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

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

II. Introduction

A. *Agent Based Artificial Intelligence*

The field of Agent-based AI is the discipline concerned with the creation of autonomous systems, capable of dynamically adapting to the constraints of some environment. From the perspective of Artificial Intelligence, an "agent" simply refers to any computationally instantiated entity capable of perceiving and acting in its environment. Agents typically have a constitutive goal of maintaining their self-organization across time.

The general paradigm of interest with these methods is that of the "sensorimotor loop"

One of the most general distinctions to be made in this field is that between "model based" and "model free" methods.

1) Model Based Methods:

2) Model Free Methods:

B. *Generative Artificial Intelligence*

One of the most fundamental tasks that an agent can perform is *prediction*.

C. *Partially Observable Markov Decision Processes (POMDPs)*

III. Active Inference

A. *An Overview*

It is highly plausible that the field of adaptive, agent-based AI is perhaps, entering its Renaissance. The last few years have made history, in terms of the the capability, sophistication and operation of what these methods have been able to achieve.

Active Inference is an emerging first-principles, account of adaptive behavior. Originating from Neuroscience: Friston, Kilner, and Harrison [18], Knill and Pouget [23] and Friston et al. [14] as a corollary of the "Free Energy Principle", the theory is increasingly making inroads into Machine Learning and Artificial Intelligence: Friston, Daunizeau, and Kiebel [21] and Millidge [29].

Active inference is a highly ambitious theory, as it purports to offer a fully unified account of action, perception and learning: Friston [17]. In Active Inference, these three cognitive processes are integrated under the rubric of variational inference via the minimization 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 minimization 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 parameterize 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...

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 [8] and Dulac-Arnold et al. [12]. 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 minimization 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 maximize a reward function of states (or states and actions), Active Inference agents select actions so as to minimize 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-maximizing" 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 behavior, in addition to that of maximizing the extrinsic value of realizing their preferred observations.

Mere extrinsic, value-maximization is typically the *only* operative imperative in Reinforcement Learning approaches; Sutton and Barto [36]. It is not impossible to endow Reinforcement Learning agents with a drive toward information-gain or "curiosity". Though typically, one has to contrive some *hd hoc* manipulation of the reward signal to hand-craft an *encoding* of the epistemic imperative: Pathak et al. [34]. 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 realize prior preferences, Active Inference agents will naturally engage in exploratory behavior, in an attempt to find as-yet unknown routes to realize 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. [37]. This is an especially intriguing aspect of the theory, since the majority of real-world problems are characterized 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....

C. Active Inference: Key Concepts

1) Variational Inference: Variational inference: Blei, Kucukelbir, and McAuliffe [3] is a technique of approximate Bayesian inference, in the case that the exact inference procedure becomes intractable, typically due to the large or infinite number of states over which it is necessary to marginalize. 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 intractable, 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)...

2) 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 [16] and perhaps even of life itself: Friston [15].

D. 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: Parr and Friston [33] and Friston et al. [19]. 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.

This is largely due to the inherently exponential search space of evaluating potential actions into some horizon into the future. Sophisticated agents require the ability to plan actions in this manner, not merely to perform the optimal action at the present time. This requires the determination of a sequence of actions (a trajectory) into the future time horizon, the ability to score this trajectory and then the selection of the optimal trajectory, with respect to the cost function. Evaluating all possible trajectories in a problem instance's state-action space, scales exponentially with the size of the state-action space: Millidge [29]. This is very much the "heart of the issue" of scaling Active Inference. In an Active Inference context, the objective function in question for the case of planning future actions is the Expected Free Energy (EFE) and it is simply the expectation of the Variational Free Energy, under the approximate posterior for a given trajectory of actions into the future.

Aside from the exponential search space over trajectories into the future, another 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. Observations can easily be on the order of milliseconds, for instance. The task of minimizing 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 caseis, for the above reasons, very much an open one. It is a problem that must be satisfactorily addressed if Active Inference is to become a serious contender as a real-world method.

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 crystalized 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 parameterize 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: Mnih et al. [32].

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 [13] 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. The Laplace Approximation and Generalised Coordinates

A common approximation scheme used to simplify the calculation of the free energy is the *Laplace Approximation*. This consists in simply assuming that the optimal approximate posterior distribution is a Gaussian distribution. The sufficient statistics of this Gaussian then become parameters which can be optimised so as to minimise the VFE.

It is almost always further assumed that the approximate posterior is sharply peaked about its mean value and that the constituent functions of the hidden states are smooth. This makes the integration problem of evaluating the VFE appreciably non-zero, only at the peaks. We can then use a Taylor series expansion of the constituent functions about their mean value.

This affords a highly tractible, analytic form for the VFE but the above assumptions are fairly strong ones, many of which are highly questionable - depending on the intended use case. For instance, the assumption that the approximate posteror is strongly peaked about its mean, will generally be appropriate in proportion to the degree that the approximate posterior *already* constitutes a good aproximation to the true posterior. In the case of a greater amount of uncertainty about the fitness of the approximate posterior, this assumption is less and less appropriate. For small or simple problems, this is often not a large issue, but in the large multivariate case, the approximate posterior need not at all resemble a multivariate Gaussian.

This approach also has highly non-trivial implications for the interpretation of brain function. To the degree that Active Inference is used as a model of brain function, this is an obvious limitation of the approximation. See Buckley et al. [4]

In conjunction with the Laplace appoximation, when implementing Active Inference in continuous state-action spaces, a related means of representing beliefs about trajectories through time is to use *generalised coordinates of motion*. This is a rather straightforward procedure, whereby inferences are drawn, not only about each hidden state variable - x, say - but also regarding each successive temporal derivative of x: x', x'', x''' and so on. These temporal derivatives can then be used to construct an approximation to the actual trajectory. This again, is a simple means of representing the trajectory of the hidden states, in an analytically tractible way.

Both these approaches are perfectly viable constituent means of implementing Active Inference on their own but it is common to use both simultaneously. These approximations culminate in a method called "Generalised Filtering": Friston et al. [20] and Balaji and Friston [2]. This is a fully online Bayesian filtering scheme that uses generalized coordinates.

While these sorts of schemes afford computationally tractible means of formuating and evaluating the VFE, they do not necessarily afford the ability to scale these procedures to larger problem instances.

B. Factor Graphs and Message Passing Methods

A recent and particularly novel approach that has enjoyed great success as of late, casts the problem of inference as a task of message passing updates on a Forney factor graph: Cox, Laar, and Vries [9], Laar and Vries [25], Vries and Friston [39] and Bagaev and Vries [1].

In this framework, the agent's generative model is constructed in such a was as to instantiate a Forney or "Normal" factor graph: Forney [13] and Loeliger [26]. Free Energy minimization is then cast as a process of message passing over this factor graph.

1) Forney-Style Factor Graphs (FFGs): A Forney-style factor graph is a graphical representation of a factorised probabilistic model. In these graphs, edges represent variables and vertices represent relationships between these variables. Consider a simple generative model - included in Cox, Laar, and Vries [9]:

$$p(x_1, ..., x_5) = p_a(x_1) \cdot p_b(x_1, x_2) \cdot p_c(x_2, x_3, x_4) \cdot p_d(x_4, x_5) \tag{1}$$

Where $p.(\cdot)$ represents a probability density function.

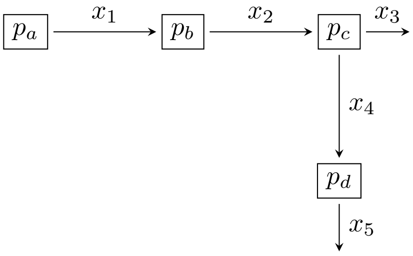

Fig. 1: An FFG representation of Equation 1. Adapted from Cox, Laar, and Vries [9]. A node (factor) connects to all edges (variables) that are arguments in that factor. For instance, p_c is connected to x_2, x_3 and x_4, since these all appear as arguments to factor p_c. Variables that only appear in one factor are represented as half-edges. Now an FFG is technically an undirected graph, however we can specify a direction for the edges to indicate the generative direction of the model.

2) Message Passing: Upon observing some particular value of one of the variables, say $x_5 = \hat{x_5}$, suppose we are interested in computing the marginal posterior probability distribution of x_2 given this observation of x_5. In an FFG formulation, observing a value for a particular variable leads to the introduction of a new factor in the model. This has the effect of "clamping" the variable to its observed value. Hence in our example, we now have:

$$p(x_1, ..., x_5) \cdot \delta(x_5 - \hat{x_5})$$

To compute the marginal posterior distribution of x_2 given an observation of $x_5 = \hat{x_5}$ we simply integrate the extended model over all variables except x_2 and renormalze:

$$p(x_2|x_5 = \hat{x}_5) \propto \int \cdots \int p(x_1, \ldots, x_5) \cdot \delta(x_5 - \hat{x}_5)\, dx_1 dx_3 dx_4 dx_5 \tag{2}$$

$$= \int p_a(x_1) p_b(x_1, x_2)\, dx_1 \cdot \int \int p_c(x_2, x_3, x_4) \cdot \left(\int p_d(x_4, x_5) \cdot \delta(x_5 - \hat{x}_5)\, dx_5 \right) dx_3 dx_4 \tag{3}$$

These nested integrals in (3) result from the substitution of the factored form of Equation 1 into (2) and then rearranging the resulting integrals via the distributive law.

The structure of the FFG can automate the the rearrangement of these integrals into a product of nested sub-integrals. The solutions to these sub-integrals can be interpreted as messages flowing over the FFG, hence this method is known as *message passing*. The massages are ordered or "scheduled" so as to only contain backward dependencies. In other words, each message can be derived from preceding messages in the schedule. Importantly, these schedules can be automatically generated by performing a depth-first search on the FFG - for instance.

Message passing is very efficient, since the computation of every message is local to each node in the FFG. Indeed, the message flowing from factor node p_b can be derived from the mere analytic expression for p_b and all messages inbound to p_b. Furthermore, if the analytic form of each incoming message is known, a pre-derived message computation rule can be used to derive the outgoing message. These rules can be easily stored in a lookup table for reuse in any model in which that specific factor-message combination is found.

The above example elaborates the fundaments of the *sum-product* message passing algorithm, however various message passing algorithms exist, such as Variational Message Passing. All message passing schemes 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 and so it avoids the computational burden associated with calculating these samples.

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 (bijective) 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, owing to the many assumptions as to the model structure just enumerated. many real-world distributions do not have conjugate prior and likelihoods and so this method cannot be applied in these cases.

C. 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 instantiates an approximate inference scheme. Indeed this "genre" of approach has already seen great success in scaling reinforcement learning methods to larger state-action spaces: Mnih et al. [31], Mnih et al. [32] and so it is a natural choice for attempting the same task in Active Inference.

Early work in this regard was condicted by: Tschantz et al. [37]. This approach uses deep neural networks to parameterise the distributions of in the generative model amortizes the inference procedure Free Energy minimization is then performed with respect to these function approximators. the use of deep neural networks to parameterise the model's distributions had been done before: Ueltzhöffer [38] and Millidge [30] but the amortization of the free energy functional over the training data was an added design choice that afforded 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. [37] is promising, its analysis was restricted in every case to fully observable environments. This potentially sold the implementaon short, since the partially observable domain is the more "natural" problem instance for which active inference was conceived as a solution strategy. Tschantz et al. [37] also embedded a reward signal into the observation space and set a prior on observing high reward outcomes. This was done simply by making the reward signal an observation specific modality.

The horizon across which Tschantz et al. [37] computed the EFE was held fixed, and dennoted as H. N sample policies are drawn from the parameterised, approximate posterior, the negative EFE is evaluated for each of these samples and then the approximate posterior is "refit" to the top M samples. Each sample is weighted in proportion to its EFE. This procedure is carried put I times, after which, the mean of the belief for the current time step is returned. This is the action selected for performance, at the current time step. Tschantz et al. [37] used $N = 1000$, $M = 100$ and $I = 10$. Hence, a distribution over policies is updated after each observation.

This approach cannot capture the exact shape of the decision sequence, though agents typicaly only need to identify the peak of the EFE landscape in order to act optimally. This is perfectly acceptable for a relatively short time horizon, but this limitation becomes prohibitive in the case of a larger one.

A more recent approach, as per: Çatal et al. [7], sucessfully learns a complex model directly from real-world pixel data. This appraoch used Deep Nerual Networks to learn the generative model from scratch, without the need to hand-craft any part of the model. Hand-crafted models are thus far the standard approach, this is obviously a time-consuming, tedious procedure. The model learned directly from observation-action sequences and constructed the agent's generative model from this raw data. In addition, the approachwas able to learn with high-dimensional pixel observations on OpenAI gym baselines. In contrast to Tschantz et al. [37], Çatal et al. [7] did not need to specify a prior on the agent's belief space. The necessity to specify such priors can be challenging, since in complicated problems, it is not necessarily obvious what these priors should be. Similarly to Tschantz et al. [37], the EFE was estimated from sampled trajectories, effectvely implementing Active Infernce as a tree search over policies.

While this appraoch is promising, the implementation uses a pre-recorded dataset of observation-action pairs to trainthe networks. Ideally, the model should be learnied in an online fashon, so that action and learning are interleaved, removing the necesity of the pre-recorded dataset. Somewhat ironicaly, one possible approach to afford this wouold be to maintain a posterior distribution over model parameters, similar to Tschantz et al. [37]. there are several other similar approaches, such as Himst and Lanillos [22], which used a Variational Autoencoder to encode the representation of continuous states and Malekzadeh and Plataniotis [27], which proposed a hybrind Active Inference and Reinforcement Learning objective called "unified inference". In the main, all these sampling approaches use deep neural networks to finesse the issue of estimating the posterior densities at issue, and most crucially the EFE itself.

Lastly, the approach of Mazzaglia, Verbelen, and Dhoedt [28], 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.

the primary disadvantage of these methods is the computational resources involved in the sampling procedure itself. it is an inherent limitation of these methods that they must cumpute a large nunber of samples in order to approximate the approximate posterior. The central question at issue in this formulaton of the problem is "how does one restrict attention to the best trajectories into the future, given the exponential search space of trajectories into the future?"

V. Gaps in The Literature

Though there has been much focus on the implementation of active inference methods for small, discrete state-action spaces: Millidge [29], Da Costa et al. [10], Smith, Friston, and Whyte [35], Da Costa et al. [11] and Friston et al. [19]. 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 [24] and Çatal et al. [6].

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 $|X| << |Y|$. 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 is 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 [29]. For large state-action spaces, evaluating all possible action trajectories quickly becomes an "unviable" solution technique.

VI. Discussion

In this section, I'll aim to settle on a particular "gap" as identified above and to justify my choice in this regard.

VII. 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.

Appendix

Remaining things that will some elaboration, or at least a definition? I'm not sure this appendix will be necessary in the end.

1) Bayesian Inference
2) Policy (Reinforcement Learning vs Active Inference framing)
3) Amortized Inference

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] Bhashyam Balaji and Karl Friston. “Bayesian state estimation using generalized coordinates”. In: *Signal Processing, Sensor Fusion, and Target Recognition XX*. Ed. by Ivan Kadar. Vol. 8050. International Society for Optics and Photonics. SPIE, 2011, 80501Y. DOI: 10.1117/12.883513. URL: https://doi.org/10.1117/12.883513.

[3] 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.

[4] Christopher L. Buckley et al. “The free energy principle for action and perception: A mathematical review”. In: *Journal of Mathematical Psychology* 81 (2017), pp. 55–79. ISSN: 0022-2496. DOI: https://doi.org/10.1016/j.jmp.2017.09.004. URL: https://www.sciencedirect.com/science/article/pii/S0022249617300962.

[5] Ozan Çatal et al. *Bayesian policy selection using active inference*. 2019. arXiv: 1904.08149 [cs.LG].

[6] Ozan Çatal et al. *Bayesian policy selection using active inference*. 2019. arXiv: 1904.08149 [cs.LG].

[7] Ozan Çatal et al. “Learning Generative State Space Models for Active Inference”. In: *Frontiers in Computational Neuroscience* 14 (2020). ISSN: 1662-5188. DOI: 10.3389/fncom.2020.574372. URL: https://www.frontiersin.org/articles/10.3389/fncom.2020.574372.

[8] Paul Christiano. “Why Reinforcement Learning is Flawed”. In: *The Gradient* (2019). URL: https://thegradient.pub/why-rl-is-flawed/.

[9] 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.

[10] 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.

[11] Lancelot Da Costa et al. *The relationship between dynamic programming and active inference: the discrete, finite-horizon case*. Sept. 2020.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] Karl Friston. *The free-energy principle: a rough guide to the brain?* 2009. URL: https://doi.org/10.1016/j.tics.2009.04.005.

[17] 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.

[18] 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.

[19] 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.

[20] Karl Friston et al. “Generalised Filtering”. In: *Mathematical Problems in Engineering* 2010 (June 2010), p. 621670. ISSN: 1024-123X. DOI: 10.1155/2010/621670. URL: https://doi.org/10.1155/2010/621670.

[21] 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.

[22] Otto van der Himst and Pablo Lanillos. “Deep Active Inference for Partially Observable MDPs”. In: *Active Inference*. Ed. by Tim Verbelen et al. Cham: Springer International Publishing, 2020, pp. 61–71. ISBN: 978-3-030-64919-7.

[23] 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.

[24] 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.

[25] 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.

[26] H.-A. Loeliger. “An introduction to factor graphs”. In: *IEEE Signal Processing Magazine* 21.1 (2004), pp. 28–41. DOI: 10.1109/MSP.2004.1267047.

[27] Parvin Malekzadeh and Konstantinos N. Plataniotis. *Combining information-seeking exploration and reward maximization: Unified inference on continuous state and action spaces under partial observability*. 2022. arXiv: 2212.07946 [cs.LG].

[28] 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.

[29] 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.

[30] Beren Millidge. “Deep Active Inference as Variational Policy Gradients”. In: (July 2019). URL: https://arxiv.org/pdf/1907.03876.pdf.

[31] Volodymyr Mnih et al. *Asynchronous Methods for Deep Reinforcement Learning*. 2016. arXiv: 1602.01783 [cs.LG].

[32] Volodymyr Mnih et al. *Playing Atari with Deep Reinforcement Learning*. 2013. arXiv: 1312.5602 [cs.LG].

[33] Thomas Parr and Karl J. Friston. “Uncertainty, epistemics and active inference”. In: *Journal of The Royal Society Interface* 14.136 (2017), p. 20170376. DOI: 10.1098/rsif.2017.0376. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2017.0376. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2017.0376.

[34] 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.

[35] 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.

[36] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. 2nd ed. Cambridge, Massachusetts: MIT Press, 2018.

[37] 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.

[38] Kai Ueltzhöffer. “Deep active inference”. In: *Biological Cybernetics* 112.6 (Oct. 2018), pp. 547–573. DOI: 10.1007/s00422-018-0785-7.

[39] 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.