

Literature Review

An Analysis of Active Inference and Reinforcement Learning Paradigms in
Partially Observable Environments.

Fraser Paterson

A review of the extant literature, pursuant to the requirements of the
Degree: Bachelor of Science (Honours).



Supervisor: Dr Tim French
Department of Computer Science and Software Engineering
The University of Western Australia
24 April 2023

CONTENTS

| | | |
|------------|---|----------|
| I | Overview | 2 |
| II | Introduction | 2 |
| II-A | Active Inference: An Overview | 2 |
| II-B | Why Care About Active Inference? | 2 |
| II-C | Variational Inference | 3 |
| II-D | The Free Energy Principle | 3 |
| II-E | Present Questions | 4 |
| II-F | Active Inference: Key Concepts | 4 |
| II-G | Overview of Research Direction | 4 |
| III | Gaps in The Literature | 5 |
| IV | Previous Work | 5 |
| IV-A | Sampling Based Approximation Methods | 5 |
| IV-B | Factor Graphs and Message Passing Methods | 6 |
| V | Conclusion | 6 |

I. OVERVIEW

This section should be an overview of the Lit review. Can possibly put some information related to the project in here.

II. INTRODUCTION

A. Active Inference: An Overview

The field of adaptive, agent-based AI is perhaps, entering its renacance. The last few years have seen major advances in the capability, sophistication and operation of agent-based methods from Deep Reinforcement Learning to Active Inference...

Active Inference is an emerging first-principles, account of adaptive behavior. Originating from Neuroscience: Friston, Kilner, and Harrison [14], Knill and Pouget [17] and Friston et al. [10] as a corollary of the “Free Energy Principle”, the theory is increasingly making inroads into Machine Learning and Artificial Intelligence: Friston, Daunizeau, and Kiebel [16] and Millidge [22].

Active inference is a highly ambitious theory, as it purports to offer a fully unified account of action, perception and learning: Friston [13]. In Active Inference, these three cognitive processes are integrated under the rubric of variational inference via the minimisation of variational free energy.

The basic postulate of the theory is that adaptive systems like living organisms, will act to fulfill prior expectations or “preferences” which encode desirable states for that system. These agents make use of a generative model to afford predictions about hidden states of their environment. These predictions are, in turn, used to produce inferences about the causes of observations and also to inform action-selection.

Under Active Inference, perception is cast as the problem of inferring the hidden states of the agent’s environment, given the agent’s sensory observations. Action-selection and planning can be described as inference on policies over trajectories into the future. Finally, learning is viewed as inference over the parameters of the agent’s generative model, that best account for the agent’s sensory data. “Learning” of this kind takes place over a slower timescale than does perceptual inference. Active Inference describes all three processes in terms of variational free energy minimisation over a single functional. This functional is composed of the agent’s beliefs about the hidden states of the environment, in addition to the sensory surprisal of its observations. It should be noted that this “divergence minus evidence” formulation of the free energy is simply one of many, there are several other ways to parameterise the free energy, such as “energy minus entropy” or “complexity minus accuracy”. I shall adhere to the “divergence minus evidence” formulation for the remainder of this review.

To elucidate these constructs in greater detail, it will be necessary to talk about variational inference.

B. Why Care About Active Inference?

Active Inference addresses a problem which has plagued value-function formulations of adaptivity since their inception. This is the issue of sample-efficiency and of learning in the presence of sparse rewards: Christiano [4] and Dulac-Arnold et al. [8]. If all that is available to the agent for action-selection, is a value function mapping from states and/or actions to an extrinsic reward signal, it is necessary to observe a great deal many state or state-action pairs to learn the optimal mapping from state/state-action to reward signal. For problems with large state-action spaces, this presents a significant challenge to such methods. Active Inference eschews this issue by placing “information-gain” on the same footing as value maximization and hence drives the agent to dynamically trade-off attention between these two goals.

Since the Free Energy is an upper bound on sensory surprisal, and since the minimisation of free energy is the sole imperative under Active inference, both action and perception have as their effect, the reduction of sensory surprisal. This is a crucial point of difference between Active Inference and Reinforcement Learning agents. While Reinforcement Learning agents select actions in an attempt to maximise a reward function of states (or states and actions), Active Inference agents select actions so as to minimise a free energy functional composed of the following two parts:

- 1) The discrepancy - KL divergence - between the agent’s prior preference for a certain observation, and the actual observation garnered as a consequence of performing a certain action.
- 2) The “surprisal” - negative log probability - of the observation. This is essentially a measure of how unlikely the observation is given the agent’s model.

This formulation of the Active Inference agent’s objective function might appear to be unnecessarily convoluted, especially since these agents have essentially the same goal as their Reinforcement Learning cousins: adaption to the constraints of some environment. There is, however, something very special about this formulation. In addition to a “pragmatic” or “reward-maximising” imperative, encoded by the divergence between the agent’s preferred and actual observations, the free energy functional affords

an additional “epistemic” imperative for information-gain, encoded via the surprisal. This means that Active Inference agents have a built-in affordance for exploratory/curious behaviour, in addition to that of maximising the extrinsic value of realising their preferred observations.

Mere extrinsic, value-maximization is typically the *only* operative imperative in Reinforcement Learning approaches; Sutton and Barto [28]. It is not impossible to endow Reinforcement Learning agents with a drive toward information-gain or “curiosity”. Though typically, one has to contrive some *ad hoc* manipulation of the reward signal to hand-craft an *encoding* of the epistemic imperative: Pathak et al. [26]. In a nutshell, Active Inference replaces value functions with variational free energy functionals of Bayesian beliefs. Active Inference agents therefore have a built-in affordance for dynamically trading-off pragmatic and epistemic imperatives. This means that in the face of sparse rewards, or a persistent failure to realise prior preferences, Active Inference agents will naturally engage in exploratory behaviour, in an attempt to find as-yet unknown routes to realise its prior preferences.

The primary appeal of the Active Inference formulation of intelligence/adaptive Behaviour is twofold. First, it unifies the study of action, perception and planning under a single imperative, that of minimizing variational - or expected - free energy. This is an efficient formulation of these problems, since one need only address a single methodological principle instead of three. Parsimony of this kind is always desirable in any scientific theory - all things being equal - since the generality of a theory is very often a good measure of its predictive power. Second, as just elaborated, Active Inference affords a much greater sample-efficiency than does Reinforcement Learning: Tschantz et al. [30]. This is an especially intriguing aspect of the theory, since the majority of real-world problems are characterised by a sparsity of rewards. Hence Active Inference is plausibly posed to inaugurate a new era of real-world agent-based AI.

Lastly, Active inference is interesting because of its promise to be such a general method and indeed owing to a growing body of empirical research to suggest that free-energy minimization is what the brain is doing. Since the brain is thought to be the seat of “natural” intelligence, evidence attesting to the brain’s function as a “free-energy minimizing machine” must surely be of interest to we who are concerned with generating instances of intelligence, artificially.

This is fine, though try to avoid “grandiose” claims. A better approach is to lead with an explanation as to what it is that Active Inference solves and perhaps how it is novel in a useful way.

Since Active Inference makes use of Variational Inference, it will be necessary to understand the fundements of the latter.

C. Variational Inference

Variational inference: Blei, Kucukelbir, and McAuliffe [2] is a technique of approximate Bayesian inference, in the case that the exact inference procedure becomes intractible, typically due to the large or infinite number of states over which it is necessary to marginalise. The goal in all Bayesian inference methods is to compute the posterior distribution over some variable, given the prior and likelihood distributions. These latter encode - respectively - the supposed causal relationship between the variable of interest and a observation, in addition to a “prior” belief about the distribution of the variable at issue. If we wish to infer the posterior distribution of x as a consequence of observing y , then this is given as the distribution: $p(x|y)$ and bayes’s rule gives:

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

For all but the smallest problems, computing the posterior exactly is intractible, due to the sheer number of states over which it is necessary to sum/integrate when computing $p(y)$. To eschew this, variational inference proposes an approximation scheme, whereby a family of “variational” or “approximate” posterior distributions is posited, each of which is a potential, approximate solution to the true posterior. These approximate posteriors are usually denoted: $q_i(x)$ where the subscript denotes that this is the i th member of the family. We select a particular approximate posterior by choosing the one with the smallest “divergence” from the true posterior. Typically this is the Kullback-Libeler Divergence (KL Divergence).

D. The Free Energy Principle

Historically speaking, Active Inference is a derivative of the “Free Energy Principle”, which is a theoretical principle thought to plausibly offer a unified, constitutive account of brain function: Friston [12] and perhaps even of life itself: Friston [11].

The Free Energy Principle (FEP) attempts to provide a principled account of adaptivity per-se. The basic insight is that if a system maintains a boundary between itself and some exterior environment,

and if this system persists over time, then this system must be capable of resisting the thermodynamic tendency toward the dissolution of the boundary which separates its internal states from the external environment. To do this, the system must maintain the configuration of its internal states so as to exist within some desired homeostatic bounds. The way the system can achieve this is by maximizing the Bayesian model-evidence for its configuration of internal states. Hence we have internal states, external states and “blanket” states, where the blanket states are themselves composed of “active” and “sensory” states. Internal states can only be directly influenced by the sensory states of the blanket, and external states can only be influenced by the active states of the blanket. The Bayesian model evidence for the system’s configuration of internal states is bounded above by the free energy of the system’s Bayesian beliefs about the sensory states. Hence the “Free Energy” principle, since the Free Energy serves as a tractable upper bound to the system’s model evidence. Thus the principle culminates in the following statement. Any system which persists across time, in the face of a tendency to thermodynamic dissolution must act as if it is soliciting evidence for its own existence, via a model of the world.

E. Present Questions

Although Active Inference offers several exciting new directions in agent-based AI and might hold the key to truly sample-efficient real-world implementations, the theory has typically only been implemented on relatively trivial problem instances, with a small number of states and/or actions, most commonly in a discrete setting: **Uncertainty Epistemics AIF Saccad** and Friston et al. [15]. Although it has been used to great effect in these “proof of concept” cases, it is not yet applicable to the same sorts of problems that reinforcement learning can currently address. Many of the sorts of real-world applications of practical interest are thus far beyond the reach of Active Inference, especially for real-time implementations. The main difficulty stems from the fact that the variational free energy is a functional of hidden states and observations. Hidden states usually exist in a very high dimensional space and observations are usually highly time-varying. The task of minimising a highly dimensional, time-varying functional is non-trivial. The goal of scaling up the method to problems with larger state and/or action spaces, such as in the continuous case, is very much an open one, and one that must be satisfactorily addressed if Active Inference is to become a serious contender as a real-world method.

F. Active Inference: Key Concepts

Provide brief overview of key concepts and theories to be discussed in the review

Things that will some elaboration, or at least a definition:

- 1) Bayesian Inference - absolutely
- 2) MDP and POMDP
- 3) Variational Inference
- 4) Variational Free Energy
- 5) Expected Free Energy
- 6) Model-Based vs Model-Free Methods - absolutely
- 7) Factor Graphs and Message Passing - absolutely
- 8) Policy (Reinforcement Learning vs Active Inference framing)
- 9) Amortized Inference

There are too many here to do full justice in the manner you desire. Maintain a glossary of key terms, in addition.

G. Overview of Research Direction

The central aim of my proposed research topic is twofold. The first constituent aim is to investigate the relative merits/demerits of active inference as a real-world control and optimization strategy, against reinforcement learning baselines. I propose to investigate this by framing the question of real-world suitability in terms of the approach’s ability to afford fast and reliable solutions in noisy, uncertain or partially observable environments. the subsequent active inference agents I develop will be compared to Reinforcement Learning baselines.

The second aim concerns the potential for “scaling up” Active Inference methods to continuous and/or higher-dimensional state-spaces. This is a natural corollary to the first aim, since if we are interested in the suitability of Active Inference as a real-world control and optimization technique, it is not enough to simply determine if it can favorably compare to an established method in the noisy case. Since the real-world tasks of interest are overwhelmingly characterized by a high degree of dimensionality, it is necessary to investigate the performance of Active inference in high-dimensional settings.

Thus are the central questions raised in this Thesis:

- Are Active Inference agents more robust to noisy observations and non-stationarity than a comparable RL baseline?
- What are the most promising avenues of investigation in the attempt to scale up active inference to larger problem instances?

III. GAPS IN THE LITERATURE

Though there has been much focus on the implementation of active inference methods for small, discrete state-action spaces: Millidge [22], Da Costa et al. [6], Smith, Friston, and Whyte [27], Da Costa et al. [7] and Friston et al. [15]. The method is not currently viable for practical use in larger or continuous state-action spaces, for which it is necessary to plan future actions over some time horizon. Owing to the relatively small size of the state-action spaces in which active inference has historically been implemented, it has been possible to simply evaluate the expected free energy of all possible actions over the specified time horizon. This owes primarily to the issue of evaluating the expected free energy, which is the expectation of the Variational free energy evaluated for future actions over some time horizon: Koudahl, Buckley, and Vries [18] and Çatal et al. [3].

Unfortunately, enumerating all possible action-trajectories over the specified time horizon does not scale well to problems with larger state-action spaces and/or longer time horizons. Hence we can now specify exactly what it is that the problem of “scaling” is supposed to be.

Let S be a solution technique. Let P_1 and P_2 be problem instances of the same type. Let X and Y be the solution spaces for P_1 and P_2 (respectfully), where $|X| \ll |Y|$. Suppose the solution technique S affords an adequate solution to problem instance P_1 , in the sense that the solution is both adequate for the task at hand and S found the solution in an adequate amount of time, consuming an acceptable amount of computational resources.

S will be said to scale (or scale well) to problem instance P_2 , if S can generate a solution to P_2 , in an acceptable amount of time, while consuming an acceptable amount of computational resources. In other words, the cost associated with generating the solution to P_2 does not outweigh the utility of being able to generate the solution to P_2 , S is a “viable” solution technique for instance P_2 .

Evaluating all possible trajectories in a problem instance’s state-action space, scales exponentially with the size of the state-action space: Millidge [22]. For large state-action spaces, evaluating all possible action trajectories quickly becomes an “unviable” solution technique.

IV. PREVIOUS WORK

Naturally, the question of scaling Active Inference to larger and more complicated state-action spaces has already begun to occupy the attention of the research community, though these endeavors are still very much in their infancy. As of the time of writing this review, two distinct approaches seem to have crystallised in the literature. The most general bifurcation is between sample-based approximation methods and distribution-based message passing methods. The latter can be either exact or approximate.

The former, sample-based methods typically make use of a function approximator, such as a neural network to parameterise the distributions of interest. This is perhaps a more “traditional” approach, very much inspired by the success of such methods in scaling Reinforcement Learning to larger state-action spaces.

The latter message-passing methods take a completely different approach to the sampling paradigm. This approach represents the generative model as a “Forney-style” factor graph: Forney [9] and inference is performed via a kind of message-passing on this graph. Instead of processing samples, the factor graph approach manipulates full distributions to produce messages which are passed around the graph. This is an extremely fast and efficient method of implementing variational inference as it completely eschews the computational expense of processing samples, which is very often the bottleneck in sample-based approximations.

Both approaches have their respective advantages and disadvantages. I’ll now turn to a detailed exposition of each approach.

A. Sampling Based Approximation Methods

A more standard approach that has seen a comparatively greater deal of attention is that of using deep neural network function approximators to either parameterize the distributions of interest, or to afford an efficient means of sampling these distributions. This affords approximate inference. Indeed this “genre” of

approach has already seen great success in scaling reinforcement learning methods to larger state-action spaces: Mnih et al. [24], Mnih et al. [25], other citations needed and available.

See: Ueltzhöffer [31], Millidge [22], Millidge [23], Tschantz et al. [29] and Çatal et al. [3].

Of particular interest are: Tschantz et al. [30], Çatal et al. [3] and Mazzaglia, Verbelen, and Dhoedt [21]. The former makes use of amortized inference, in the form of neural network function approximators to parameterize the relevant distributions. Free Energy minimization is then performed with respect to the function approximators. In addition, the free energy functional is amortized over the training data. This affords several advantages. For example, the number of parameters remains constant with respect to the size of the data and inference can be achieved via a single forward pass through the network. This contrasts with the iterative approach, where the VFE must be scored for every sample, individually. The resulting algorithm was able to explore a much greater proportion of the state space in a simple stationary environment, in comparison with two Reinforcement Learning baseline agents. In addition, the agent was able to learn to control the continuous inverted pendulum task with a far greater sample efficiency than the baseline agents. Although the approach offered in Tschantz et al. [30] is promising, its analysis was restricted in every case to fully observable environments. This potentially sold the implementation short, since the partially observable domain is the more "natural" problem instance for which active inference was conceived as a solution strategy. Active Inference has a built-in drive to effect uncertainty reduction, this is not so with standard reinforcement learning, for which only ad-hoc strategies exist to afford the same sort of epistemic drive that exists in active inference. This is a salient point of departure between active inference and reinforcement learning, since both already implement strategies for realizing pragmatic value.

Pragmatic value is encoded as the sum of discounted reward across time, in the case of reinforcement learning. In the case of active inference, the drive to realize pragmatic value is afforded by the choice of action/s that realize the agent's prior preferences. The realization of prior preferences in active inference is analogous to maximizing the reward signal in reinforcement learning. However there is no analogous process for a drive to realize epistemic value, going from active inference to reinforcement learning, at least not without some aforementioned ad-hoc contrivance of the reward signal.

Lastly, the approach of Mazzaglia, Verbelen, and Dhoedt [21], implemented a contrastive method for their Active Inference agent, which significantly reduced the computational burden of learning the parameters for the generative model and planning future actions. This method performed substantially better than the usual, likelihood-based "reconstructive" means of implementing Active Inference and it was also computationally cheaper to train. Importantly, this method offered a unique way to afford increased model-robustness in the face of environmental distractors.

B. Factor Graphs and Message Passing Methods

A particularly novel approach that has enjoyed some success as of late, casts the problem of inference as a species of message passing updates on a Forney factor graph: Cox, Laar, and Vries [5], Laar and Vries [19], Vries and Friston [32] and Bagaev and Vries [1].

In this framework, the agent's generative model is constructed in such a way as to instantiate a Forney or "Normal" factor graph: Forney [9]. Free Energy minimization is then cast as a process of message passing over this factor graph. Various message passing algorithms exist, such as Belief Propagation and Variational Message Passing. This message passing scheme greatly reduces the number of terms over which it is necessary to sum, when computing the approximate marginal and posterior distributions; affording much more efficient inference and a great potential for scaling up to larger state-action spaces. Indeed, this method does not make use of any approximation by means of a sampling procedure. Since this method relies upon a particular schedule of message-passing update rules on the underlying factor graph, all functions used need to be invertible and an inference is performed via a closed form update where the prior and likelihood distributions must be conjugate. The model passes around full distributions instead of mere samples. This results in a very fast and efficient implementation - when applicable, but the issue is that it is not a completely generic method as of yet, owing to the many assumptions as to the model structure just enumerated.

V. CONCLUSION

Here I think I'll reiterate why this problem of scaling active inference is important at all and suggest potential implications for being able to make some headway in on this problem.

REFERENCES

- [1] Dmitry Bagaev and Bert de Vries. “Reactive Message Passing for Scalable Bayesian Inference”. In: *CoRR* abs/2112.13251 (2021). arXiv: 2112.13251. URL: <https://arxiv.org/abs/2112.13251>.
- [2] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. “Variational Inference: A Review for Statisticians”. In: *Journal of the American Statistical Association* 112.518 (Apr. 2017), pp. 859–877. DOI: 10.1080/01621459.2017.1285773.
- [3] Ozan Çatal et al. *Bayesian policy selection using active inference*. 2019. arXiv: 1904.08149 [cs.LG].
- [4] Paul Christiano. “Why Reinforcement Learning is Flawed”. In: *The Gradient* (2019). URL: <https://thegradient.pub/why-rl-is-flawed/>.
- [5] Marco Cox, Thijs van de Laar, and Bert de Vries. “A factor graph approach to automated design of Bayesian signal processing algorithms”. In: *International Journal of Approximate Reasoning* 104 (Jan. 2019), pp. 185–204. DOI: 10.1016/j.ijar.2018.11.002.
- [6] Lancelot Da Costa et al. “Active inference on discrete state-spaces: A synthesis”. In: *Journal of Mathematical Psychology* 99 (2020), p. 102447. ISSN: 0022-2496. DOI: <https://doi.org/10.1016/j.jmp.2020.102447>. URL: <https://www.sciencedirect.com/science/article/pii/S0022249620300857>.
- [7] Lancelot Da Costa et al. *The relationship between dynamic programming and active inference: the discrete, finite-horizon case*. Sept. 2020.
- [8] Gabriel Dulac-Arnold et al. “Challenges of real-world reinforcement learning: definitions, benchmarks and analysis”. In: *Machine Learning* 110.9 (Sept. 2021), pp. 2419–2468. ISSN: 1573-0565. DOI: 10.1007/s10994-021-05961-4. URL: <https://doi.org/10.1007/s10994-021-05961-4>.
- [9] G.D. Forney. “Codes on graphs: normal realizations”. In: *IEEE Transactions on Information Theory* 47.2 (2001), pp. 520–548. DOI: 10.1109/18.910573.
- [10] K Friston et al. “Action and behavior: a free-energy formulation”. In: *Biological cybernetics* 102.3 (2010), pp. 227–260. DOI: 10.1007/s00422-010-0364-z. URL: <https://doi.org/10.1007/s00422-010-0364-z>.
- [11] Karl Friston. “Life as we know it”. In: *Journal of The Royal Society Interface* 10.86 (2013), p. 20130475. DOI: 10.1098/rsif.2013.0475. eprint: <https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2013.0475>. URL: <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2013.0475>.
- [12] Karl Friston. *The free-energy principle: a rough guide to the brain?* 2009. URL: <https://doi.org/10.1016/j.tics.2009.04.005>.
- [13] Karl Friston. “The free-energy principle: a unified brain theory?” In: *Nature Reviews Neuroscience* 11.2 (Feb. 2010), pp. 127–138. ISSN: 1471-0048. DOI: 10.1038/nrn2787. URL: <https://doi.org/10.1038/nrn2787>.
- [14] Karl Friston, James Kilner, and Lee Harrison. “A free energy principle for the brain”. In: *Journal of Physiology-Paris* 100.1 (2006). Theoretical and Computational Neuroscience: Understanding Brain Functions, pp. 70–87. ISSN: 0928-4257. DOI: <https://doi.org/10.1016/j.jphysparis.2006.10.001>. URL: <https://www.sciencedirect.com/science/article/pii/S092842570600060X>.
- [15] Karl Friston et al. “Active inference and epistemic value”. In: *Cognitive Neuroscience* 6.4 (2015). PMID: 25689102, pp. 187–214. DOI: 10.1080/17588928.2015.1020053. eprint: <https://doi.org/10.1080/17588928.2015.1020053>. URL: <https://doi.org/10.1080/17588928.2015.1020053>.
- [16] Karl J. Friston, Jean Daunizeau, and Stefan J. Kiebel. “Reinforcement Learning or Active Inference?” In: *PLOS ONE* 4.7 (July 2009), pp. 1–13. DOI: 10.1371/journal.pone.0006421. URL: <https://doi.org/10.1371/journal.pone.0006421>.
- [17] David C. Knill and Alexandre Pouget. “The Bayesian brain: the role of uncertainty in neural coding and computation”. In: *Trends in Neurosciences* 27.12 (2004), pp. 712–719. ISSN: 0166-2236. DOI: <https://doi.org/10.1016/j.tins.2004.10.007>. URL: <https://www.sciencedirect.com/science/article/pii/S0166223604003352>.
- [18] Magnus Koudahl, Christopher L. Buckley, and Bert de Vries. “A Message Passing Perspective on Planning Under Active Inference”. In: *Active Inference*. Ed. by Christopher L. Buckley et al. Cham: Springer Nature Switzerland, 2023, pp. 319–327. ISBN: 978-3-031-28719-0.
- [19] Thijs W. van de Laar and Bert de Vries. “Simulating Active Inference Processes by Message Passing”. In: *Frontiers in Robotics and AI* 6 (2019). ISSN: 2296-9144. DOI: 10.3389/frobt.2019.00020. URL: <https://www.frontiersin.org/articles/10.3389/frobt.2019.00020>.
- [20] Dimitrije Marković et al. “An empirical evaluation of active inference in multi-armed bandits”. In: *Neural Networks* 144 (2021), pp. 229–246. ISSN: 0893-6080. DOI: <https://doi.org/10.1016/j.neunet.2021.08.018>. URL: <https://www.sciencedirect.com/science/article/pii/S0893608021003233>.

- [21] Pietro Mazzaglia, Tim Verbelen, and Bart Dhoedt. “Contrastive Active Inference”. In: *Advances in Neural Information Processing Systems*. Ed. by M. Ranzato et al. Vol. 34. Curran Associates, Inc., 2021, pp. 13870–13882. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/73c730319cf839f143bf40954448ce39-Paper.pdf.
- [22] Beren Millidge. “Applications of the free energy principle to machine learning and neuroscience”. PhD thesis. University of Edinburgh, 2021. URL: <https://era.ed.ac.uk/handle/1842/38235>.
- [23] Beren Millidge. “Deep Active Inference as Variational Policy Gradients”. In: (July 2019). URL: <https://arxiv.org/pdf/1907.03876.pdf>.
- [24] Volodymyr Mnih et al. *Asynchronous Methods for Deep Reinforcement Learning*. 2016. arXiv: 1602.01783 [cs.LG].
- [25] Volodymyr Mnih et al. *Playing Atari with Deep Reinforcement Learning*. 2013. arXiv: 1312.5602 [cs.LG].
- [26] Deepak Pathak et al. “Curiosity-driven Exploration by Self-supervised Prediction”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, June 2017, pp. 2778–2787. URL: <https://proceedings.mlr.press/v70/pathak17a.html>.
- [27] Ryan Smith, Karl J. Friston, and Christopher J. Whyte. “A step-by-step tutorial on active inference and its application to empirical data”. In: *Journal of Mathematical Psychology* 107 (2022), p. 102632. ISSN: 0022-2496. DOI: <https://doi.org/10.1016/j.jmp.2021.102632>. URL: <https://www.sciencedirect.com/science/article/pii/S0022249621000973>.
- [28] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. 2nd ed. Cambridge, Massachusetts: MIT Press, 2018.
- [29] Alexander Tschantz et al. *Reinforcement Learning through Active Inference*. 2020. arXiv: 2002.12636 [cs.LG].
- [30] Alexander Tschantz et al. “Scaling Active Inference”. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. 2020, pp. 1–8. DOI: 10.1109/IJCNN48605.2020.9207382.
- [31] Kai Ueltzhöffer. “Deep active inference”. In: *Biological Cybernetics* 112.6 (Oct. 2018), pp. 547–573. DOI: 10.1007/s00422-018-0785-7.
- [32] Bert de Vries and Karl J. Friston. “A Factor Graph Description of Deep Temporal Active Inference”. In: *Frontiers in Computational Neuroscience* 11 (2017). ISSN: 1662-5188. DOI: 10.3389/fncom.2017.00095. URL: <https://www.frontiersin.org/articles/10.3389/fncom.2017.00095>.