=1}^T \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_t, x_{t-1}, u_t)] + \mathbb{H}[q(y_t, x_t, \theta)] - \mathbb{H}[q(x_t, \theta)]$

Interestingly enough, the contributions from [Lemma 3](#) and [Lemma 5](#) contain the same terms with their signs flipped, where all contributions from $\tilde{p}(x_t)$ are canceled out. This warrants a revision of the epistemic priors $\tilde{p}(x_t)$ and $\tilde{p}(y_t, x_t)$. In [Appendix D](#) we provide a proof that these two priors can be replaced by $\tilde{p}(x_t, \theta) = \exp(-\mathbb{H}[q(y_t | x_t, \theta)])$ without changing the inference objective. This re-arrangement is theoretically useful because it shows that parameters and states are actually not distinguished by entropic priors, and the only possible distinction could come from the generative model itself. With the landscape of entropic inference set up, we are in a position to derive a message-passing procedure corresponding to Active Inference.

3 Deriving Message Passing

To obtain a *local* objective amenable to message passing, we replace the global VFE in [Theorem 1](#) by its Bethe approximation presented in detail in [subsection A.3](#). On tree-structured instances of (1) this replacement is *exact* (for instance, a θ -free model); otherwise, it is a standard variational approximation. But to define the Bethe objective, we need to identify our model with a factor graph

(shown in Figure 1). We start with the node set \mathcal{V}

$$f_\theta(\theta) = p(\theta), \quad f_{x_0}(x_0) = p(x_0), \quad (3a)$$

$$f_{y,t}(y_t, x_t, \theta) = p(y_t | x_t, \theta), \quad f_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) = p(x_t | x_{t-1}, \theta, u_t), \quad (3b)$$

$$f_{u,t}(u_t) = p(u_t), \quad f_{x,\hat{p},t}(x_t) = \hat{p}_x(x_t), \quad f_{y,\hat{p},t}(y_t) = \hat{p}_y(y_t). \quad (3c)$$

With the natural set of edges \mathcal{E}

$$\mathcal{E} := \{\theta, x_0\} \cup \bigcup_{t=1}^T \{x_t, y_t, u_t\}. \quad (4)$$

Each node $a \in \mathcal{V}$ has scope $s_a \subseteq \mathcal{E}$.

In Bethe terminology, each node $a \in \mathcal{V}$ has an associated *local* belief $q_a(s_a)$ over its scope. Additionally, to impose consistency constraints between nodes, we introduce singleton beliefs $q_i(s_i)$ for each edge $i \in \mathcal{E}$. Together, these node beliefs $\{q_a(s_a)\}_{a \in \mathcal{V}}$ and edge beliefs $\{q_i(s_i)\}_{i \in \mathcal{E}}$ form what we call the **Bethe coordinates**: a collection of normalized probability distributions that must satisfy the local consistency constraints:

$$\int q_a(s_a) ds_{a \setminus i} = q_i(s_i) \quad (5)$$

whenever $i \in s_a$.

With this notation, the Bethe Free Energy specialized to (1) takes the standard form

$$F_f[q] = \sum_{a \in \mathcal{V}} \mathbb{D}[q_a(s_a) || f_a(s_a)] + \sum_{i \in \mathcal{E}} (d_i - 1) \mathbb{H}[q_i(s_i)],$$

whose fully expanded expression and d_i are given in Equation E.1 and Equation E.2 respectively.

But to make the objective from Theorem 1 local, we must express all its terms using local marginals; otherwise, it is a global objective. Intuitively, this means we need to find a node in our factor graph to which we can attach each new term. For instance, $q(y_t, x_t, \theta)$ can be attached to the node $f_{y,t}$ and be identified with $q_{y,t}$ Bethe coordinate.

However, the Bethe coordinates alone are *insufficient* for the adjusted objective in Theorem 1, because the entropic correction contains

$$- \mathbb{H}[q(x_t, \theta)] + \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_t, x_{t-1}, u_t)].$$

These three entropy terms involve marginal distributions that are not Bethe coordinates: $q(x_t, \theta)$, $q(x_{t-1}, u_t)$, and $q(x_t, x_{t-1}, u_t)$. None of these distributions correspond to the scope of any node in our factor graph, nor are they singleton beliefs over edges. Note, the same exact reasoning applies to Planning and MaxAbm Planning from Table 1.

To keep the objective local, we therefore introduce three auxiliary *region beliefs*:

$$q_{\text{sep},t}(x_t, \theta) := \int q_{y,t}(y_t, x_t, \theta) dy_t = \int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) dx_{t-1} du_t, \quad (6a)$$

$$q_{\text{trip},t}(x_t, x_{t-1}, u_t) := \int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) d\theta, \quad (6b)$$

$$q_{\text{pair},t}(x_{t-1}, u_t) := \int q_{\text{trip},t}(x_t, x_{t-1}, u_t) dx_t \quad (6c)$$

with the natural projections (e.g., $\int q_{\text{trip},t} dx_t = q_{\text{pair},t}$, $\int q_{\text{sep},t} dx_t = q_\theta$) enforcing consistency with existing beliefs. We will refer to this coordinate system as the *region-extended Bethe coordinates*. In the region-extended Bethe coordinates, the objective from Theorem 1 can be expressed as follows

$$F_f[q] + \sum_{t=1}^T \left(\mathbb{H}[q_{y,t}] - \mathbb{H}[q_{\text{sep},t}] + \mathbb{H}[q_{\text{pair},t}] - \mathbb{H}[q_{\text{trip},t}] \right).$$

However, after adding the new region marginals, we obtain the local objective that is degenerate because of the term $\mathbb{H}[q_{y,t}]$; we prove this supporting result Theorem 7 in subsection E.1. To have

both locality and full-identifiability, we augment the coordinates with a single *channel* variable $r_{y|x\theta,t}(y_t | x_t, \theta)$ with the natural normalization constraint

$$\int r_{y|x\theta,t}(y_t | x_t, \theta) dy_t = 1 \quad (7)$$

and rewrite the global conditional entropy as a local cross-entropy,

$$H[q(y_t | x_t, \theta)] = \min_{r_{y|x\theta,t}} \mathbb{E}_{q_{y,t}(y_t, x_t, \theta)} [-\log r_{y|x\theta,t}(y_t | x_t, \theta)].$$

Under this reparameterization, $q_{\text{sep},t}(x_t, \theta)$ is no longer needed as a free coordinate in the objective: the global term $H[q(y_t, x_t, \theta)] - H[q(x_t, \theta)]$ collapses into the conditional form and depends only on $(q_{y,t}, r_{y|x\theta,t})$ locally. The following theorem shows the stationary conditions in the r -adjusted coordinate system. The proof of [Theorem 2](#) is provided in [Appendix E](#).

Theorem 2 (The stationary scheme for Active Inference). *Consider the Bethe objective (E.1) for the model (1) augmented by the Active Inference correction of [Theorem 1](#), and adopt the adjusted coordinate system*

$$\left\{ q_{y,t}(y_t, x_t, \theta), q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t), q_{\text{sep},t}(x_t, \theta), \right. \\ \left. q_{\text{trip},t}(x_t, x_{t-1}, u_t), q_{\text{pair},t}(x_{t-1}, u_t), r_{y|x\theta,t}(y_t | x_t, \theta) \right\}_{t=1}^T$$

with the projection constraints [Equation 6](#) and the row-normalization [Equation 7](#). Then any stationary point satisfies, for each $t = 1, \dots, T$, the following local equations (all equalities are up to normalizers):

$$q_{y,t}(y_t, x_t, \theta) \propto p(y_t | x_t, \theta) r_{y|x\theta,t}(y_t | x_t, \theta) \hat{p}_y(y_t) \exp\{-\Lambda_{x\theta}(x_t, \theta)\}, \quad (8)$$

$$r_{y|x\theta,t}(y_t | x_t, \theta) \propto \frac{q_{y,t}(y_t, x_t, \theta)}{q_{\text{sep},t}(x_t, \theta)} \quad (9)$$

$$\exp\{-\Lambda_{x\theta}(x_t, \theta)\} \propto \frac{\int p(y_t | x_t, \theta) r_{y|x\theta,t}(y_t | x_t, \theta) \hat{p}_y(y_t) dy_t}{q_{\text{sep},t}(x_t, \theta)}. \quad (10)$$

$$q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) \propto p(x_t | x_{t-1}, \theta, u_t) \exp\{-\Lambda_{x\theta}(x_t, \theta)\} \exp\{-\Lambda_{\text{trip}}(x_t, x_{t-1}, u_t)\}, \quad (11)$$

$$\exp\{-\Lambda_{\text{trip}}(x_t, x_{t-1}, u_t)\} \propto \frac{q_{\text{pair},t}(x_{t-1}, u_t)}{q_{\text{trip},t}(x_t, x_{t-1}, u_t)}. \quad (12)$$

The region beliefs are tied by the projections $\int q_{\text{dyn},t} d\theta = q_{\text{trip},t}$ and $\int q_{\text{trip},t} dx_t = q_{\text{pair},t}$.

All remaining coordinates. Singletons $q_{x_t}, q_{y_t}, q_{u_t}, q_\theta, q_{x_0}$ and unary factor beliefs (including \hat{p}_x, \hat{p}_y) satisfy the classical Bethe equations, i.e. the standard belief-propagation fixed-point conditions on the generative model; equivalently, their multipliers/messages are exactly those of BP on (1) (with degrees (E.2)) and are not modified by the entropic correction.

4 Discussion

While our theoretical framework provides principled planning with epistemic objectives, its computational implementation faces significant challenges that warrant careful analysis.

To implement our scheme, we must address the nontrivial factor graph structure shown in [Figure 2](#). This is a Forney-style factor graph [[Forney, 2001](#)] representing a single time slice, where variables are represented on edges and factors are represented as nodes. Message passing algorithms are generally implemented on factor graphs to leverage their locality [[Bagaev and De Vries, 2023](#)]. However, the factor graph corresponding to the scheme introduced in [Section 3](#) is nontrivial. Unlike standard factor graphs derived directly from generative models, our approach requires region-based representations [[Yedidia et al., 2005](#)] with edges representing multiple variables. In [Figure 2](#), we

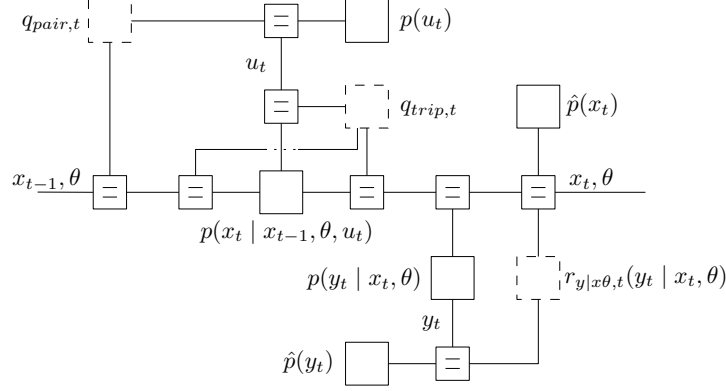


Figure 2: Time-slice factor graph corresponding to the scheme introduced in section 3. To form a full generative model to run inference, we chain T of these slices to form a terminated factor graph.

see the edges representing (x_{t-1}, θ) and (x_t, θ) . Furthermore, there are additional nodes (shown in dotted lines in Figure 2) that are necessary to compute the new coordinates in our optimization procedure. The functional form of these nodes is currently unknown, but the messages that are sent are defined by the schedule derived in Theorem 2. This means we cannot interpret what these nodes do concretely, highlighting a gap between our theoretical framework and its practical implementation.

Previous attempts have been made to implement a minimization of the objective presented in De Vries et al. [2025]. These attempts have also implemented a message-passing procedure on a factor graph [Nuijten et al., 2025]. The scheme previously derived manually recomputes the epistemic priors for each iteration of the variational inference procedure. Still, the scheme can be viewed as a linearization of the true message-passing scheme, which is derived in this work.

Beyond structural challenges, the computational complexity of our approach is quadratic in the state space size. Previous derivations of Planning-as-Inference express the computational cost of the procedure in terms of the number of variables [Lázaro-Gredilla et al., 2024]. We argue that this is a misleading way to express cost, as the cost of computing joint marginal distributions greatly depends on the size and dimensionality of the state space. The scheme introduced in Section 3 warrants the computation of $q_{y,t}(y_t, x_t, \theta)$, $q_{pair,t}(x_{t-1}, u_t)$ and $q_{trip,t}(x_t, x_{t-1}, u_t)$. In discrete state and action spaces, the computational complexity of computing these quantities is polynomial in the state and action spaces. The computation of $q_{trip,t}(x_t, x_{t-1}, u_t)$ is the most expensive, since it takes $\mathcal{O}(|X|^2 \cdot |U| \cdot D_\Theta)$ time, where D_Θ is the dimensionality of the parameter space. Note that, if we would not localize the inference procedure and not introduce our message passing scheme, the computational complexity would be exponential in the planning horizon T . Interestingly enough, the scheme for Planning as Inference also requires this time complexity, as the same $q_{trip,t}(x_t, x_{t-1}, u_t)$ is computed. The epistemic drive, however, elicits an additional complexity cost. Computing $q_{y,t}(y_t, x_t, \theta)$ takes $\mathcal{O}(|Y| \cdot |X| \cdot D_\Theta)$ time. This quadratic dependence on the state space is limiting for interesting problems where the state space quickly grows with the size of the system, such as Minigrid environments [Chevalier-Boisvert et al., 2023].

These complexity limitations suggest that for this approach to truly scale, hierarchical state-space partitioning becomes essential. The scheme derived in this paper warrants a partition of the state-space, hinting to a hierarchical generative model [Palacios et al., 2020, Beck and Ramstead, 2025]. Such hierarchical partitioning would allow us to avoid the quadratic complexity within state-space partitions, dramatically reducing computational costs while maintaining the principled epistemic objectives of our framework.

Acknowledgements

This publication is part of the project ROBUST: Trustworthy AI-based Systems for Sustainable Growth with project number KICH3.LTP.20.006, which is (partly) financed by the Dutch Research Council (NWO), GN Hearing, and the Dutch Ministry of Economic Affairs and Climate Policy (EZK) under the program LTP KIC 2020-2023.

References

- Hagai Attias. Planning by probabilistic inference. In *International workshop on artificial intelligence and statistics*, pages 9–16. PMLR, 2003. URL <https://proceedings.mlr.press/r4/attias03a.html>.
- Dmitry Bagaev and Bert De Vries. Reactive Message Passing for Scalable Bayesian Inference. *Scientific Programming*, 2023:1–26, May 2023. ISSN 1875-919X, 1058-9244. doi: 10.1155/2023/6601690. URL <https://www.hindawi.com/journals/sp/2023/6601690/>.
- Jeff Beck and Maxwell J. D. Ramstead. Dynamic Markov Blanket Detection for Macroscopic Physics Discovery, February 2025. URL <http://arxiv.org/abs/2502.21217>. arXiv:2502.21217 [q-bio].
- Richard Bellman. The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6):503–515, 1954. ISSN 0002-9904, 1936-881X. doi: 10.1090/S0002-9904-1954-09848-8. URL <https://www.ams.org/bull/1954-60-06/S0002-9904-1954-09848-8/>.
- Maxime Chevalier-Boisvert, Bolun Dai, Mark Towers, Rodrigo Perez-Vicente, Lucas Willems, Salem Lahlou, Suman Pal, Pablo Samuel Castro, and Jordan Terry. Mini-igrid & miniworld: Modular & customizable reinforcement learning environments for goal-oriented tasks. *Advances in Neural Information Processing Systems*, 36:73383–73394, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/e8916198466e8ef218a2185a491b49fa-Abstract-Datasets_and_Benchmarks.html.
- Richard R. Cutler and B. L. Ramaker. Dynamic Matrix Control-A Computer Control Algorithm. *Proc. Joint Automatic Control Conference*, 1979, 1979. URL <https://cir.nii.ac.jp/crid/1570291225777284224>.
- Lancelot Da Costa, Thomas Parr, Noor Sajid, Sebastijan Veselic, Victorita Neacsu, and Karl Friston. Active inference on discrete state-spaces: A synthesis. *Journal of Mathematical Psychology*, 99: 102447, December 2020. ISSN 0022-2496. doi: 10.1016/j.jmp.2020.102447. URL <https://www.sciencedirect.com/science/article/pii/S0022249620300857>.
- Bert De Vries, Wouter Nuijten, Thijs van de Laar, Wouter Kouw, Sepideh Adamiat, Tim Nisslbeck, Mykola Lukashchuk, Hoang Minh Huu Nguyen, Marco Hidalgo Araya, Raphael Tresor, Thijs Jenneskens, Ivana Nikoloska, Raaja Ganapathy Subramanian, Bart van Erp, Dmitry Bagaev, and Albert Podusenko. Expected Free Energy-based Planning as Variational Inference, April 2025. URL <http://arxiv.org/abs/2504.14898>. arXiv:2504.14898 [stat].
- G. David Forney. Codes on graphs: Normal realizations. *IEEE Transactions on Information Theory*, 47(2):520–548, 2001. URL <https://ieeexplore.ieee.org/abstract/document/910573/>. Publisher: IEEE.
- Karl Friston, Francesco Rigoli, Dimitri Ognibene, Christoph Mathys, Thomas Fitzgerald, and Giovanni Pezzulo. Active inference and epistemic value. *Cognitive Neuroscience*, 6(4):187–214, October 2015. ISSN 1758-8928, 1758-8936. doi: 10.1080/17588928.2015.1020053. URL <http://www.tandfonline.com/doi/full/10.1080/17588928.2015.1020053>.
- Karl J. Friston, Jean Daunizeau, James Kilner, and Stefan J. Kiebel. Action and behavior: a free-energy formulation. *Biological Cybernetics*, 102(3):227–260, March 2010. ISSN 1432-0770. doi: 10.1007/s00422-010-0364-z. URL <https://doi.org/10.1007/s00422-010-0364-z>.
- Tuomas Haarnoja, Haoran Tang, Pieter Abbeel, and Sergey Levine. Reinforcement Learning with Deep Energy-Based Policies. In *Proceedings of the 34th International Conference on Machine Learning*, pages 1352–1361. PMLR, July 2017. URL <https://proceedings.mlr.press/v70/haarnoja17a.html>. ISSN: 2640-3498.
- Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In *Proceedings of the 35th International Conference on Machine Learning*, pages 1861–1870. PMLR, July 2018. URL <https://proceedings.mlr.press/v80/haarnoja18b.html>. ISSN: 2640-3498.

Sergey Levine. Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review, May 2018. URL <http://arxiv.org/abs/1805.00909>. arXiv:1805.00909 [cs].

Miguel Lázaro-Gredilla, Li Yang Ku, Kevin P. Murphy, and Dileep George. What type of inference is planning?, November 2024. URL <http://arxiv.org/abs/2406.17863>. arXiv:2406.17863.

Wouter W. L. Nuijten, Mykola Lukashchuk, Thijs van de Laar, and Bert de Vries. A Message Passing Realization of Expected Free Energy Minimization, August 2025. URL <http://arxiv.org/abs/2508.02197>. arXiv:2508.02197 [cs].

Enzor Rafael Palacios, Adeel Razi, Thomas Parr, Michael Kirchhoff, and Karl Friston. On Markov blankets and hierarchical self-organisation. *Journal of Theoretical Biology*, 486:110089, February 2020. ISSN 0022-5193. doi: 10.1016/j.jtbi.2019.110089. URL <https://www.sciencedirect.com/science/article/pii/S0022519319304588>.

Thomas Parr and Karl J. Friston. Generalised free energy and active inference. *Biological Cybernetics*, 113(5):495–513, December 2019. ISSN 1432-0770. doi: 10.1007/s00422-019-00805-w. URL <https://doi.org/10.1007/s00422-019-00805-w>.

Richard S. Sutton and Andrew G. Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998. URL <https://www.cambridge.org/core/journals/robotica/article/robot-learning-edited-by-jonathan-h-connell-and-sridhar-mahadevan-kluwer-boston-19931997-xii2737FD21CA908246DF17779E9C20B6DF6>.

Emanuel Todorov. General duality between optimal control and estimation. In *2008 47th IEEE Conference on Decision and Control*, pages 4286–4292, December 2008. doi: 10.1109/CDC.2008.4739438. URL <https://ieeexplore.ieee.org/abstract/document/4739438>. ISSN: 0191-2216.

Marc Toussaint. Robot trajectory optimization using approximate inference. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1049–1056, Montreal Quebec Canada, June 2009. ACM. ISBN 978-1-60558-516-1. doi: 10.1145/1553374.1553508. URL <https://dl.acm.org/doi/10.1145/1553374.1553508>.

Jonathan S. Yedidia, W.T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7): 2282–2312, July 2005. ISSN 0018-9448. doi: 10.1109/TIT.2005.850085.

İsmail Şenöz. *Message Passing Algorithms for Hierarchical Dynamical Models*. Phd Thesis 1 (Research TU/e / Graduation TU/e), Eindhoven University of Technology, Eindhoven, June 2022. ISBN: 9789038655321.

A Background

A.1 Variational Inference and the Posterior Factorization

In standard Variational Inference, we minimize the Variational Free Energy (VFE) between a variational posterior q and a generative model p :

$$F_p[q] = \int q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) \log \frac{q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u})}{p(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u})} d\mathbf{y} d\mathbf{x} d\theta d\mathbf{u}. \quad (\text{A.1})$$

We are considering the factorized generative model defined in (1). Since (A.1) defines a functional objective that we can minimize, we should also specify a family \mathcal{Q} over which we are optimizing the VFE. We will choose the elements q of this family such that they decompose as follows

$$q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) = q(x_0, \theta) \prod_{t=1}^T q(y_t | x_t, \theta) q(x_t | x_{t-1}, u_t, \theta) q(u_t | x_{t-1}, \theta). \quad (\text{A.2})$$

This factorization of the posterior distribution is required for the definition of the augmented generative model in subsection A.2 [Nuijten et al., 2025].

A.2 The Epistemic-Prior-Augmented Generative Model

The key insight of [De Vries et al. \[2025\]](#) is that Active Inference can be recovered through an *adjusted* or *augmented* generative model that includes additional factors called *epistemic priors*. From the generative model, an augmented model is constructed

$$\tilde{p}(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) = p(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) \cdot \prod_{t=1}^T \tilde{p}(x_t) \tilde{p}(u_t) \tilde{p}(y_t, x_t). \quad (\text{A.3})$$

Here, the additional \tilde{p} terms are epistemic priors, which are functions of the variational distribution q itself, creating a self-consistent optimization problem. The epistemic priors take the following forms:

$$\tilde{p}(u_t) = \exp(\mathbb{H}[q(x_t, x_{t-1}|u_t)] - \mathbb{H}[q(x_{t-1}|u_t)]) \quad (\text{action prior}); \quad (\text{A.4a})$$

$$\tilde{p}(x_t) = \exp(-\mathbb{H}[q(y_t|x_t)]) \quad (\text{state prior}); \quad (\text{A.4b})$$

$$\tilde{p}(y_t, x_t) = \exp(D_{KL}[q(\theta|y_t, x_t) \| q(\theta|x_t)]) \quad (\text{observation prior}), \quad (\text{A.4c})$$

where the entropy of a distribution q over variables z_1, \dots, z_n defined as follows

$$\mathbb{H}[q(z_1, \dots, z_n)] = - \int q(z_1, \dots, z_n) \log q(z_1, \dots, z_n) dz_1 \dots dz_n, \quad (\text{A.5})$$

and the conditional entropy has the following functional form

$$\mathbb{H}[q(z_1, \dots, z_n | \omega)] = - \int q(z_1, \dots, z_n | \omega) \log q(z_1, \dots, z_n | \omega) dz_1 \dots dz_n. \quad (\text{A.6})$$

The epistemic priors admit a practical interpretation: the action prior encourages actions that resolve ambiguity, the state prior favors informative states, and the observation prior encourages observations that are informative about the parameters.

The *adjusted VFE* is then defined in the following way:

$$F_{\tilde{p}}[q] = \int q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) \log \frac{q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u})}{\tilde{p}(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u})} d\mathbf{y} d\mathbf{x} d\theta d\mathbf{u}. \quad (\text{A.7})$$

Crucially, [De Vries et al. \[2025\]](#) show that minimizing this adjusted VFE is equivalent to minimizing a bound on the Expected Free Energy (EFE) objective from Active Inference [[Friston et al., 2015](#)].

A.3 Bethe Free Energy

Let $(\mathcal{V}, \mathcal{E})$ be the (Forney-style) factor graph of a positive function $f(s) = \prod_{a \in \mathcal{V}} f_a(s_a)$, where s_a collects the variables incident on factor a , and each edge $i \in \mathcal{E}$ carries a variable s_i . The *Bethe variational family* consists of *factor beliefs* $\{q_a(s_a)\}_{a \in \mathcal{V}}$ and *edge beliefs* $\{q_i(s_i)\}_{i \in \mathcal{E}}$ constrained to the *marginal manifold*:

$$\int q_a(s_a) ds_a = 1 \quad \forall a \in \mathcal{V}, \quad (\text{A.8a})$$

$$\int q_i(s_i) ds_i = 1 \quad \forall i \in \mathcal{E}, \quad (\text{A.8b})$$

$$\int q_a(s_a) ds_{a \setminus i} = q_i(s_i) \quad \forall a \in \mathcal{V}, \forall i \in a, \quad (\text{A.8c})$$

where $s_{a \setminus i}$ denotes the variables in s_a except s_i . Let d_i be the number of factors incident on edge i (the degree of variable s_i). The *Bethe Free Energy* (BFE) [[Yedidia et al., 2005](#)] is

$$F_f[q] = \sum_{a \in \mathcal{V}} \mathbb{D}[q_a(s_a) \| f_a(s_a)] + \sum_{i \in \mathcal{E}} (d_i - 1) \mathbb{H}[q_i(s_i)], \quad (\text{A.9})$$

where

$$\mathbb{D}[q_a(s_a) \| f_a(s_a)] = \int q_a(s_a) \log \frac{q_a(s_a)}{f_a(s_a)} ds_a. \quad (\text{A.10})$$

The *Bethe approximation* takes $\log Z_f \approx -\min_{q \in \mathcal{M}} F_f[q]$, where \mathcal{M} is the manifold (A.8). The approximation is *exact* when the factor graph is a tree; on loopy graphs, stationary points of (A.9) coincide with fixed points of (loopy) belief propagation [[Yedidia et al., 2005](#)]. We refer the reader to [Şenöz \[2022\]](#) for a modern variational calculus exposition on the message passing algorithms derivation from (A.9).

B Proof of Theorem 1

To prove the theorem, we will require three lemmas:

Lemma 3. *Under the assumption that our posterior distribution factorizes as in Equation A.2, and $\tilde{p}(x_t) = \exp(-\mathbb{H}[q(y_t | x_t)])$:*

$$-\int q(x_t) \log \tilde{p}(x_t) dx_t = \mathbb{H}[q(y_t, x_t)] - \mathbb{H}[q(x_t)] \quad (\text{B.1})$$

Proof.

$$-\int q(x_t) \log \tilde{p}(x_t) dx_t \quad (\text{B.2a})$$

$$= -\int q(x_t) \int q(y_t | x_t) \log q(y_t | x_t) dy_t dx_t \quad (\text{B.2b})$$

$$= -\int q(y_t | x_t) q(x_t) \log q(y_t | x_t) dy_t dx_t \quad (\text{B.2c})$$

$$= -\int q(y_t, x_t) \log \frac{q(y_t, x_t)}{q(x_t)} dy_t dx_t \quad (\text{B.2d})$$

$$= -\int q(y_t, x_t) \log \frac{q(y_t, x_t)}{q(x_t)} dy_t dx_t = \mathbb{H}[q(y_t, x_t)] - \mathbb{H}[q(x_t)]. \quad (\text{B.2e})$$

□

Lemma 4. *Under the assumption that our posterior distribution factorizes as in Equation A.2, and $\tilde{p}(u_t) = \exp(\mathbb{H}[q(x_t, x_{t-1} | u_t)] - \mathbb{H}[q(x_{t-1} | u_t)])$ the following identity holds:*

$$-\int q(u_t) \log \tilde{p}(u_t) du_t = \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_t, x_{t-1}, u_t)] \quad (\text{B.3})$$

Proof.

$$-\int q(u_t) \log \tilde{p}(u_t) du_t \quad (\text{B.4a})$$

$$= \int q(u_t) \left(\int q(x_t, x_{t-1} | u_t) \log q(x_t, x_{t-1} | u_t) dx_t dx_{t-1} - \int q(x_{t-1} | u_t) \log q(x_{t-1} | u_t) dx_{t-1} \right) du_t \quad (\text{B.4b})$$

$$= \int q(u_t) \left(\int \frac{q(x_t, x_{t-1}, u_t)}{q(u_t)} \log \frac{q(x_t, x_{t-1}, u_t)}{q(u_t)} dx_t dx_{t-1} - \int \frac{q(x_{t-1}, u_t)}{q(u_t)} \log \frac{q(x_{t-1}, u_t)}{q(u_t)} dx_{t-1} \right) du_t \quad (\text{B.4c})$$

$$= \int q(x_t, x_{t-1}, u_t) \log \frac{q(x_t, x_{t-1}, u_t)}{q(u_t)} dx_t dx_{t-1} du_t - \int q(x_{t-1}, u_t) \log \frac{q(x_{t-1}, u_t)}{q(u_t)} dx_{t-1} du_t \quad (\text{B.4d})$$

$$= \underbrace{\int q(x_t, x_{t-1}, u_t) \log q(x_t, x_{t-1}, u_t) dx_t dx_{t-1} du_t}_{-\mathbb{H}[q(x_t, x_{t-1}, u_t)]} + \underbrace{\int q(x_t, x_{t-1}, u_t) \log \frac{1}{q(u_t)} dx_t dx_{t-1} du_t}_{\mathbb{H}[q(u_t)]}$$

$$- \underbrace{\int q(x_{t-1}, u_t) \log q(x_{t-1}, u_t) dx_{t-1} du_t}_{-\mathbb{H}[q(x_{t-1}, u_t)]} - \underbrace{\int q(x_{t-1}, u_t) \log \frac{1}{q(u_t)} dx_{t-1} du_t}_{\mathbb{H}[q(u_t)]} \quad (\text{B.4e})$$

$$= \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_t, x_{t-1}, u_t)] \quad (\text{B.4f})$$

□

Lemma 5. Under the assumption that our posterior distribution factorizes as in Equation A.2, and $\tilde{p}(y_t, x_t) = \exp D_{KL}[q(\theta | y_t, x_t) || q(\theta | x_t)]$:

$$- \int q(y_t, x_t) \log \tilde{p}(y_t, x_t) dy_t dx_t = \mathbb{H}[q(y_t, x_t, \theta)] + \mathbb{H}[q(x_t)] - \mathbb{H}[q(y_t, x_t)] - \mathbb{H}[q(x_t, \theta)] \quad (\text{B.5})$$

Proof.

$$- \int q(y_t, x_t) \log \tilde{p}(y_t, x_t) dy_t dx_t \quad (\text{B.6a})$$

$$= - \int q(y_t, x_t) \left(\int q(\theta | y_t, x_t) \log \frac{q(\theta | y_t, x_t)}{q(\theta | x_t)} d\theta \right) dy_t dx_t \quad (\text{B.6b})$$

$$= - \int q(y_t, x_t) \left(\int q(\theta | y_t, x_t) \log q(y_t, x_t, \theta) - \log q(y_t, x_t) d\theta \right) dy_t dx_t \\ + \int q(y_t, x_t) \left(\int q(\theta | y_t, x_t) \log q(\theta, x_t) - \log q(x_t) d\theta \right) dy_t dx_t \quad (\text{B.6c})$$

$$= - \underbrace{\int q(y_t, x_t, \theta) \log q(y_t, x_t, \theta) dy_t dx_t d\theta}_{\mathbb{H}[q(y_t, x_t, \theta)]} + \underbrace{\int q(y_t, x_t, \theta) \log q(y_t, x_t) dy_t dx_t d\theta}_{-\mathbb{H}[q(y_t, x_t)]} \\ + \underbrace{\int q(y_t, x_t, \theta) \log q(x_t, \theta) dy_t dx_t d\theta}_{-\mathbb{H}[q(x_t, \theta)]} - \underbrace{\int q(y_t, x_t, \theta) \log q(x_t) dy_t dx_t d\theta}_{\mathbb{H}[q(x_t)]} \quad (\text{B.6d})$$

$$= \mathbb{H}[q(y_t, x_t, \theta)] - \mathbb{H}[q(y_t, x_t)] - \mathbb{H}[q(x_t, \theta)] + \mathbb{H}[q(x_t)] \quad (\text{B.6e})$$

□

Now that we have our lemmas in place, we can continue with the proof of Theorem 1.

Proof.

$$\tilde{F}_{\tilde{p}}[q] = \int q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) \log \frac{q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u})}{\tilde{p}(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u})} d\mathbf{y} d\mathbf{x} d\theta d\mathbf{u} \quad (\text{B.7a})$$

$$= \int q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) \log \frac{q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u})}{p(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) \prod_{t=1}^T \tilde{p}(x_t) \tilde{p}(u_t) \tilde{p}(y_t, x_t)} d\mathbf{y} d\mathbf{x} d\theta d\mathbf{u} \quad (\text{B.7b})$$

$$= \underbrace{\int q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) \log \frac{q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u})}{p(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u})} d\mathbf{y} d\mathbf{x} d\theta d\mathbf{u}}_{F_p[q]} \\ + \int q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) \log \frac{1}{\prod_{t=1}^T \tilde{p}(x_t) \tilde{p}(u_t) \tilde{p}(y_t, x_t)} d\mathbf{y} d\mathbf{x} d\theta d\mathbf{u} \quad (\text{B.7c})$$

$$= F_p[q] - \sum_{t=1}^T \int q(\mathbf{y}, \mathbf{x}, \theta, \mathbf{u}) (\log \tilde{p}(x_t) + \log \tilde{p}(u_t) + \log \tilde{p}(y_t, x_t)) \quad (\text{B.7d})$$

$$= F_p[q] + \sum_{t=1}^T \\ - \int q(x_t) \log \tilde{p}(x_t) dx_t - \int q(u_t) \log \tilde{p}(u_t) du_t - \int q(y_t, x_t) \log \tilde{p}(y_t, x_t) dy_t dx_t. \quad (\text{B.7e})$$

Here, we can recognize the identities from [Lemma 3](#), [Lemma 4](#) and [Lemma 5](#).

$$F_p[q] + \sum_{t=1}^T - \int q(x_t) \log \tilde{p}(x_t) dx_t - \int q(u_t) \log \tilde{p}(u_t) du_t - \int q(y_t, x_t) \log \tilde{p}(y_t, x_t) dy_t dx_t \quad (\text{B.8a})$$

$$= F_p[q] + \sum_{t=1}^T \mathbb{H}[q(y_t, x_t)] - \mathbb{H}[q(x_t)] + \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_t, x_{t-1}, u_t)] \\ + \mathbb{H}[q(y_t, x_t, \theta)] + \mathbb{H}[q(x_t)] - \mathbb{H}[q(y_t, x_t)] - \mathbb{H}[q(x_t, \theta)] \quad (\text{B.8b})$$

$$= F_p[q] + \sum_{t=1}^T \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_t, x_{t-1}, u_t)] + \mathbb{H}[q(y_t, x_t, \theta)] - \mathbb{H}[q(x_t, \theta)] \quad (\text{B.8c})$$

□

C Additional Entropy Terms in Planning Objective

The adjusted inference objective in [Lázaro-Gredilla et al. \[2024\]](#) is phrased as a maximization problem including conditional entropies. In this appendix, we will rephrase the planning-as-inference objective into the vocabulary used in this paper. [Lázaro-Gredilla et al. \[2024\]](#) phrases their optimization problem as a maximization of a variational bound, whereas we pose the problem as a minimization of the Variational Free Energy. This means the entropy terms in [Lázaro-Gredilla et al. \[2024\]](#) have their sign flipped, as this entropy term is subtracted from the objective. Note that we also have a slightly different indexing for actions, as u_t leads to x_t , instead of leading to x_{t+1} as is the notation used by [Lázaro-Gredilla et al. \[2024\]](#). Here, we use $\mathbb{H}_q[z \mid \omega] = - \int q(z, \omega) \log q(z \mid \omega) dz d\omega$ as

the notation of the conditional entropy.

$$\mathbb{H}[q(x_0)] + \sum_{t=1}^T \mathbb{H}_q[x_t \mid x_{t-1}, u_t] \quad (\text{C.1a})$$

$$= \mathbb{H}[q(x_0)] + \sum_{t=1}^T - \int q(x_t, x_{t-1}, u_t) \log q(x_t \mid x_{t-1}, u_t) dx_t dx_{t-1} du_t \quad (\text{C.1b})$$

$$= \mathbb{H}[q(x_0)] + \sum_{t=1}^T - \int q(x_t, x_{t-1}, u_t) \log \frac{q(x_t, x_{t-1}, u_t)}{q(u_t \mid x_{t-1})q(x_{t-1})} dx_t dx_{t-1} du_t \quad (\text{C.1c})$$

$$= \mathbb{H}[q(x_0)] + \sum_{t=1}^T - \int q(x_t, x_{t-1}, u_t) \log \frac{q(x_t, x_{t-1}, u_t)}{q(x_{t-1})} dx_t dx_{t-1} du_t \\ + \int q(x_t, x_{t-1}, u_t) \log q(u_t \mid x_{t-1}) dx_t dx_{t-1} du_t \quad (\text{C.1d})$$

$$= \mathbb{H}[q(x_0)] + \sum_{t=1}^T \underbrace{- \int q(x_t, x_{t-1}, u_t) \log q(x_t, u_t \mid x_{t-1}) dx_t dx_{t-1} du_t}_{\mathbb{H}_q[x_t, u_t \mid x_{t-1}]} \\ + \sum_{t=1}^T \int q(x_{t-1}, u_t) \log \frac{q(x_{t-1}, u_t)}{q(x_{t-1})} dx_{t-1} du_t \quad (\text{C.1e})$$

$$= \mathbb{H}[q(x_0)] + \underbrace{\sum_{t=1}^T \mathbb{H}_q[x_t, u_t \mid x_{t-1}]}_{\mathbb{H}[q]} + \sum_{t=1}^T \underbrace{\int q(x_{t-1}, u_t) \log q(x_{t-1}, u_t) dx_{t-1} du_t}_{-\mathbb{H}[q(x_{t-1}, u_t)]} \\ - \underbrace{\int q(x_{t-1}) \log q(x_{t-1}) dx_{t-1}}_{-\mathbb{H}[q(x_{t-1})]} \quad (\text{C.1f})$$

$$= \mathbb{H}[q] + \sum_{t=1}^T \mathbb{H}[q(x_{t-1})] - \mathbb{H}[q(x_{t-1}, u_t)] \quad (\text{C.1g})$$

Since in our minimization scheme this entropy is subtracted from the objective, we subtract the additional terms in (C.1g) from the inference objective, and we obtain $\sum_{t=1}^T \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_{t-1})]$ as a final additional term.

D Consolidating State and Observation Epistemic Priors

Lemma 6. *Under the assumption that our posterior distribution factorizes as in Equation A.2, with*

$$\tilde{p}(x_t) = \exp\{-\mathbb{H}[q(y_t \mid x_t)]\}, \quad \tilde{p}(y_t, x_t) = \exp D_{\text{KL}}[q(\theta \mid y_t, x_t) \parallel q(\theta \mid x_t)],$$

define $\tilde{p}(x_t, \theta) = \exp\{-\mathbb{H}[q(y_t \mid x_t, \theta)]\}$. Then

$$- \int q(x_t) \log \tilde{p}(x_t) dx_t - \int q(y_t, x_t) \log \tilde{p}(y_t, x_t) dy_t dx_t = - \int q(x_t, \theta) \log \tilde{p}(x_t, \theta) dx_t d\theta. \quad (\text{D.1})$$

Proof.

$$S_t := - \int q(x_t) \log \tilde{p}(x_t) dx_t - \int q(y_t, x_t) \log \tilde{p}(y_t, x_t) dy_t dx_t \quad (\text{D.2a})$$

$$= \mathbb{H}[q(y_t, x_t, \theta)] - \mathbb{H}[q(x_t, \theta)] \quad (\text{by Lemmas 3 and 5}) \quad (\text{D.2b})$$

$$. \quad (\text{D.2c})$$

On the other hand,

$$R_t := - \int q(x_t, \theta) \log \tilde{p}(x_t, \theta) dx_t d\theta = \int q(x_t, \theta) \mathbb{H}[q(y_t | x_t, \theta)] dx_t d\theta \quad (\text{D.3a})$$

$$= - \int q(y_t, x_t, \theta) \log q(y_t | x_t, \theta) dy_t dx_t d\theta \quad (\text{D.3b})$$

$$= - \int q(y_t, x_t, \theta) (\log q(y_t, x_t, \theta) - \log q(x_t, \theta)) dy_t dx_t d\theta \quad (\text{D.3c})$$

$$= \mathbb{H}[q(y_t, x_t, \theta)] - \mathbb{H}[q(x_t, \theta)]. \quad (\text{D.3d})$$

Thus $S_t = R_t$, which proves (D.1). \square

E Proof of Theorem 2

In this appendix, we establish the stationary conditions for the Active Inference message-passing scheme presented in Theorem 2. The proof proceeds by deriving the first-order optimality conditions for each coordinate in the adjusted system.

We begin by expanding the Bethe Free Energy for the generative model (1) and establishing the necessary consistency constraints in (E.3). subsection E.1 demonstrates that the standard region coordinates lead to a degenerate optimization problem (Theorem 7), motivating the introduction of the channel variable $r_{y|x\theta,t}$. subsection E.2 and subsection E.3 then derive the stationary conditions with respect to each coordinate—the observation factor belief $q_{y,t}$ (Theorem 8), the channel $r_{y|x\theta,t}$ (Theorem 9), and the dynamics factor belief $q_{\text{dyn},t}$ (Theorem 10)—as well as the identifications of the Lagrange multipliers $\Lambda_{x\theta}$ (Theorem 11) and Λ_{trip} (Theorem 12).

Theorem 2 follows directly from these results: equations (8)–(12) are simply the collected stationary conditions established in Theorem 8–Theorem 12 below, expressed in the notation of the main text with the goal priors $\hat{p}_y(y_t)$ explicitly included.

BFE for the model (1) then reads

$$\begin{aligned} F_p^{\text{Bethe}}[q] &= \mathbb{D}[q_\theta(\theta) || p(\theta)] + \mathbb{D}[q_{x_0}(x_0) || p(x_0)] \\ &+ \sum_{t=1}^T \left[\mathbb{D}[q_{y,t}(y_t, x_t, \theta) || p(y_t | x_t, \theta)] + \mathbb{D}[q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) || p(x_t | x_{t-1}, \theta, u_t)] \right. \\ &\quad \left. + \mathbb{D}[q_{u,t}(u_t) || p(u_t)] + \mathbb{D}[q_{x,\hat{p},t}(x_t) || \hat{p}_x(x_t)] + \mathbb{D}[q_{y,\hat{p},t}(y_t) || \hat{p}_y(y_t)] \right] \\ &+ (d_\theta - 1) \mathbb{H}[q_\theta(\theta)] + (d_{x_0} - 1) \mathbb{H}[q_{x_0}(x_0)] \\ &+ \sum_{t=1}^T \left[(d_{x_t} - 1) \mathbb{H}[q_{x_t}(x_t)] + (d_{y_t} - 1) \mathbb{H}[q_{y_t}(y_t)] + (d_{u_t} - 1) \mathbb{H}[q_{u_t}(u_t)] \right]. \end{aligned} \quad (\text{E.1})$$

The variable-node degrees implied by (1) and the factorization above are

$$\begin{aligned} d_\theta &= 1 + 2T \quad (\text{prior}, T \text{ obs}, T \text{ dyn}), & d_{x_0} &= 2 \quad (\text{prior}, \text{dyn } t=1), \\ d_{y_t} &= 2 \quad (\text{obs}, \text{goal prior on } y_t), & d_{u_t} &= 2 \quad (\text{action prior}, \text{dyn } t), \\ d_{x_t} &= \begin{cases} 4, & 1 \leq t \leq T-1 \quad (\text{obs } t, \text{dyn } t, \text{dyn } t+1, \text{goal prior on } x_t), \\ 3, & t = T \quad (\text{obs } T, \text{dyn } T, \text{goal prior on } x_T). \end{cases} \end{aligned} \quad (\text{E.2})$$

(For unary factors $f_\theta, f_{x_0}, f_{u,t}, f_{x,\hat{p},t}, f_{y,\hat{p},t}$ we identify the factor belief with the adjacent singleton.) With this notation, the Bethe Free Energy in (E.1) is the specialization of the general BFE in subsection A.3 to (1).

Local consistency requires that, for every factor a and every variable $i \in s_a$,

$$\int q_a(s_a) ds_{a \setminus i} = q_i(s_i).$$

$$\textbf{Observation } (y, t) : \int q_{y,t}(y_t, x_t, \theta) dx_t d\theta = q_{y_t}(y_t), \quad (\text{E.3a})$$

$$\int q_{y,t}(y_t, x_t, \theta) dy_t d\theta = q_{x_t}(x_t), \quad (\text{E.3b})$$

$$\int q_{y,t}(y_t, x_t, \theta) dy_t dx_t = q_\theta(\theta). \quad (\text{E.3c})$$

$$\textbf{Dynamics } (\text{dyn}, t) : \int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) dx_{t-1} du_t d\theta = q_{x_t}(x_t), \quad (\text{E.3d})$$

$$\int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) dx_t du_t d\theta = q_{x_{t-1}}(x_{t-1}), \quad (\text{E.3e})$$

$$\int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) dx_t dx_{t-1} d\theta = q_{u_t}(u_t), \quad (\text{E.3f})$$

$$\int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) dx_t dx_{t-1} du_t = q_\theta(\theta). \quad (\text{E.3g})$$

$$\textbf{Unary priors/goals} : q_{u,t}(u_t) = q_{u_t}(u_t), \quad q_{x,\hat{p},t}(x_t) = q_{x_t}(x_t), \quad q_{y,\hat{p},t}(y_t) = q_{y_t}(y_t). \quad (\text{E.3h})$$

(For unary factors we identify the factor belief with the adjacent singleton belief; the equality is enforced by the corresponding consistency constraint.)

E.1 Degeneracy of the marginal scheme

Theorem 7 (Degeneracy persists under the augmented region coordinates). *Fix t . Consider the Bethe-form objective specialized to (1), augmented by the Active Inference entropic correction*

$$+ \mathbb{H}[q(y_t, x_t, \theta)] - \mathbb{H}[q(x_t, \theta)] + \mathbb{H}[q(x_{t-1}, u_t)] - \mathbb{H}[q(x_t, x_{t-1}, u_t)].$$

Introduce the auxiliary region beliefs $q_{\text{sep},t}(x_t, \theta)$ (defined in Equation 6) with the consistency constraints

$$\int q_{y,t}(y_t, x_t, \theta) dx_t d\theta = q_{y_t}(y_t), \quad (\text{E.4a})$$

$$\int q_{y,t}(y_t, x_t, \theta) dy_t = q_{\text{sep},t}(x_t, \theta). \quad (\text{E.4b})$$

(Region beliefs $q_{\text{pair},t}$ and $q_{\text{trip},t}$ are tied to the dynamics block and do not appear in (E.4).) Then any stationary point of the Lagrangian with respect to the observation-factor belief $q_{y,t}(y_t, x_t, \theta)$ satisfies

$$-\log p(y_t | x_t, \theta) + \lambda_y(y_t) + \lambda_{\text{sep}}(x_t, \theta) = 0, \quad (\text{E.5})$$

for some multipliers $\lambda_y(\cdot)$, $\lambda_{\text{sep}}(\cdot, \cdot)$. Consequently:

1. If $p(y_t | x_t, \theta)$ is not separable as $a_t(y_t) b_t(x_t, \theta)$, the system is infeasible (no interior solution for $q_{y,t}$).
2. If $p(y_t | x_t, \theta) = a_t(y_t) b_t(x_t, \theta)$ is separable, the observation block is affine in $q_{y,t}$ (its second variation w.r.t. $q_{y,t}$ vanishes), hence stationary points are non-unique (a flat face of the feasible polytope).

Proof. The $q_{y,t}$ -dependent part of the objective is

$$\int q_{y,t} \log \frac{q_{y,t}}{p(y_t | x_t, \theta)} + \mathbb{H}[q(y_t, x_t, \theta)],$$

since $-\mathbb{H}[q(x_t, \theta)]$, $+\mathbb{H}[q(x_{t-1}, u_t)]$, and $-\mathbb{H}[q(x_t, x_{t-1}, u_t)]$ do not depend on $q_{y,t}$. The $\int q_{y,t} \log q_{y,t}$ from the KL term cancels exactly with $+\mathbb{H}[q(y_t, x_t, \theta)]$, leaving the linear functional $-\int q_{y,t} \log p(y_t | x_t, \theta)$. Add constraints (E.4a)–(E.4b) with multipliers $\lambda_y(\cdot)$, $\lambda_{\text{sep}}(\cdot, \cdot)$. Taking the functional derivative yields (E.5). Exponentiating shows that a solution exists only if $p(y_t | x_t, \theta) \propto e^{\lambda_y(y_t)} e^{\lambda_{\text{sep}}(x_t, \theta)}$, i.e., it factorizes as $a_t(y_t) b_t(x_t, \theta)$. In the separable case, the absence of a $\int q_{y,t} \log q_{y,t}$ term implies zero curvature in the $q_{y,t}$ -direction and thus non-uniqueness; otherwise the system is infeasible. \square

Consequence. [Theorem 7](#) shows that even after introducing the augmented region coordinates $(q_{\text{sep},t}, q_{\text{pair},t}, q_{\text{trip},t})$, the Active Inference correction leaves the observation block degenerate: feasibility requires a separable likelihood in y_t and (x_t, θ) , and otherwise the block is flat.

E.2 Stationary conditions in the conditional adjusted system

Lemma 8 (Stationary condition for the observation factor with channel augmentation). *Fix t and introduce the channel $r_{y|x\theta,t}(y_t | x_t, \theta)$ with the normalization $\int r_{y|x\theta,t}(y_t | x_t, \theta) dy_t = 1$. Assume the separator consistency on the overlap (x_t, θ) ,*

$$\int q_{y,t}(y_t, x_t, \theta) dy_t = q_{x\theta}(x_t, \theta) = \int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) dx_{t-1} du_t,$$

and the y_t -singleton consistency

$$\int q_{y,t}(y_t, x_t, \theta) dx_t d\theta = q_{y_t}(y_t).$$

Consider the observation-block objective (holding $r_{y|x\theta,t}$ fixed):

$$\int q_{y,t} \log \frac{q_{y,t}}{p(y_t | x_t, \theta)} + \int q_{y,t} [-\log r_{y|x\theta,t}(y_t | x_t, \theta)],$$

plus Lagrange terms for the two consistency constraints. Then the stationarity with respect to $q_{y,t}(y_t, x_t, \theta)$ is

$$\log q_{y,t}(y_t, x_t, \theta) - \log p(y_t | x_t, \theta) - \log r_{y|x\theta,t}(y_t | x_t, \theta) + \lambda_y(y_t) + \lambda_{x\theta}(x_t, \theta) = 0, \quad (\text{E.6})$$

for some multipliers $\lambda_y(\cdot)$ and $\lambda_{x\theta}(\cdot, \cdot)$. Equivalently,

$$q_{y,t}(y_t, x_t, \theta) \propto p(y_t | x_t, \theta) r_{y|x\theta,t}(y_t | x_t, \theta) \exp\{-\lambda_y(y_t)\} \exp\{-\lambda_{x\theta}(x_t, \theta)\}, \quad (\text{E.7})$$

with $\lambda_y, \lambda_{x\theta}$ chosen to satisfy the two projection constraints above.

Proof. Form the partial Lagrangian (omitting the redundant normalization of $q_{y,t}$):

$$\begin{aligned} \mathcal{L}[q_{y,t}] = & \int q_{y,t} \log \frac{q_{y,t}}{p(y_t | x_t, \theta)} + \int q_{y,t} [-\log r_{y|x\theta,t}] \\ & + \int \lambda_y(y_t) \left(\int q_{y,t} dx_t d\theta - q_{y_t}(y_t) \right) dy_t \\ & + \iint \lambda_{x\theta}(x_t, \theta) \left(\int q_{y,t} dy_t - q_{x\theta}(x_t, \theta) \right) dx_t d\theta. \end{aligned}$$

Taking the functional derivative w.r.t. $q_{y,t}$ and setting it to zero yields

$$\log q_{y,t} - \log p(y_t | x_t, \theta) - \log r_{y|x\theta,t} + 1 + \lambda_y(y_t) + \lambda_{x\theta}(x_t, \theta) = 0,$$

where the constant +1 can be absorbed into either multiplier. This is (E.6). Exponentiating gives (E.7), and the multipliers are determined by enforcing the two linear projection constraints. \square

Lemma 9 (Stationary condition for the observation channel). *Fix t . Let the channel $r_{y|x\theta,t}(y_t | x_t, \theta)$ satisfy the row-normalization*

$$\int r_{y|x\theta,t}(y_t | x_t, \theta) dy_t = 1 \quad \text{for all } (x_t, \theta), \quad (\text{E.8})$$

and assume separator consistency on the overlap (x_t, θ) :

$$\int q_{y,t}(y_t, x_t, \theta) dy_t = q_{\text{sep},t}(x_t, \theta). \quad (\text{E.9})$$

Consider the channel objective (holding $q_{y,t}$ and $q_{\text{sep},t}$ fixed)

$$\mathcal{J}[r] = \int q_{y,t}(y_t, x_t, \theta) [-\log r_{y|x\theta,t}(y_t | x_t, \theta)] dy_t dx_t d\theta, \quad (\text{E.10})$$

subject to (E.8). Then a stationary point is given pointwise by

$$r_{y|x\theta,t}^*(y_t | x_t, \theta) = \frac{q_{y,t}(y_t, x_t, \theta)}{q_{\text{sep},t}(x_t, \theta)} \quad \text{whenever } q_{\text{sep},t}(x_t, \theta) > 0, \quad (\text{E.11})$$

with the convention that if $q_{\text{sep},t}(x_t, \theta) = 0$, the entire row $r_{y|x\theta,t}(\cdot | x_t, \theta)$ can be chosen arbitrarily as any probability distribution (it does not affect \mathcal{J}).

Proof. Form the Lagrangian with a row multiplier $\lambda(x_t, \theta)$ enforcing (E.8):

$$\mathcal{L}[r] = \int q_{y,t}(y_t, x_t, \theta) [-\log r(y_t | x_t, \theta)] + \iint \lambda(x_t, \theta) \left(\int r(y_t | x_t, \theta) dy_t - 1 \right) dx_t d\theta.$$

Taking the functional derivative and setting it to zero yields, pointwise in (y_t, x_t, θ) ,

$$-\frac{q_{y,t}(y_t, x_t, \theta)}{r^*(y_t | x_t, \theta)} + \lambda(x_t, \theta) = 0 \quad \Rightarrow \quad r^*(y_t | x_t, \theta) = \frac{q_{y,t}(y_t, x_t, \theta)}{\lambda(x_t, \theta)}.$$

Imposing the row-normalization (E.8) and using (E.9),

$$1 = \int r^*(y_t | x_t, \theta) dy_t = \frac{\int q_{y,t}(y_t, x_t, \theta) dy_t}{\lambda(x_t, \theta)} = \frac{q_{\text{sep},t}(x_t, \theta)}{\lambda(x_t, \theta)},$$

so $\lambda(x_t, \theta) = q_{\text{sep},t}(x_t, \theta)$, giving (E.11). If $q_{\text{sep},t}(x_t, \theta) = 0$, then the row of $q_{y,t}$ is identically zero and \mathcal{J} is unaffected by the choice of $r(\cdot | x_t, \theta)$. \square

Lemma 10 (Stationary condition for the dynamics factor (minimal projections)). *Fix t . Let $q_{\text{dyn},t}(x_t, x_{t-1}, u_t, \theta)$ be the dynamics-factor belief and introduce the region beliefs $q_{x\theta,t}(x_t, \theta)$ and $q_{\text{trip},t}(x_t, x_{t-1}, u_t)$ with projection constraints*

$$\int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) dx_{t-1} du_t = q_{x\theta,t}(x_t, \theta), \quad (\text{E.12a})$$

$$\int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) d\theta = q_{\text{trip},t}(x_t, x_{t-1}, u_t). \quad (\text{E.12b})$$

(Separately, enforce the pair-trip relation $\int q_{\text{trip},t}(x_t, x_{t-1}, u_t) dx_t = q_{\text{pair},t}(x_{t-1}, u_t)$ in the $q_{\text{trip},t}$ block.) The $q_{\text{dyn},t}$ -dependent part of the objective is the KL term

$$\int q_{\text{dyn},t} \log \frac{q_{\text{dyn},t}}{p(x_t | x_{t-1}, \theta, u_t)} dx_t dx_{t-1} du_t d\theta,$$

while the entropic corrections only involve the region beliefs. Form the partial Lagrangian with multipliers $\Lambda_{x\theta}(x_t, \theta)$ and $\Lambda_{\text{trip}}(x_t, x_{t-1}, u_t)$ enforcing (E.12a)–(E.12b). Then the stationarity w.r.t. $q_{\text{dyn},t}$ is

$$\log q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) - \log p(x_t | x_{t-1}, \theta, u_t) + \Lambda_{x\theta}(x_t, \theta) + \Lambda_{\text{trip}}(x_t, x_{t-1}, u_t) = 0, \quad (\text{E.13})$$

up to an additive constant, hence

$$q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) \propto p(x_t | x_{t-1}, \theta, u_t) \exp\{-\Lambda_{x\theta}(x_t, \theta)\} \exp\{-\Lambda_{\text{trip}}(x_t, x_{t-1}, u_t)\}. \quad (\text{E.14})$$

The multipliers are determined by the two projection constraints (E.12), while the consistency $\int q_{\text{trip},t} dx_t = q_{\text{pair},t}$ is enforced in the $q_{\text{trip},t}$ variation and does not appear in (E.14).

E.3 Identification of the Lagrange multipliers

Lemma 11 (Identification of the dynamics-side separator multiplier). *Fix t . Assume the channel $r_{y|x\theta,t}(y_t | x_t, \theta)$ is row-normalized and impose separator consistency*

$$\int q_{y,t}(y_t, x_t, \theta) dy_t = q_{\text{sep},t}(x_t, \theta) = \int q_{\text{dyn},t}(x_t, x_{t-1}, \theta, u_t) dx_{t-1} du_t.$$

At any stationary point of the observation block, there exists a slice-constant $C_t > 0$ such that

$$\exp\{-\Lambda_{x\theta}(x_t, \theta)\} = C_t \frac{\int p(y_t | x_t, \theta) r_{y|x\theta,t}(y_t | x_t, \theta) e^{-\lambda_y(y_t)} dy_t}{q_{\text{sep},t}(x_t, \theta)}, \quad (\text{E.15})$$

where $\lambda_y(\cdot)$ is the Lagrange multiplier enforcing $\int q_{y,t}(y_t, x_t, \theta) dx_t d\theta = q_{y_t}(y_t)$. Equivalently,

$$\Lambda_{x\theta}(x_t, \theta) = \log q_{\text{sep},t}(x_t, \theta) - \log \left(\int p(y_t | x_t, \theta) r_{y|x\theta,t}(y_t | x_t, \theta) e^{-\lambda_y(y_t)} dy_t \right) + c_t,$$

with $c_t = -\log C_t$ independent of (x_t, θ) .

Proof. From the observation–block stationarity (Lemma 8), there exists a slice–constant $\kappa_t > 0$ such that

$$q_{y,t}(y_t, x_t, \theta) = \kappa_t p(y_t | x_t, \theta) r_{y|x\theta,t}(y_t | x_t, \theta) e^{-\lambda_y(y_t)} e^{-\lambda_{x\theta}(x_t, \theta)}. \quad (\text{E.16})$$

Integrate (E.16) over y_t and use the left separator equality to obtain

$$q_{\text{sep},t}(x_t, \theta) = \kappa_t e^{-\lambda_{x\theta}(x_t, \theta)} \underbrace{\int p(y_t | x_t, \theta) r_{y|x\theta,t}(y_t | x_t, \theta) e^{-\lambda_y(y_t)} dy_t}_{=: I_t(x_t, \theta)}.$$

Solve for $e^{-\lambda_{x\theta}(x_t, \theta)}$:

$$e^{-\lambda_{x\theta}(x_t, \theta)} = \frac{q_{\text{sep},t}(x_t, \theta)}{\kappa_t I_t(x_t, \theta)}.$$

On the other hand, varying the separator $q_{\text{sep},t}$ in the Lagrangian for the two equalities gives $\lambda_{x\theta}(x_t, \theta) + \Lambda_{x\theta}(x_t, \theta) = 0$, hence $e^{-\Lambda_{x\theta}(x_t, \theta)} = e^{\lambda_{x\theta}(x_t, \theta)}$. Combining the two displays,

$$e^{-\Lambda_{x\theta}(x_t, \theta)} = e^{\lambda_{x\theta}(x_t, \theta)} = \frac{\kappa_t I_t(x_t, \theta)}{q_{\text{sep},t}(x_t, \theta)} = C_t \frac{\int p(y_t | x_t, \theta) r_{y|x\theta,t}(y_t | x_t, \theta) e^{-\lambda_y(y_t)} dy_t}{q_{\text{sep},t}(x_t, \theta)},$$

with $C_t = \kappa_t$ independent of (x_t, θ) . Taking logs yields the equivalent additive form with $c_t = -\log C_t$. \square

Lemma 12 (Identification of the triplet multiplier from the dynamics-side entropic correction). *Fix t . Let the region beliefs $q_{\text{trip},t}(x_t, x_{t-1}, u_t)$ and $q_{\text{pair},t}(x_{t-1}, u_t)$ satisfy the coupling constraint*

$$\int q_{\text{trip},t}(x_t, x_{t-1}, u_t) dx_t = q_{\text{pair},t}(x_{t-1}, u_t), \quad (\text{E.17})$$

and let the projection from the dynamics factor be enforced via the multiplier $\Lambda_{\text{trip}}(x_t, x_{t-1}, u_t)$ on $\int q_{\text{dyn},t} d\theta - q_{\text{trip},t} = 0$. Suppose the objective contains the dynamics-side entropic correction $+H[q_{\text{pair},t}] - H[q_{\text{trip},t}]$. Then, at any stationary point, there exists a slice-constant $C_t > 0$ such that

$$\exp\{-\Lambda_{\text{trip}}(x_t, x_{t-1}, u_t)\} = C_t \frac{q_{\text{pair},t}(x_{t-1}, u_t)}{q_{\text{trip},t}(x_t, x_{t-1}, u_t)}, \quad (\text{E.18})$$

equivalently,

$$\Lambda_{\text{trip}}(x_t, x_{t-1}, u_t) = \log q_{\text{trip},t}(x_t, x_{t-1}, u_t) - \log q_{\text{pair},t}(x_{t-1}, u_t) + c_t,$$

with $c_t = -\log C_t$ independent of (x_t, x_{t-1}, u_t) .

Proof. Consider the part of the Lagrangian involving $q_{\text{trip},t}$ and $q_{\text{pair},t}$:

$$\begin{aligned} \mathcal{L} = & \underbrace{\int q_{\text{trip},t} \log q_{\text{trip},t}}_{\text{from } -H[q_{\text{trip},t}]} - \underbrace{\int q_{\text{pair},t} \log q_{\text{pair},t}}_{\text{from } +H[q_{\text{pair},t}]} \\ & + \iint \Xi_{\text{pair}}(x_{t-1}, u_t) \left(\int q_{\text{trip},t} dx_t - q_{\text{pair},t} \right) dx_{t-1} du_t \\ & - \int \Lambda_{\text{trip}} q_{\text{trip},t} \end{aligned}$$

where the last term comes from the projection $\int q_{\text{dyn},t} d\theta - q_{\text{trip},t} = 0$. Varying w.r.t. $q_{\text{trip},t}$ gives

$$\frac{\delta \mathcal{L}}{\delta q_{\text{trip},t}} : \log q_{\text{trip},t} + 1 + \Xi_{\text{pair}}(x_{t-1}, u_t) - \Lambda_{\text{trip}}(x_t, x_{t-1}, u_t) = 0.$$

Varying w.r.t. $q_{\text{pair},t}$ yields

$$\frac{\delta \mathcal{L}}{\delta q_{\text{pair},t}} : -(\log q_{\text{pair},t} + 1) - \Xi_{\text{pair}}(x_{t-1}, u_t) = 0 \Rightarrow \Xi_{\text{pair}}(x_{t-1}, u_t) = -(\log q_{\text{pair},t}(x_{t-1}, u_t) + 1).$$

Eliminating Ξ_{pair} in the first equation gives

$$\Lambda_{\text{trip}}(x_t, x_{t-1}, u_t) = \log q_{\text{trip},t}(x_t, x_{t-1}, u_t) - \log q_{\text{pair},t}(x_{t-1}, u_t) + c_t,$$

where the slice-constant c_t (absorbing the $+1$ terms and any normalization) is independent of (x_t, x_{t-1}, u_t) . Exponentiating yields (E.18) with $C_t = e^{-c_t}$. \square