Review

The Free Energy Principle for Perception and Action: A Deep Learning Perspective

Pietro Mazzaglia * [ID]**, Tim Verbelen** [ID]**, Ozan Çatal and Bart Dhoedt**

IDLab, Ghent University, 9052 Gent, Belgium; tim.verbelen@ugent.be (T.V.); ozan.catal@ugent.be (O.Ç.); bart.dhoedt@ugent.be (B.D.)
* Correspondence: pietro.mazzaglia@ugent.be

Abstract: The free energy principle, and its corollary active inference, constitute a bio-inspired theory that assumes biological agents act to remain in a restricted set of preferred states of the world, i.e., they minimize their free energy. Under this principle, biological agents learn a generative model of the world and plan actions in the future that will maintain the agent in an homeostatic state that satisfies its preferences. This framework lends itself to being realized in silico, as it comprehends important aspects that make it computationally affordable, such as variational inference and amortized planning. In this work, we investigate the tool of deep learning to design and realize artificial agents based on active inference, presenting a deep-learning oriented presentation of the free energy principle, surveying works that are relevant in both machine learning and active inference areas, and discussing the design choices that are involved in the implementation process. This manuscript probes newer perspectives for the active inference framework, grounding its theoretical aspects into more pragmatic affairs, offering a practical guide to active inference newcomers and a starting point for deep learning practitioners that would like to investigate implementations of the free energy principle.

Keywords: free energy principle; active inference; deep learning; machine learning

1. Introduction

Understanding the processes that sentient and reasoning beings play out mentally in order to perceive the world they live and act in is as compelling as it is complex. The free energy principle hypothesizes all brain processes can be understood as subserving one unicum imperative: the minimization of free energy [1,2]. This principle, and its corollary active inference, assumes that agents act to contrast forces from the environment that obstruct them from remaining in a restricted set of preferred states of the world. Under this assumption, biological agents develop a variety of skills, such as perception, action, planning, and learning, that agents continuously adapt along their lives.

Active inference and the free energy principle have been employed to explain and simulate several complex processes across different disciplines. In psychology, they have been used to ground a computational account of neuropsychological syndromes [3] and to develop emotional recognition devices, which allow resolving uncertainty about emotional states by interaction and learning [4]. In economics, the free energy principle has been exploited to reformulate the agents' optimization process in terms of their beliefs [5]. Variational approaches have been employed to explain niche construction, based on the free energy principle [6,7]. Active inference has been used to model smooth and saccadic eye movements [8,9] and to conceptualize attention [10], salience, and memory [11]. In the context of scene construction, active inference provides an explanation of how agents infer a higher-order visual pattern of a scene through visual foraging [12,13].

Under some sets of assumptions [14,15], the free energy principle can also be used to explain how all biological organisms and processes that subserve perception and action naturally emerge and are continuously adjusted through a natural model selection process. On shorter (somatic) timescales, minimizing free energy leads to the development of

the single organisms' brains, giving rise to learning and memory functions. On longer (evolutionary) time scales, free energy minimization fosters the evolution process of the species [16,17]. Establishing a statistical boundary for the Earth's climate system, this can also be interpreted as a self-producing system that is performing active inference [18]. Systems that are capable of self-production, i.e., continuously generating and maintaining themselves by creating their own parts, are called autopoietic [19]. It can be shown that the process of autopoiesis minimizes free energy, given that for an organism to maintain a model of itself and its environment it must minimally self-produce the components required to carry out the process of preserving its generative model [20].

In active inference, the agent minimizes a variational free energy objective with respect to past experience that causes it to learn a generative model of the world, which allows predicting what will happen in the future in order to avoid surprising states. Variational inference [21], from which originates the variational free energy functional, makes possible to cast the process of perception as an optimization process, which subsumes a set of choices in terms of modeling and learning, such as the choice of the state distribution or the way the model copes with uncertainty. The generative world model is then used in the future to plan actions that will maintain the agent in a homeostatic state that satisfies the agent preferences. The agent operates an (amortized) Bayesian selection of actions that will have the least surprise with respect to its preferred state. This decision-making process takes into consideration several aspects of learning, such as epistemics, habit learning, and preference learning.

Recent developments in deep learning have opened new frontiers for studying and experimenting with different perception and behavioral theories; enabling the practical analyses of artificial implementations, either in simulations or in real environments. One popular example in this regard is reinforcement learning (RL) [22], a theory that links the dopamine signals in the brain to reward signals that can be used to reinforce correct behaviors [23,24], and describes how intelligent behaviors can be learned through reward maximization [25]. Combining RL with deep learning models for function estimation [26] has led to several empirical successes, allowing the training of artificial agents to play video games [27,28], master board games [29], or execute robotic tasks [30]. Similarly, deep learning techniques are also starting to arise in the context of active inference [31–33].

This work aims to survey the current state-of-the-art of deep learning models for active inference. At the same time, we want to provide a reference and a starting point for machine learning practitioners to become acquainted with active inference, drawing parallels between active inference and recent advances in RL. Other reviews of active inference have previously been presented for expressing generative models in continuous spaces [34] or discrete spaces [35]; however, the context, the practices presented, and the scope of our work strongly differ from theirs, in that we focus specifically on deep-learning-based techniques to scale active inference to large continuous state and action spaces or high-dimensional settings, such as robotics and visual control. Concurrently to our work, a review on active inference for robotics has been released [36], including references to deep learning methods for active inference. Our work differs from theirs in that they specifically focus on methods that enable applying active inference in robotics, while we more broadly discuss active inference techniques that can be employed to develop active inference artificial agents, explaining in details how each component could be implemented, highlighting the challenges to overcome, and establishing connections between methods in different machine learning areas.

A non-exhaustive summary of the aspects that are considered in the active inference framework for learning perception and action is presented in Figure 1. We divide into model learning on the one hand, and action selection on the other hand. The first comprises minimizing variational free energy for learning generative models based on past experiences, modeling belief states and priors, learning state representations and uncertainty, whereas the latter deals with selecting actions in the future, which trade-off epistemic foraging and realizing preferences, either by planning or learning habits. Two recurrent

patterns in active inference are variational inference and the amortization of Bayesian selection. Variational inference allows casting inference as an optimization problem, i.e., finding the distribution closest to the actual one. Amortization allows a faster computation of the inference process, by reusing previous computations [37]. In active inference, amortized inference is applied for the choice of variational parameters, in the model learning process, and for the formation of habits, in the action selection process. Combining the two techniques strongly reduces the computational requirements of active inference and makes the framework promising for implementation in silico; however, without adequate models, it is unfeasible to scale active inference in complex scenarios, with continuous and/or high-dimensional state/action spaces.
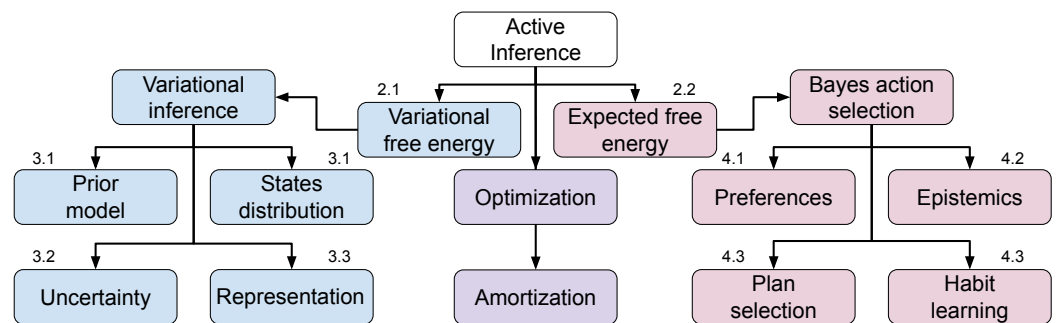

Figure 1. The free energy functional minimized by active inference takes two forms: variational free energy, with respect to past experience, and expected free energy, for selecting future behaviors. For each of the two, an (amortized) Bayesian optimization scheme is followed that needs to consider several aspects, as summarized in the diagram. The numbering indicates the section of the paper discussing each aspect.

In deep learning, generative models have been widely studied, obtaining outstanding results in several domains, such as image generation [38–40], text prediction [41–43], and video modeling [44–47]. In particular, temporal deep generative models that allow predicting the dynamics of a system, i.e., the environment or world, have been studied for control [48–50], curiosity and exploration [51–53], and anomaly detection [54]. Several of these models have been used in settings that are similar to the active inference one, and some of them even share some similarities with the active inference objective of minimizing variational free energy. As for action selection, several works that use deep learning have improved upon more classical methods (e.g., $\alpha - \beta$ pruning, A*, beam search) allowing to search much larger state and action spaces. Some examples are dynamic programming-related techniques in RL [26], evolutionary strategies [55], and Monte Carlo Tree Search (MCTS) [29,56]. These methods all combine more general and classical planning strategies with deep learning for function estimation, enabling the scaling of behavior learning and action selection for complex environments.

The remainder of this work is organized as follows: in Section 2, we present the free energy principle and explain what minimizing free energy entails in terms of perception and action; in Section 3, we establish the connection between model learning, according to active inference, and deep generative models, analyzing the different ingredients involved. In Section 4, we discuss the implementation and design choices that underlie the action selection process, relating existing works on habit learning, exploration and model-based control. Finally, we conclude our discussion, remarking the work conducted until this moment, addressing the implications of learning with deep learning, and proposing some future perspectives.

2. The Free Energy Principle and Active Inference

The free energy principle is at the core of the active inference framework, as it conceptualizes the development of embodied perception as the result of minimizing a free energy

objective. As we show in this section, the free energy is a function of the agent's beliefs about the environment, representing a (variational) upper bound on surprisal from sensorial stimuli. This entails that reducing free energy additionally reduces the agent's model surprise, restricting its existence to a limited set of craved beliefs. The free energy principle originated from the work of von Helmholtz on 'unconscious inference' [57], postulating that humans inevitably perform inference in order to perform perception. This implies that the human perceptual system continuously adjusts beliefs about the hidden states of the world in an unconscious way. The variational formulation of the free energy [1,58], along with the introduction of actions as part of the inference process, expanded the original free energy principle leading to the development of active inference.

In Figure 2, we illustrate the interplay between the main factors that determine the embodied perception process as described in active inference. At any time, the environment is in a certain state η, which is external to the agent and not directly observable. The agent interacts with the environment in two ways: either through (passive) sensorial perception, which is characterized by the observation of sensorial states o, or by actions, which can be cast as a set of active states a that the agent imposes on the environment. According to the free energy principle, in order to minimize free energy, the agent learns an internal model of potential states of the environment. Crucially, these internal states do not need to be isomorphic to the external ones, as their purpose is explaining sensorial states in accordance with active states, rather than replicating the exact dynamics of the environment Isomorphism, in this context, refers to considering a structure-preserving mapping of the state space. According to active inference, internal and environment states are not necessarily equal and the way the internal states are organized may even differ from agent to agent, despite having to deal with similar concepts/observations/sensory states. From a biological perspective, this finds evidence in the fact that different living systems have developed different organs/tissues along their evolutionary process [16]. The role of the internal state representation is, in fact, to provide the sufficient statistics that allow a 'best guess' about the causes of the agent's observations and the selection of adaptive action policies [59].
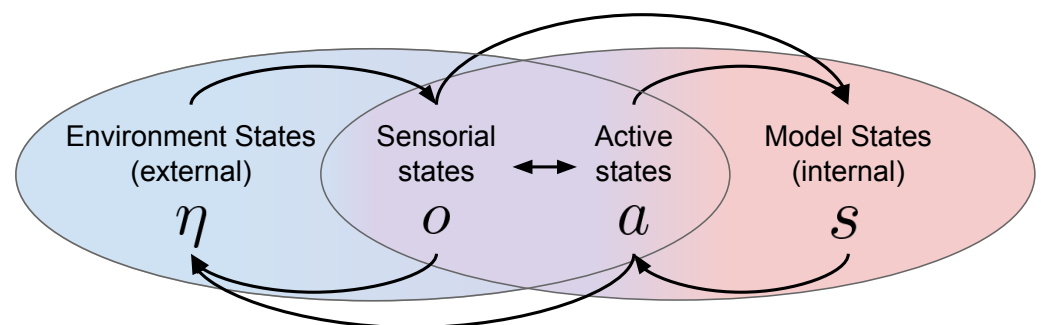

Figure 2. The external environment states η are the hidden causes of sensorial states o (observations). The environment attempts to represents such hidden causes through its internal model states s. Crucially, internal states may or may not correspond to external states, which means that hidden causes in the brain do not need to be represented in the same way as in the environment. Active states a (actions), which are developed according to internal states, allow the agent to condition the environment states.

As a consequence of minimizing free energy, the agent possesses beliefs about the process generating outcomes, but also about action policies that lead to generating those outcomes [60]. This corresponds to a probabilistic model of how sensations are caused and how states should be actively sampled to drive the environment's generation of sensory data. Because of these assumptions, the concept of 'reward' in active inference is very different from rewards in RL, as rewards are not signals used to attract trajectories, but rather sensory states that the agents aims to frequently visit in order to minimize its free

energy [61]. From an engineering perceptive, this difference is reflected in the fact that rewards in RL are part of the environment and, thus, each environment should provide its unique reward signal, while in active inference 'rewards' are intrinsic to the agent, which would pursue its preferences in any environment, developing a set of most frequently visited states.

In the remainder of this section, we discuss how the agent's probabilistic model for perception and action is learned by minimizing free energy, providing a mathematical synthesis. We consider the environment as a partially observable Markov decision process (POMDP), represented in Figure 3. Using subscripts to specify the discrete time steps, we indicate the observation or outcome at time t with o_t. To indicate sequences that span over an undefined number of time steps, we use the superscript $\sim$, i.e., for outcomes $\tilde{o} = \{o_1, o_2, \ldots, o_t\}$. The succession of states $\tilde{s} = \{s_1, s_2, \ldots, s_t\}$ is influenced by sequences of actions, or policies, that we indicate with $\pi = \tilde{a} = \{a_1, a_2, \ldots, a_t\}$. Parameterization of the state-outcome likelihood mapping is indicated with θ. A precision parameter ζ influences action selection working as an inverse temperature over policies.

Generative model

$$p(\tilde{o}, \tilde{s}, \pi, \theta, \zeta) = p(\pi|\zeta)p(\theta)p(\zeta)\prod_{t=1}^{T} p(o_t|s_t, \theta)p(s_t|s_{t-1}, \pi)$$

Factors

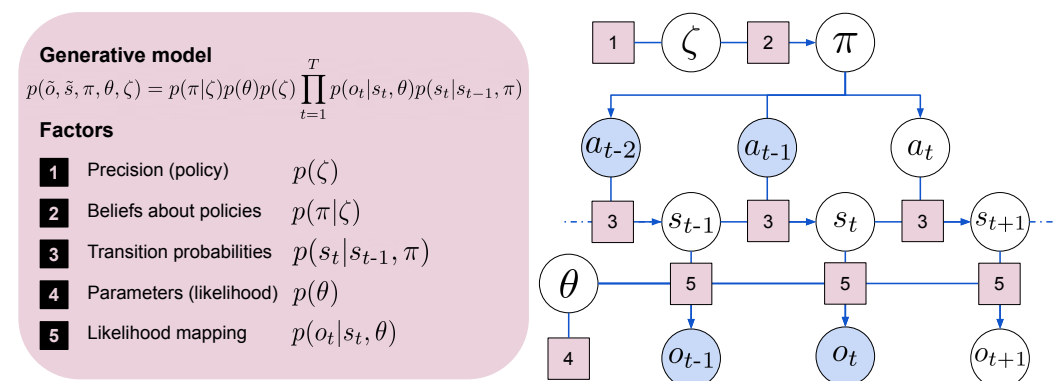|		
1	Precision (policy)	$p(\zeta)$
2	Beliefs about policies	$p(\pi|\zeta)$
3	Transition probabilities	$p(s_t|s_{t-1}, \pi)$
4	Parameters (likelihood)	$p(\theta)$
5	Likelihood mapping	$p(o_t|s_t, \theta)$

Figure 3. The diagram illustrates the interplay between the different factors that compose the graphical model. (1) Policy precision; (2) beliefs about policies; (3) transition probabilities, also known as dynamics; (4) parameters of the likelihood mapping; (5) likelihood model.

The section is divided in two parts: the first explains how the internal model is learned with respect to past experience, minimizing a variational free energy functional that explains the dynamics of the environment's outcomes, given a sequence of actions. In the second part, we discuss the minimization of expected free energy with respect to the future, when actions are selected to reduce surprise with respect to the agent's preferred outcomes. Importantly, our treatment refers to a discrete-time instantiation of active inference. For a discussion on continuous time, the reader may refer to [34].

2.1. Variational Free Energy

In order to minimize the negative log evidence (also known as "surprisal" or "surprise") of observations from the external world, the agent exploits its past sensory experiences to learn a generative model of the environment using variational inference. Under the free energy principle [62,63], an upper bound on surprise is established as:

$$\underbrace{-\log p(\tilde{o})}_{\text{surprise}} = -\log \mathbb{E}_{q(\tilde{s}, \pi, \theta, \zeta)}\left[\frac{p(\tilde{o}, \tilde{s}, \pi, \theta, \zeta)}{q(\tilde{s}, \pi, \theta, \zeta)}\right] \leq \underbrace{\mathbb{E}_{q(\tilde{s}, \pi, \theta, \zeta)}\left[-\log \frac{p(\tilde{o}, \tilde{s}, \pi, \theta, \zeta)}{q(\tilde{s}, \pi, \theta, \zeta)}\right]}_{\text{variational free energy } \mathcal{F}} \quad (1)$$

where, left to right, the following operations are performed: (i) the surprise over the sequence of observations is marginalized (i.e., summed over / integrated) with respect to the other factors of the generative model, which are the sequence of states, the policy of actions, the model parameters, and the policy precision parameter, (ii) a variational

posterior over states and policies $q(\tilde{s}, \pi, \theta, \zeta)$ is introduced for variational inference [21], (iii) Jensen's inequality is applied to push the logarithm inside the expectation operator.

The variational free energy $\mathcal{F}$ can be reformulated as:

$$
\begin{aligned}
\mathcal{F} &= \mathbb{E}_{q(\pi,\theta)}[\mathcal{F}_{\pi,\theta}] + D_{\text{KL}}[q(\pi,\zeta)\|p(\pi,\zeta)] + D_{\text{KL}}[q(\theta)\|p(\theta)], \\
\mathcal{F}_{\pi,\theta} &= \mathbb{E}_{q(\tilde{s}|\pi,\theta)}[\log q(\tilde{s}|\pi,\theta) - \log p(\tilde{o}, \tilde{s}|\pi,\theta)].
\end{aligned}
\tag{2}
$$

When minimizing the variational free energy with respect to the past, the second equation is generally adopted as the two Kullback–Leibler (KL) divergence terms in the first equation can be neglected. The former KL divergence, referring to the policies and the precision parameter, can be overlooked because policies in the past are observed. The latter KL divergence, referring to the model parameters, can be later considered to work on top of the model through regularization techniques, e.g., similarly to sleep, where redundant synaptic parameters are eliminated to minimize model complexity [64].

Omitting conditioning on π and θ for brevity, allows the expression of the free energy in its two typical forms:

$$
\begin{aligned}
\mathcal{F}_{\pi,\theta} &= \underbrace{D_{\text{KL}}[q(\tilde{s})\|p(\tilde{s})]}_{\text{complexity}} - \underbrace{\mathbb{E}_{q(\tilde{s})}[\log p(\tilde{o}|\tilde{s})]}_{\text{accuracy}} \\
&= \underbrace{D_{\text{KL}}[q(\tilde{s})\|p(\tilde{s}|\tilde{o})]}_{\text{approx vs true posterior}} - \underbrace{\log p(\tilde{o})}_{\text{log evidence}}
\end{aligned}
\tag{3}
$$

On the one hand, minimizing free energy implies maximizing the accuracy of a likelihood model $p(\tilde{o}|\tilde{s})$ while reducing the complexity of the posterior distribution. On the other hand, it implies optimizing the variational evidence bound, reminding that the KL divergence is always non-negative, namely $D_{\text{KL}}[\cdot\|\cdot] \geq 0$ for any distribution. The KL divergence is zero when the agent's model perfectly matches the environment dynamics, corresponding to the optimal scenario.

Though there is no expectation operator over past experiences (both for brevity and to comply with the typical way of expressing this functional), it should be clear that the agent minimizes variational free with respect to known sequences of past observations and policies from the environment. As we discuss in the following paragraphs, this is fundamental and is the main aspect that differentiates the variational free energy, computed with respect to past states of the environment, from the expected free energy, which considers future states and unobserved data.

2.2. Expected Free Energy

In order to minimize free energy in the future, the agent should adapt its behavior, i.e. the active states, to confine its existence within a limited set of states. These states correspond to the so called *preferences* of the agent, or preferred observations/outcomes/states, depending on the context. The objective of the agent is to exploit its knowledge about the environment, available through the internal model, to perpetually fulfill preferred perceptions in the upcoming future. While minimizing free energy about future sequences, the agent imagines how the future would look like, given a certain sequence of actions or policy. This is reflected by an expectation over future states and observations generated by the model, which exploits both the internal states model and the likelihood mapping from the environment as $\tilde{q} = q(\tilde{o}, \tilde{s}, \theta|\pi) = p(\tilde{o}|\tilde{s}, \theta)q(\tilde{s}|\pi)q(\theta)$.

Starting from this assumption, the expected free energy $\mathcal{G}$ can be expressed as:

$$
\begin{aligned}
\mathcal{G}_\pi &= \mathbb{E}_{\tilde{q}}[\log q(\tilde{s}, \theta|\pi) - \log p(\tilde{o}, \theta, \tilde{s}|\pi)] \\
&= -\underbrace{\mathbb{E}_{\tilde{q}}[\log p(\tilde{s}|\tilde{o}, \pi) - \log q(\tilde{s}|\pi)]}_{\text{information gain (hidden states)}} - \underbrace{\mathbb{E}_{\tilde{q}}[\log p(\theta|\tilde{s}, \tilde{o}, \pi) - \log q(\theta)]}_{\text{information gain (parameters)}} - \underbrace{\mathbb{E}_{\tilde{q}}[\log p(\tilde{o})]}_{\text{extrinsic value}}
\end{aligned}
\tag{4}
$$

Hence, minimizing the expected free energy implies that the agent: (i) maximizes epistemic value, i.e., mutual information between hidden states and sensory data, (ii) maximizes parameter information gain, i.e., mutual information between parameters and states, and (iii) maximizes extrinsic value, i.e., the log likelihood of outcomes under a preferred, prior distribution $\log p(\tilde{o})$, or rewards.

Again, omitting conditioning on the policy for brevity, it is possible to rewrite the expected free energy as a minimization of risk and ambiguity:

$$\mathcal{G}_\pi = \underbrace{\mathbb{E}_{\tilde{q}}[H[p(\tilde{o}|\tilde{s})]}_{\text{ambiguity}} + \underbrace{D_{\text{KL}}[q(\tilde{s})\|p(\tilde{s})]}_{\text{state control}} - \underbrace{\mathbb{E}_{\tilde{q}}[\log p(\theta|\tilde{s},\tilde{o}) - \log q(\theta)]}_{\text{information gain (parameters)}} \tag{5}$$

$$\approx \underbrace{\mathbb{E}_{\tilde{q}}[H[p(\tilde{o}|\tilde{s})]}_{\text{ambiguity}} + \underbrace{D_{\text{KL}}[q(\tilde{o})\|p(\tilde{o})]}_{\text{risk}} - \underbrace{\mathbb{E}_{\tilde{q}}[\log p(\theta|\tilde{s},\tilde{o}) - \log q(\theta)]}_{\text{information gain (parameters)}} \tag{6}$$

Here, we assume the bound is tight and the approximate posterior is a good approximation of the true posterior to express risk in terms of outcomes; however, the agent may also express its homeostatic preferences in terms of internal states rather than on its sensorial perceptions, and formulate the expected free energy in terms of state control.

Despite the variety of objectives that can be considered for the expected free energy, by either changing some assumptions or reordering its factors, the imperative of the functional stays the same. The goal of the agent is to restrict itself to its preferred set of states/sensorial perceptions, while minimizing the ambiguity of its internal model. Maximizing epistemic value and/or parameter information gain, as for Equation (4), also implicitly adheres to this hypothesis, as finding informative states in the environment will minimize the uncertainty of the model in the future [63,65].

3. Variational World Models

In active inference, the objective of minimizing surprise of the internal model with respect to sensory inputs induces a continuous model learning process that happens inside the brain. This assumes a predictive coding interpretation of the brain, where the internal model is used to generate predictions of sensory inputs that are compared to the actual sensory inputs. The internal model attempts to explain the dynamics of the world and thus, as performed in related work [50,66], we also refer to it as the 'world model'. The reaction of the internal model, tending to minimize free energy with respect to the sensory inputs, accounts for perception of the agent, which learns to predict the sensory inputs and the causal structure of their generation.

In machine learning, such a learning process, which requires no human supervision or labelling, is generally referred to as self-supervised or unsupervised learning. In contrast to biological agents, deep learning systems typically use a batch learning scheme, where a dataset of past trajectories is collected, and models are parameterized as deep neural networks that are optimized by training on batches of data sampled from this dataset. Such a dataset, indicated by $\mathcal{D}_{\text{env}}$, can be seen as an ordered set of triplets containing an environment observation, the agent's action, and the following observation (caused by the action), namely the triplet (o_t, a_t, o_{t+1}). Training the model is then typically alternated with collecting new data by interacting with the environment, using the model for planning [49], or using an amortized (habitual) policy [33,67]. In practice, one can also train a model upfront using a dataset of collected trajectories from a random agent [68] or an expert [32]. The latter is especially relevant in contexts where collecting experience online with the agent can be expensive or unsafe [69].

The free energy loss to minimize for one time step, i.e., $\pi = a_t$, can then be written under the expectations of the data (observations and actions) from the replay buffer:

$$\mathcal{F}_{a_t} = \mathbb{E}_{\mathcal{D}_{\text{env}}} \left[\mathrm{D}_{\mathrm{KL}}[q(s_{t+1}|s_t, a_t) \| p(s_{t+1}|s_t, a_t)] - \mathbb{E}_{q(s_{t+1}|s_t, a_t)}[\log p(o_{t+1}|s_{t+1})] \right] \quad (7)$$

Given the above loss, three distributions or models need to be instantiated: (i) a likelihood model $p(o_{t+1}|s_{t+1})$ that allows generating (also known as reconstructing) sensory data from the model's internal states, (ii) a prior model $p(s_{t+1}|s_t, a_t)$, which encodes information about the transition probabilities of the dynamics of the internal states, (iii) a posterior model $q(s_{t+1}|s_t, a_t)$, which is chosen to minimize the upper bound on surprise, according to variational inference [21]. In machine learning, this formulation is better known as the (negative) evidence lower bound (ELBO), it is the same as the loss used to train variational autoencoders (VAE) [70,71]) and it is shown to optimize a variational information bottleneck, trading off between the accuracy and the complexity of summarizing the sensory information in the internal state representation [72,73]. As it is assumed that external states are not observed by the agent and that the dynamics of the environment is unknown, there is no way to ensure that the representation learned is the same as external one; however, the relationship with information compression techniques ensures that minimizing free energy entails optimizing the information contained in the internal states about the sensory states [62,74].

Theoretically, deep learning models can be employed to approximate any function with an arbitrary degree of accuracy [75], which in our case means predicting both the internal states and sensory data distributions with an arbitrary degree of accuracy. In practice, though, obtaining an highly accurate model is difficult and, while neural models can find useful approximations of these models, if one of them is well-known in advance, directly using it and adapting the other models accordingly could lead to more satisfactory results. For instance, if the actual likelihood model is known for a certain state space, the state space of the prior and posterior models can be adapted accordingly. This is the case for differentiable simulators, where the environment's observations are the result of a differentiable generation process that can be integrated in the world model [76,77]. Or again, if the dynamics of the environment is known, it is possible to use that as a prior, forcing the internal states of the model and the external states of the environment to have the same structure. When the environment is represented as a POMDP, having complete knowledge of the dynamics is the only case that ensures an optimal behavior can be found [22,78]; however, even knowing the dynamics in advance, solving the POMDP problem remains computationally intractable. Function approximations techniques and procedures to construct an improved state representation, as the model learning approaches we describe here, are often used to find nearly optimal policies in more computationally efficient ways [79–81].

In a general scenario, all distributions are unknown and must be either learned by the agent or assumed having a certain form, according to some design choices. There are indeed numerous options to consider when instantiating the different models, and some of them are important to carry out a stable optimization and/or a well-thought amortization of the Bayesian inference process. Other design choices consider different aspects of the generative model (Figure 3), such as the parameters of the likelihood mapping θ and the sensitivity of the model, which are certainly relevant but that can often be assumed fixed and neglected.

3.1. Models

To instantiate the deep neural networks for the agent's generative model, first, it is important to consider the nature of the variables involved. For the hidden states, active inference assumes a probabilistic model. Unless the environment state space nature is known, the internal state distribution may have no predefined structure, and neural networks can be trained to output distributions of different kinds; however, in order to compute the expectations, as in Equation (7), it is important that the sampling process of the distribution is differentiable, as the objective needs to be backpropagated through the model for

computing the gradients that update the model [82]. As the sampling process is generally non-differentiable, the gradients of the samples should be estimated with ad hoc techniques. Some widely known examples are the reparametrization trick for Gaussian distributions [70,71], the straight-through gradient method [83], the likelihood-ratios method [84], also known as REINFORCE gradients [85], for Bernoulli and categorical variables.

In the active inference literature, multivariate Gaussian (also known as normal) distributions with a diagonal covariance matrix have been largely adopted since the initial works on VAEs [32,33,68,86]. Similarly, numerous latent state space models have adopted a Gaussian structure of their latent space [48,49,87,88], but also more complex mixture models have been proposed [50]. For Bernoulli and categorical distributions, there has been both work on general-purpose generative models, such as discrete VAE [38,89], on latent dynamics models for planning [66,90], and recently they have also been used in an active inference setting [91]. Some other alternatives to the above methods, which have yet to be explored for training world models, are: piecewise distributions [92], Markov chains [93], and normalizing flows [94].

Posterior Model. The choice of the distribution is particularly important for the posterior model, which is the variational distribution. In theory, one could search for an optimal distribution of parameters for each of the environment's transitions/observations, though that is a slow and difficult process. To speed up training, but also guaranteeing a legitimate choice of the posterior, it is possible to amortize the selection of the posterior parameters, as presented in the original VAE work [70,71]. The autoencoding amortization scheme employs the observation corresponding to a certain state to infer the parameters of the variational distribution, $q(s_{t+1}|s_t, a_t, o_{t+1})$. This allows optimizing the parameters of the posterior to compress information optimally, as the posterior has access to the observation that the likelihood model wants to generate. In VAE terms the posterior model is typically called the "encoder", whereas the likelihood model is dubbed the "decoder".

The choice of the encoder architecture, which allows the flow of information from observations to the posterior, depends on the environment. For instance, for two-dimensional matrices of data, such as images, convolutional neural networks (CNN) [95], or other architectures for computer vision such as vision transformers [96] are common choices. Other potentially useful models can be multilayer perceptrons (MLP) for vector-structured data [97] or graph neural networks, for graph-structured data [98]. Similarly, the choice of the likelihood model depends on the format of the observation data, e.g., a transposed CNN can be useful in the case of visual data.

Prior Model. The prior model can either be known or learned. To learn the prior model, one can adopt a recurrent neural network architecture, i.e., using memory cells such as long short term memories (LSTM) [99] or gated recurrent units (GRU) [100]. In other cases, the environment dynamics is known upfront, or assumptions about the prior can be made, such as assuming that the prior is a uniform probability distribution. For instance, an isotropic multivariate Gaussian $\mathcal{N}(0, I)$, with zero mean and an identity covariance matrix I, can be employed as a fixed prior, as performed in standard VAE architectures [70,71]. Alternatively, assuming the laws that govern the dynamics are known (e.g., physics laws), the environment's physics could be exploited as a strong prior [101]. In a similar fashion, in [102], the authors used the internal state of the robot to force a known prior structure on the posterior. Finally, the prior could also be ignored/considered constant, treating the model as an entropy-regularized autoencoder [103].

3.2. Uncertainty

In the active inference perception model, precision, or sensitivity [104,105], is generally associated to the uncertainty of the transitions between hidden states of the prior (beliefs precision) or the mapping from hidden states to outcomes of the likelihood (sensory precision), and can be expressed as the inverse variance of the distribution [106]. In active inference implementations, precision has been employed as a form of attention, to decide on which transitions the model should focus on learning from [33], though this aspect has

been generally less studied in the literature. Similar mechanisms of precision have been employed for VAE models to control disentanglement of the latent state space [107] or posterior collapse [108].

Another source of uncertainty in the model arises from uncertainty in the parameters. In the deep learning community, uncertainty about model parameters has been studied employing Bayesian neural networks [109], dropout [110], or ensembles [111], and used in RL to study the exploration problem [53,112,113]. In active inference, considering the generative model in Figure 3, uncertainty is treated with respect to the distribution over the parameters of the likelihood model. Similarly to RL, dropout [33] and ensembles [114] have been studied in the active inference literature, although several implementations until now have neglected this aspect, assuming confidence over a single set of parameters.

3.3. Representation

Following the variational free energy formulation from Equation (7), the agent's generative model is assumed capable of generating imaginary outcomes that match closely with the sensory perceptions through the likelihood model. This is presented on the left in Figure 4, for an environment with visual sensory data, i.e., images. The model depicted uses a sequential VAE-like setup, with the posterior encoding information from the observation (red) in the state, and the likelihood model generating observations from the state with a decoder (blue).
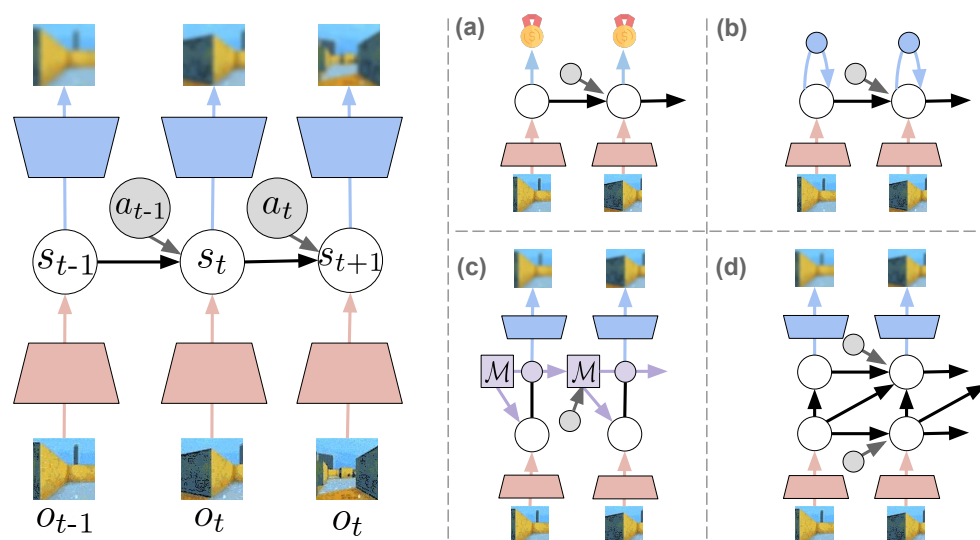

Figure 4. Representation learning approaches for world models with a latent dynamics. On the left, the base approach with the likelihood-model that reconstructs sensory information. On the right: (**a**) Task-oriented representation; (**b**) State-consistent representation; (**c**) Memory-equipped model (memory cell indicated with $\mathcal{M}$); (**d**) Hierarchical states structure.

However, learning a likelihood model from high-dimensional sensory observations, such as in pixel-based environments, is not a trivial problem. In this case, the likelihood model needs to generate images that match with the original observations pixel by pixel, requiring both a high-capacity model and a considerable accuracy, especially for high-resolution images. Most often, the probability distribution of an image is represented as a product of independent Gaussians over each pixel with fixed standard deviation, in which the log-likelihood loss in Equation (7) becomes the pixel-wise mean squared error between two images; however, this can be problematic, as it might lead to the model ignoring small but important features in the environment (as the pixel-wise loss is low) and wasting a lot of capacity in encoding potentially irrelevant information (i.e., the exact textures of a wall).

One potential solution (a) is to train the model in sight of the future task to accomplish, considering only the states related to accomplishing the agent's goal, i.e., rewards. Finding a hidden state representation that allows predicting such information, without necessarily

generating observations greatly lightens the representational burden of the model, though it would make the internal dynamics less informed. An example is illustrated in Figure 4a. Similar representations have been employed for RL [29], and could perhaps be also adapted for active inference.

Another proposed solution (b) is to replace the likelihood component of the loss with a state-consistency loss. These kinds of representations, which enforce some form of consistency between states and their corresponding sensory observations, have increasingly gained popularity in deep learning as self-supervised learning methods, such as contrastive learning [115], clustering/prototypical methods [116], distillation/self-consistency methods [117,118], and redundancy reduction/concept whitening [119,120]. These representation techniques, depicted in Figure 4b, have also been shown successful in training dynamics models for RL [121,122] and, recently, for active inference as well [123].

Rather than replacing the likelihood model, some other approaches (c–d) have instead focused on improving the capacity of the model. This is the case of the memory-equipped models (Figure 4c) and hierarchical models (Figure 4d). Using memory allows to preserve more information about other (past) observations, and has shown encouraging results in training latent dynamics models with deep learning memory models, such as LSTMs and GRUs [47–49,124]. The memory increases the capacity of the model and allows more accurate predictions of states that are far in the future, especially when the prior model is unknown and must be learned.

Hierarchical models, which in the active inference community are also referred to as deep active inference models [13,125] (with an unfortunate confusion caused by using "deep active inference" as a term for active inference methods using deep neural networks in the generative model [31,32,102,126]), use a multi-layer structure of the hidden states of the model that facilitates the modeling of part-whole or temporal hierarchies. Similarly to using a memory, a hierarchy of states can increase the representational capacity of the model and allow more accurate predictions. Some deep learning examples have already implemented this [127,128] as well as some active inference implementations for long-term navigation [69].

3.4. Summary

There are several choices to consider when designing a variational world model. In this section, we explain some of them along with providing references to a variety of studies that actually implement these mechanism, considering contributions both in the larger deep learning literature as well as in an active inference context. A summary of the design choices is presented in Table 1.

Table 1. Implementation and design choices for learning the variational world model of the agent. The table displays one or two examples for each aspect–modality pair, both in the active inference and in the more general active inference literature, when applicable.

	Modality	Active Inference	Deep Learning
States distribution	Gaussian	[32,33,114,123,129]	[49,87]
	Categorical	[91]	[66,90]
	Others	-	[92–94]
Prior model	No prior	-	[103]
	Known prior	[86,102,114]	[70,71,101]
	Learned prior	[31,33,68,91,123]	[48,49,66]
Uncertainty	Precision	[33]	[107,108]
	Ensemble	[91,114]	[112,113]
	Dropout	[33]	[130]
Representation	Task-oriented	-	[29]
	State consistency	[123]	[116,119,121,122,131]
	Memory-equipped	[32,124]	[47–50,66]
	Hierarchical	[132]	[127,128]

In the future, it will be important to continue investigating these design choices and analyzing their interplay. We believe that synergies between advances in the deep learning community and active inference adopters will be crucial to further develop generative models for a wide range of use cases. We also look forward towards novel aspects of model learning, such as the representation of time [133] or the model reduction happening during sleep [134].

It is also important to note that some of the design choices discussed, such as the internal state representation, have an impact on the model learning part but also on the agent's action selection. For instance, if there is no likelihood model, how should the agent recognize whether the preferred outcomes are being satisfied? Or, again, if the agent's model is hierarchical, how should granular actions relate to the hierarchical state structure? In the next section, we provide an overview of the techniques proposed to adopt the generative model for action selection, and elaborate on these issues.

4. Bayesian Action Selection

In order to select future actions in active inference, the agent exploits the learned model in order to match its preferred outcomes by minimizing the expected free energy. More formally, the agent's belief over which policy or sequence of actions to follow is given by:

$$p(\pi|\zeta) = \sigma(-\zeta G_\pi), \tag{8}$$

where σ is the softmax function and ζ a precision parameter. Hence, when precision is high, the agent is most likely to engage in the policy with the lowest expected free energy, whereas for (very) low precision the agent will rather randomly explore. As discussed in Section 2, the expected free energy G_π is calculated by taking expectations with respect to outcomes in the future, inside the model's predictions. Hence, the agent minimizes the expected free energy by evaluating the predicted outcomes against the preferred distribution before deploying actions in reality.

Whereas the generative model is trained to match the real outcomes of the world with past experience, future outcomes are not yet available to the agent. As an active inference agent adopts prior expectations of reaching preferred outcomes, one can interpret this as having a biased generative model of the future towards one's preferences. The self-evidencing behavior that emerges is that of a 'crooked scientist' [7], searching active states that will provide evidence for its biased hypothesis.

From a biological perspective, we could assume that every agent possesses a unique set of preferences, i.e., to maintain homeostasis [135]. These preferences could, for instance, associate internal signals, such as body temperature, hunger, happiness, and satisfaction, to the preferred states of the world. For artificial agents, defining the correct set of preferences can instead be problematic. Different ways of addressing this problem are presented in the first subsection. We also analyze the problem of dealing with the agent's uncertainty and how to learn and/or amortize the action selection process.

4.1. Preferences Modeling

As summarized in Section 2, the expected free energy objective can be factored in several ways, each highlighting different emergent properties of the agent's behavior (Equations (4)–(6)). While this aspect of active inference has been the target of critics [136], this allows for greater flexibility in designing the agent selection process.

Observation Preferences. If the agent's objective is to match a set of preferred outcomes, the preferred distribution is over the environment's observations $p(o)$. Matching outcomes can be seen as a form of goal-directed behavior, where the agent plans its actions to achieve certain outcomes from the environment. Goal-directed behavior has been widely studied in the context of RL, both in low-dimensional [137] and visual domains [138,139]. Preferences defined in the observation space can be handy, as they just require observations from "snapshots" of the environment in the correct state. Nevertheless, artificial active inference implementations have rarely used them, as they are generally hard to match in

the high-dimensional settings. Strategies that overcome such limitations [123] could be the subject of future studies.

Internal State Preferences. Instead of defining preferences in observation space, these could be directly instantiated in the internal state space of the agent. This form of state matching [140] assumes that the agent knows both the preferred states distribution $p(s)$ and the model in advance, or as typical in RL, that sensory states are used as internal states. Alternatively, if a set of preferred outcomes is available, preferred states can be inferred from those using an inference model $p(s|o)$. This approach has been applied in robotics simulated and realistic setups [32,68].

Rewards as Preferences. Another way to circumvent the problem of defining preferences is to use a reward function that represents the agent's probability of observing the preferred outcomes. The RL problem can be cast as probabilistic inference, by introducing an optimality variable $\mathcal{O}_t$, which denotes whether the time step t is optimal [141]. The distribution over the optimality variable is defined in terms of rewards as $p(\mathcal{O}_t = 1|s_t, a_t) = \exp(r(s_t, a_t))$. As discussed in [142], RL works alike active inference but it encodes utility value in the optimality likelihood rather than in a prior over observations. Assuming $\log p(o_t) = \log p(\mathcal{O}_t|s_t, a_t)$, the environment rewards can be used for active inference as well. This possibility has allowed some active inference work [33,114] to reuse reward functions from RL environments [22]. Concretely, it is possible to consider rewards as a part of the observable aspects of the environment, and define their maximum values as the preferred observations [143]. Nonetheless, defining reward functions is also problematic [144] as they are not naturally available, and this setup works well only for well-engineered environments.

Learned Preferences. Finally, state preferences can also be learned from previous experience using conjugate priors [91], or from expert demonstrations [68]. In a RL context, demonstrations can be used in an inverse reinforcement learning fashion [145,146], where a reward signal is inferred from correct behaviors, which is then optimized using RL techniques.

4.2. Epistemics, Exploration, and Ambiguity

While active inference agents seek to realize their preferences, they also aim to reduce the uncertainty of their model. For instance, if an agent has to manipulate some objects in a dark room, it would first search for the light switch to increase the confidence of its model and reduce the resulting ambiguity of its actions. As also shown in Section 2, the causes of the agent's ambiguity can be twofold: on the one hand, it can be due to the incapability of inferring its state with certainty, referring to uncertainty in the state-observation mapping, e.g., likelihood entropy or mutual information; on the other hand, the uncertainty can be caused by the agent's lack of confidence with respect to the model's parameters. As depicted in Equation (4), an agent's drive for epistemic foraging is caused by maximizing two information gain terms: information gain on model parameters and information gain on hidden states.

Parameter-driven Exploration. Maximizing mutual information in parameter space has been studied in RL as a way to encourage exploration, computing the information gain given by the distribution over parameters with ensembles [112,113,147] or Bayesian neural networks [53]. In particular, in [113], they use the model to both evaluate the states/actions to explore and to plan the exploratory behavior, which is close to what envisioned in active inference. Ensemble methods have also been employed in some active inference works [114,129] along with dropout [33].

State-driven Exploration. Maximizing mutual information between states and observations has also been studied in RL for exploration, using the Bayesian surprise signal given by the D_{KL} divergence between the (autoencoding) posterior and the prior of the model as a reward [51]. Alternatively, the surprisal with respect to future observations has also been used in RL to generate an intrinsic motivation signal that rewards exploration [52,148,149].

In active inference, the majority of works have instead focused on using multiple samples from the likelihood model [32,33].

Uncertainty Tradeoffs. It is worth mentioning that, during different stages of training, uncertainty related to parameters and uncertainty related to sensory/internal states may overlap. Particularly, given that the distributions that represent the agents' states are inferred by employing the model parameters, uncertainty in the model strongly influences uncertainty with respect to the state. This highlights the importance of considering both kinds of uncertainty, especially when the model is imperfect or its learning process is incomplete.

From an engineering perspective, having to deal with multiple signals, as in the active inference objective, poses additional optimization problems. Different parts of the objectives may provide values on different scales, depending on the different models, distributions and on the sensory data processed by the agent. In the RL landscape, how to combine the environments' rewards with exploration bonuses to obtain the best performance is an ongoing research problem. While one might consider using the 'vanilla' objective, with no weighting of the different components, weighting could lead to different behaviors that might come in use for practical purposes, e.g., reducing the weight on the exploration/ambiguity terms might lead to faster convergence when there is no ambiguity/need to explore in the environment.

4.3. Plans, Habits, and Search Optimization

From a computational point of view, the most complex aspect of minimizing the expected free energy consists of how selecting the actions that will accomplish the agent's belief. In practice, optimizing G_π becomes a tree search, optionally pruning away from the search all policies that fall outside of an Occam's window, which are the policies that have a very low posterior probability. Nonetheless, depending on the way policies are defined, the search can still be significantly expensive, especially in high-dimensional and continuous actions domains.

We distinguish three ways of establishing action selection, summarized in Figure 5. The first is the typical active inference's definition, with the policy being a sequence of actions $\pi = \{a_1, a_2, a_3, \ldots\}$, and we will refer to these policies as *plans* for distinction (Figure 5a). Each plan is evaluated by its expected free energy, and the next action is selected from the best plan according to Equation (8). The second way of defining a policy is by learning a state-action mapping $\pi(s_t)$, which is amortizing policy selection by finding an optimal *habit* policy that outputs the expected best action for each state (Figure 5b). This is also the notion of a policy that is accustomed in a typical RL setting. Finally, it is possible to combine both worlds by first estimating the expected free energy for a given state and action, and then performing a search over the reduced search space (Figure 5c).

Plan-based policies. Assuming a complete search over all potential sequences of actions, the plan-based method should yield the optimal policy. Unfortunately, in most domains, considering all sequences of actions is an intractable problem and more engineered random shooting methods are used to search only over the most promising sequences of actions, such as [55]. Similar methods have been employed both for RL [49] and active inference [32,68]. In particular, when the search over policies takes into account recursive beliefs about the future, this scheme is referred to as sophisticated inference [74]. Sophistication describes the degree to which an agent has beliefs about beliefs. A sophisticated agent, when evaluating a sequence of actions, instead of directly considering the sequence of outcomes, recursively evaluates outcomes in terms of the beliefs it would have when applying each action of the sequence.

Habit Policies. For habit policies, we consider a one-action version of the expected free energy $\mathcal{G}$ that can be obtained by considering one-action plans $\pi = a_t$ for all time steps:

$$\mathcal{G}_{a_t} = \mathbb{E}_{q(s_{t+1}, o_{t+1}|s_t, a_t)}[\log q(s_{t+1}|s_t, a_t) - \log p(o_{t+1}, s_{t+1}|s_t, a_t)] \tag{9}$$
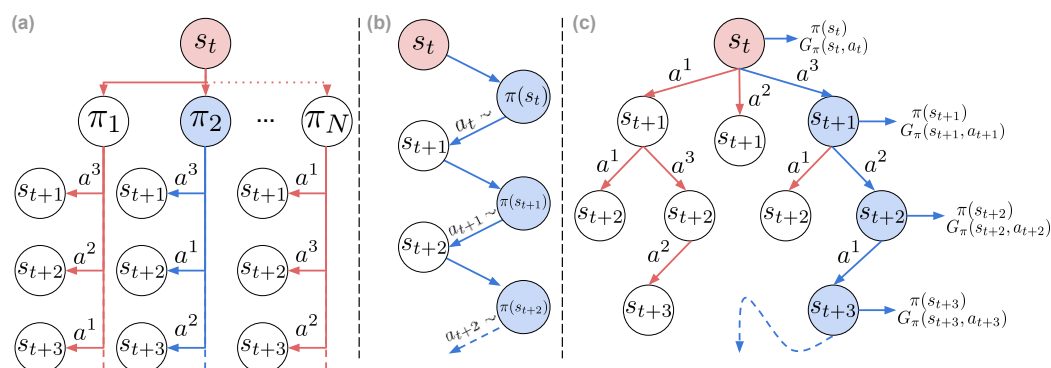
Figure 5. Different approaches for selecting actions. Blue circles represent the path selected by the agent. (**a**) Deep search via action plans: the path selected has the lowest free energy. (**b**) Habit learning via state-action policies: the agent always samples from the same conditional distribution. (**c**) Tree search guided by value and policy: the agent selects the actions according to the prior and the expected free energy.

A state-action policy $\pi(s_t)$ can then be trained to maximize the above signal over multiple time steps, as typically performed in policy gradient methods [150,151]. In order to plan for longer horizons, deep RL methods adopt value functions that allow estimating the expected sum of rewards over time, over a potentially infinite horizon. For these long-term estimates, the value functions utilize a dynamic programming approach, where values are continuously updated bootstrapping current estimates with actual data. From an active inference perspective, it is also possible to estimate the expected free energy for a longer horizon by applying dynamic programming, similarly to what was studied in [126]. The expected free energy can then be rewritten and optimized recursively as follows:

$$G_\pi(s_t, a_t) = \mathcal{G}(s_t, a_t) + \gamma \, \mathbb{E}_{a_{t+1} \sim \pi(s_{t+1})} [G_\pi(s_{t+1}, a_{t+1})], \tag{10}$$

where G_π represents an estimate of the expected free energy following the policy π and the expectation over π means the actions are sampled from the state-action policy distribution. The above equation resembles the Bellman equation known from RL with gamma being an (optional) discount factor that is used to avoid infinite sum. This optimization scheme leads to an habit policy that can achieve optimal behavior, when the sources of uncertainty of the environment are stationary; however, habitual learning can be insufficient in realistic scenarios, where rare and unexpected events are common. In this action selection scheme, the precision parameter ζ controls the entropy of the state-action policy distribution, similarly to maximum-entropy control approaches [151,152].

Hybrid Search Policies. Finally, hybrid search schemes (c) combine the use of a learned prior with computing the expected free energy for sequences of actions. The search space is greatly limited by using the prior, which influences the choice of the nodes to select and expand. One of the most popular applications in RL of these methods is by employing variants of Monte Carlo Tree Search (MCTS) [29,153], which use both a prior over actions and estimates of the expected utility over long horizons, as in Equation (10). Similar approaches have recently been applied for active inference [33,154]. While these methods are generally applicable only for discrete action spaces, extensions of MCTS for continuous domains have been developed as well [56]. The precision parameter ζ in these methods can be used to control the influence of the prior relative to the expected free energy (computed a posteriori, with respect to a certain action/plan).

4.4. Summary

Similar as when designing the agent's model, there are several aspects to consider for implementing action selection in an active inference fashion. In this section, we covered a number of important aspects and provided references to existing implementations, both in

the active inference and in the deep reinforcement learning landscape. A summary of these methods is presented in Table 2.

Table 2. Implementation and design choices for the action selection process, minimizing the expected free energy. The table displays one or two examples for each aspect–modality pair, both in the active inference and in the deep learning (mainly, reinforcement learning) literature, when applicable. * All active inference methods generally consider hidden states exploration.

	Modality	Active Inference	Deep Learning
Preferences	Observations	[123]	[137–139]
	States	[68]	[140]
	Rewards	[33,114,126,129]	[66,67,87,155]
	Learned	[91]	[145,146]
Exploration	Hidden states	*	[51,52,148,149]
	Likelihood parameters	[33,91,114]	[112,113]
Action selection	Action plans	[32,68]	[55]
	State-action policy	[123,126]	[150,151]
	Amortized search	[33,154]	[29]

There are still several challenges to overcome in addition to the discussion so far. For instance, defining the preferences for an artificial agent is still unresolved for many practical applications. Future work should also address hierarchical implementations for action selection, to accompany hierarchical models [133], allowing to amortize and abstract the action selection further. Another interesting interesting avenue consists of investigating episodic control (currently less studied in the active inference literature), since this has played an important role for improving performance in RL [156].

5. Discussion and Perspectives

Developing artificial intelligence is a complex and intriguing problem. Among the capabilities that artificial intelligent agents should possess, the ability to sense and to act consistently is crucial. Intelligent agents should be able to exhibit their intelligence, manipulating the environment according to their will or purpose, and to understand the consequences of their actions, in order to provide a closed-loop feedback to their acting system and to acknowledge the accomplishment of their desires.

Active inference is a neuro-inspired framework that encompasses both a perception process, through learning a variational world model, and a Bayesian action selection process, which considers both preference satisfaction and uncertainty in the environment and in the agent's model. The variational inference optimization scheme and the amortization of inference, when learning the model parameters and selecting actions, make the framework promising for practical implementations, however, without scalable models, it is unfeasible to apply active inference in complex scenarios, with continuous and/or high-dimensional state/action spaces. We showed how active inference can be combined with deep learning models for function approximation to provide implementations that scale to more complex environments, with the potential of applying it in realistic scenarios.

As mentioned in Section 3, neural networks can learn to approximate any function with an arbitrary degree of accuracy. Considering multi-layer perceptron models, this should theoretically achievable by using shallow networks with two hidden layers and sufficient capacity; however, empirical evidence has shown that deep neural networks tend to converge to more accurate models more easily than shallow networks [157] and that overparameterized neural networks generally lead to more accurate predictions [158]. One of the major issues of deep learning might be attributed to its gradient-based learning rule. While there are, indeed, proofs that an arbitrarily accurate neural network might exists, there is no certainty about a way to find the accurate model given a set of data, and thus the learning problem might be undecidable [159,160]. Nonetheless, we believe that the strong empirical results obtained by using deep neural networks justify their practical utility for implementing active inference. Furthermore, several of the modeling choices

here discussed, such as the definition of the internal state distribution or the modality to amortize the action selection process, might be useful in other learning/modeling frameworks as well.

One of the intentions of this work is to provide an introduction to active inference and guidelines for deep learning researchers to hit the ground running in a given field by exploiting concepts that are in common between the two areas. At the same time, this article can be used as a reference for scientists intending to address some of the issues that hinder artificial implementations of the active inference framework.

We presented several design choices that need to be addressed to instantiate artificial active inference agents with deep learning models, attempting to relate them to well-established studies in both fields. In particular, we found that some aspects of active inference are well reflected in some areas of deep learning, such as unsupervised learning, representation learning, and reinforcement learning, whose findings can be used to push the boundaries of active inference further. In turn, active inference provides a framework for perception and action, from which individual approaches could obtain insights to both expand their scope or understand the implications of their work from a larger perspective.

Author Contributions: P.M. and T.V. contributed to the conceptualization and the writing. P.M. took care of figures and visualizations. O.Ç. and B.D. helped reviewing and editing the article. All authors have read and agreed to the published version of the manuscript.

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

References

1. Friston, K.J.; Stephan, K.E. Free-energy and the brain. *Synthese* **2007**, *159*, 417–458. doi: 10.1007/s11229-007-9237-y. [CrossRef] [PubMed]
2. Friston, K.; FitzGerald, T.; Rigoli, F.; Schwartenbeck, P.; Doherty, J.O.; Pezzulo, G. Active inference and learning. *Neurosci. Biobehav. Rev.* **2016**, *68*, 862–879. [CrossRef] [PubMed]
3. Parr, T.; Rees, G.; Friston, K.J. Computational Neuropsychology and Bayesian Inference. *Front. Hum. Neurosci.* **2018**, *12*, 61. [CrossRef]
4. Demekas, D.; Parr, T.; Friston, K.J. An Investigation of the Free Energy Principle for Emotion Recognition. *Front. Comput. Neurosci.* **2020**, *14*, 30. [CrossRef]
5. Henriksen, M. Variational Free Energy and Economics Optimizing with Biases and Bounded Rationality. *Front. Psychol.* **2020**, *11*, 549187. [CrossRef]
6. Constant, A.; Ramstead, M.J.D.; Veissière, S.P.L.; Campbell, J.O.; Friston, K.J. A variational approach to niche construction. *J. R. Soc. Interface* **2018**, *15*, 20170685. [CrossRef]
7. Bruineberg, J.; Rietveld, E.; Parr, T.; van Maanen, L.; Friston, K.J. Free-energy minimization in joint agent-environment systems: A niche construction perspective. *J. Theor. Biol.* **2018**, *455*, 161–178. [CrossRef]
8. Perrinet, L.U.; Adams, R.A.; Friston, K.J. Active inference, eye movements and oculomotor delays. *Biol. Cybern.* **2014**, *108*, 777–801. [CrossRef]
9. Parr, T.; Friston, K.J. Active inference and the anatomy of oculomotion. *Neuropsychologia* **2018**, *111*, 334–343. [CrossRef]
10. Brown, H.; Friston, K.; Bestmann, S. Active Inference, Attention, and Motor Preparation. *Front. Psychol.* **2011**, *2*, 218. [CrossRef]
11. Parr, T.; Friston, K.J. Working memory, attention, and salience in active inference. *Sci. Rep.* **2017**, *7*, 14678. [CrossRef] [PubMed]
12. Mirza, M.B.; Adams, R.A.; Mathys, C.D.; Friston, K.J. Scene Construction, Visual Foraging, and Active Inference. *Front. Comput. Neurosci.* **2016**, *10*, 56. [CrossRef] [PubMed]
13. Heins, R.C.; Mirza, M.B.; Parr, T.; Friston, K.; Kagan, I.; Pooresmaeili, A. Deep Active Inference and Scene Construction. *Front. Artif. Intell.* **2020**, *3*, 81. [CrossRef] [PubMed]
14. Biehl, M.; Pollock, F.A.; Kanai, R. A Technical Critique of Some Parts of the Free Energy Principle. *Entropy* **2021**, *23*, 293. [CrossRef]
15. Friston, K.J.; Da Costa, L.; Parr, T. Some Interesting Observations on the Free Energy Principle. *Entropy* **2021**, *23*, 1076. [CrossRef]
16. Friston, K. Life as we know it. *J. R. Soc. Interface* **2013**, *10*, 20130475. [CrossRef]
17. Kirchhoff, M.; Parr, T.; Palacios, E.; Friston, K.; Kiverstein, J. The Markov blankets of life: Autonomy, active inference and the free energy principle. *J. R. Soc. Interface* **2018**, *15*, 20170792. [CrossRef]
18. Rubin, S.; Parr, T.; Da Costa, L.; Friston, K. Future climates: Markov blankets and active inference in the biosphere. *J. R. Soc. Interface* **2020**, *17*, 20200503. [CrossRef]
19. Maturana, H.R.; Varela, F.J.; Maturana, H.R. *Autopoiesis and Cognition: The Realization of the Living*; D. Reidel Pub. Co: Dordrecht, The Netherlands, 1980.

20. Kirchhoff, M.D. Autopoiesis, free energy, and the life–mind continuity thesis. *Synthese* **2018**, *195*, 2519–2540. [CrossRef]

21. Blei, D.M.; Kucukelbir, A.; McAuliffe, J.D. Variational Inference: A Review for Statisticians. *J. Am. Stat. Assoc.* **2017**, *112*, 859–877. [CrossRef]

22. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.

23. Wise, R.A. Dopamine, learning and motivation. *Nat. Rev. Neurosci.* **2004**, *5*, 483–494. [CrossRef] [PubMed]

24. Glimcher, P.W. Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 15647–15654. [CrossRef] [PubMed]

25. Silver, D.; Singh, S.; Precup, D.; Sutton, R.S. Reward is enough. *Artif. Intell.* **2021**, *299*, 103535. [CrossRef]

26. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing Atari with Deep Reinforcement Learning. *arXiv* **2013**, arXiv:1312.5602.

27. Badia, A.P.; Piot, B.; Kapturowski, S.; Sprechmann, P.; Vitvitskyi, A.; Guo, D.; Blundell, C. Agent57: Outperforming the Atari Human Benchmark. *arXiv* **2020**, arXiv:2003.13350.

28. Vinyals, O.; Babuschkin, I.; Czarnecki, W.M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D.H.; Powell, R.; Ewalds, T.; Georgiev, P.; et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* **2019**, *575*, 350–354. [CrossRef]

29. Schrittwieser, J.; Antonoglou, I.; Hubert, T.; Simonyan, K.; Sifre, L.; Schmitt, S.; Guez, A.; Lockhart, E.; Hassabis, D.; Graepel, T.; et al. Mastering Atari, Go, chess and shogi by planning with a learned model. *Nature* **2020**, *588*, 604–609. [CrossRef]

30. Akkaya, I.; Andrychowicz, M.; Chociej, M.; Litwin, M.; McGrew, B.; Petron, A.; Paino, A.; Plappert, M.; Powell, G.; OpenAI; et al. Solving Rubik's Cube with a Robot Hand. *arXiv* **2019**, arXiv:1910.07113.

31. Ueltzhöffer, K. Deep active inference. *Biol. Cybern.* **2018**, *112*, 547–573. [CrossRef]

32. Çatal, O.; Verbelen, T.; Nauta, J.; De Boom, C.; Dhoedt, B. Learning Perception and Planning with Deep Active Inference. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 3952–3956. doi: 10.1109/ICASSP40776.2020.9054364. [CrossRef]

33. Fountas, Z.; Sajid, N.; Mediano, P.; Friston, K. Deep active inference agents using Monte-Carlo methods. In *Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2020; Volume 33, pp. 11662–11675.

34. Buckley, C.L.; Kim, C.S.; McGregor, S.; Seth, A.K. The free energy principle for action and perception: A mathematical review. *J. Math. Psychol.* **2017**, *81*, 55–79. [CrossRef]

35. Da Costa, L.; Parr, T.; Sajid, N.; Veselic, S.; Neacsu, V.; Friston, K. Active inference on discrete state-spaces: A synthesis. *J. Math. Psychol.* **2020**, *99*, 102447. [CrossRef] [PubMed]

36. Lanillos, P.; Meo, C.; Pezzato, C.; Meera, A.A.; Baioumy, M.; Ohata, W.; Tschantz, A.; Millidge, B.; Wisse, M.; Buckley, C.L.; et al. Active Inference in Robotics and Artificial Agents: Survey and Challenges. *arXiv* **2021**, arXiv:2112.01871.

37. Gershman, S.; Goodman, N. Amortized inference in probabilistic reasoning. *Proc. Annu. Meet. Cogn. Sci. Soc.* **2014**, *36*, 516–522.

38. Razavi, A.; van den Oord, A.; Vinyals, O. Generating Diverse High-Fidelity Images with VQ-VAE-2. *arXiv* **2019**, arXiv:1906.00446.

39. Karras, T.; Aittala, M.; Laine, S.; Härkönen, E.; Hellsten, J.; Lehtinen, J.; Aila, T. Alias-Free Generative Adversarial Networks. *arXiv* **2021**, arXiv:2106.12423.

40. Vahdat, A.; Kautz, J. NVAE: A Deep Hierarchical Variational Autoencoder. *arXiv* **2021**, arXiv:2007.03898.

41. Zilly, J.G.; Srivastava, R.K.; Koutník, J.; Schmidhuber, J. Recurrent Highway Networks. *arXiv* **2017**, arXiv:1607.03474.

42. Melis, G.; Kočiský, T.; Blunsom, P. Mogrifier LSTM. *arXiv* **2020**, arXiv:1909.01792.

43. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models Are Few-Shot Learners. *arXiv* **2020**, arXiv:2005.14165.

44. Xingjian, S.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 802–810.

45. Wang, Y.; Long, M.; Wang, J.; Gao, Z.; Yu, P.S. PredRNN: Recurrent Neural Networks for Predictive Learning using Spatiotemporal LSTMs. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.

46. Denton, E.; Fergus, R. Stochastic Video Generation with a Learned Prior. *arXiv* **2018**, arXiv:1802.07687.

47. Lotter, W.; Kreiman, G.; Cox, D. Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. *arXiv* **2017**, arXiv:1605.08104.

48. Buesing, L.; Weber, T.; Racaniere, S.; Eslami, S.M.A.; Rezende, D.; Reichert, D.P.; Viola, F.; Besse, F.; Gregor, K.; Hassabis, D.; et al. Learning and Querying Fast Generative Models for Reinforcement Learning. *arXiv* **2018**, arXiv:1802.03006.

49. Hafner, D.; Lillicrap, T.; Fischer, I.; Villegas, R.; Ha, D.; Lee, H.; Davidson, J. Learning Latent Dynamics for Planning from Pixels. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 5–9 June 2019; Chaudhuri, K., Salakhutdinov, R., Eds.; PMLR: Brookline, MA, USA, 2019; Volume 97, pp. 2555–2565.

50. Ha, D.; Schmidhuber, J. Recurrent World Models Facilitate Policy Evolution. *arXiv* **2018**, arXiv:1809.01999.

51. Mazzaglia, P.; Catal, O.; Verbelen, T.; Dhoedt, B. Self-Supervised Exploration via Latent Bayesian Surprise. *arXiv* **2021**, arXiv:2104.07495.

52. Pathak, D.; Agrawal, P.; Efros, A.A.; Darrell, T. Curiosity-driven Exploration by Self-supervised Prediction. *arXiv* **2017**, arXiv:1705.05363.

53. Houthooft, R.; Chen, X.; Duan, Y.; Schulman, J.; De Turck, F.; Abbeel, P. VIME: Variational Information Maximizing Exploration. In Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS'16, Barcelona, Spain, 5–10 December 2016; pp. 1117–1125.

54. Çatal, O.; Leroux, S.; De Boom, C.; Verbelen, T.; Dhoedt, B. Anomaly Detection for Autonomous Guided Vehicles using Bayesian Surprise. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October–24 January 2020; pp. 8148–8153. [CrossRef]

55. Hansen, N. The CMA Evolution Strategy: A Tutorial. *arXiv* **2016**, arXiv:1604.00772.

56. Hubert, T.; Schrittwieser, J.; Antonoglou, I.; Barekatain, M.; Schmitt, S.; Silver, D. Learning and Planning in Complex Action Spaces. *arXiv* **2021**, arXiv:2104.06303.

57. Von Helmholtz, H. *Handbuch der Physiologischen Optik: Mit 213 in den Text Eingedruckten Holzschnitten und 11 Tafeln*; Wentworth Press: Sydney, Australia, 1867; Volume 9.

58. Friston, K. The free-energy principle: A rough guide to the brain? *Trends Cogn. Sci.* **2009**, *13*, 293–301. [CrossRef]

59. Ramstead, M.J.; Kirchhoff, M.D.; Friston, K.J. A tale of two densities: Active inference is enactive inference. *Adapt. Behav.* **2020**, *28*, 225–239. [CrossRef]

60. Friston, K.J.; Parr, T.; de Vries, B. The graphical brain: Belief propagation and active inference. *Netw. Neurosci.* **2017**, *1*, 381–414. [CrossRef] [PubMed]

61. Friston, K.J.; Daunizeau, J.; Kiebel, S.J. Reinforcement Learning or Active Inference? *PLoS ONE* **2009**, *4*, e6421. [CrossRef] [PubMed]

62. Karl, F. A Free Energy Principle for Biological Systems. *Entropy* **2012**, *14*, 2100–2121. [CrossRef]

63. Schwartenbeck, P.; Passecker, J.; Hauser, T.U.; FitzGerald, T.H.; Kronbichler, M.; Friston, K.J. Computational mechanisms of curiosity and goal-directed exploration. *eLife* **2019**, *8*, e41703. [CrossRef]

64. Friston, K.J.; Lin, M.; Frith, C.D.; Pezzulo, G.; Hobson, J.A.; Ondobaka, S. Active Inference, Curiosity and Insight. *Neural Comput.* **2017**, *29*, 2633–2683. [CrossRef]

65. Friston, K.; Rigoli, F.; Ognibene, D.; Mathys, C.; Fitzgerald, T.; Pezzulo, G. Active inference and epistemic value. *Cogn. Neurosci.* **2015**, *6*, 187–214. [CrossRef]

66. Hafner, D.; Lillicrap, T.; Norouzi, M.; Ba, J. Mastering Atari with Discrete World Models. *arXiv* **2021**, arXiv:2010.02193.

67. Hafner, D.; Lillicrap, T.P.; Ba, J.; Norouzi, M. Dream to Control: Learning Behaviors by Latent Imagination. In Proceedings of the ICLR Conference, Addis Abeba, Ethiopia, 26 April–1 May 2020.

68. Çatal, O.; Nauta, J.; Verbelen, T.; Simoens, P.; Dhoedt, B. Bayesian policy selection using active inference. *arXiv* **2019**, arXiv:1904.08149.

69. Çatal, O.; Verbelen, T.; Van de Maele, T.; Dhoedt, B.; Safron, A. Robot navigation as hierarchical active inference. *Neural Netw.* **2021**, *142*, 192–204. [CrossRef]

70. Kingma, D.P.; Welling, M. Auto-Encoding Variational Bayes. *arXiv* **2014**, arXiv:1312.6114.

71. Rezende, D.J.; Mohamed, S.; Wierstra, D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In Proceedings of the 31st International Conference on Machine Learning (ICML), Beijing, China, 21–26 June 2014; Volume 32, pp. 1278–1286.

72. Tishby, N.; Pereira, F.C.; Bialek, W. The information bottleneck method. *arXiv* **2000**, arXiv:physics/0004057.

73. Alemi, A.A.; Fischer, I.; Dillon, J.V.; Murphy, K. Deep Variational Information Bottleneck. *arXiv* **2019**, arXiv:1612.00410.

74. Friston, K.; Da Costa, L.; Hafner, D.; Hesp, C.; Parr, T. Sophisticated Inference. *Neural Comput.* **2021**, *33*, 713–763. [CrossRef]

75. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **1991**, *4*, 251–257. [CrossRef]

76. Heiden, E.; Millard, D.; Coumans, E.; Sheng, Y.; Sukhatme, G.S. NeuralSim: Augmenting Differentiable Simulators with Neural Networks. *arXiv* **2021**, arXiv:2011.04217.

77. Freeman, C.D.; Frey, E.; Raichuk, A.; Girgin, S.; Mordatch, I.; Bachem, O. Brax—A Differentiable Physics Engine for Large Scale Rigid Body Simulation. *arXiv* **2021**, arXiv:2106.13281.

78. Lovejoy, W.S. A survey of algorithmic methods for partially observed Markov decision processes. *Ann. Oper. Res.* **1991**, *28*, 47–65. [CrossRef]

79. Roy, N.; Gordon, G.; Thrun, S. Finding Approximate POMDP solutions Through Belief Compression. *J. Artif. Intell. Res.* **2005**, *23*, 1–40. [CrossRef]

80. Kurniawati, H.; Hsu, D.; Lee, W.S. Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. In *Robotics: Science and Systems*; Citeseer: Pennsylvania, PA, USA, 2008; Volume 2008.

81. Heess, N.; Hunt, J.J.; Lillicrap, T.P.; Silver, D. Memory-based control with recurrent neural networks. *arXiv* **2015**, arXiv:1512.04455.

82. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]

83. Bengio, Y.; Léonard, N.; Courville, A. Estimating or Propagating Gradients through Stochastic Neurons for Conditional Computation. *arXiv* **2013**, arXiv:1308.3432.

84. Glynn, P.W. Likelilood ratio gradient estimation: An overview. In Proceedings of the 19th Conference on Winter Simulation, Atlanta, GA, USA, 14–16 December 1987; pp. 366–375.

85. Williams, R.J. Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Mach. Learn.* **1992**, *8*, 229–256. [CrossRef]

86. Van de Maele, T.; Verbelen, T.; Çatal, O.; De Boom, C.; Dhoedt, B. Active Vision for Robot Manipulators Using the Free Energy Principle. *Front. Neurorobot.* **2021**, *15*, 14. [CrossRef] [PubMed]

87. Lee, A.X.; Nagabandi, A.; Abbeel, P.; Levine, S. Stochastic Latent Actor-Critic: Deep Reinforcement Learning with a Latent Variable Model. *arXiv* **2020**, arXiv:1907.00953.

88. Igl, M.; Zintgraf, L.; Le, T.A.; Wood, F.; Whiteson, S. Deep Variational Reinforcement Learning for POMDPs. *arXiv* **2018**, arXiv:1806.02426.

89. Rolfe, J.T. Discrete variational autoencoders. *arXiv* **2016**, arXiv:1609.02200.

90. Ozair, S.; Li, Y.; Razavi, A.; Antonoglou, I.; van den Oord, A.; Vinyals, O. Vector Quantized Models for Planning. *arXiv* **2021**, arXiv:2106.04615.

91. Sajid, N.; Tigas, P.; Zakharov, A.; Fountas, Z.; Friston, K. Exploration and preference satisfaction trade-off in reward-free learning. *arXiv* **2021**, arXiv:2106.04316.

92. Serban, I.V.; Ororbia, A.G.; Pineau, J.; Courville, A. Piecewise Latent Variables for Neural Variational Text Processing. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 7–11 September 2017; Association for Computational Linguistics: Copenhagen, Denmark, 2017; pp. 422–432. [CrossRef]

93. Rezende, D.J.; Mohamed, S. Variational Inference with Normalizing Flows. *arXiv* **2016**, arXiv:1505.05770.

94. Salimans, T.; Kingma, D.P.; Welling, M. Markov Chain Monte Carlo and Variational Inference: Bridging the Gap. *arXiv* **2015**, arXiv:1410.6460.

95. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]

96. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.

97. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386. [CrossRef] [PubMed]

98. Scarselli, F.; Gori, M.; Tsoi, A.C.; Hagenbuchner, M.; Monfardini, G. The Graph Neural Network Model. *IEEE Trans. Neural Netw.* **2009**, *20*, 61–80. [CrossRef]

99. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

100. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. In Proceedings of the NIPS 2014 Workshop on Deep Learning, Montreal, QC, Canada, 12–13 December 2014.

101. Toth, P.; Rezende, D.J.; Jaegle, A.; Racanière, S.; Botev, A.; Higgins, I. Hamiltonian Generative Networks. *arXiv* **2020**, arXiv:1909.13789.

102. Sancaktar, C.; van Gerven, M.A.J.; Lanillos, P. End-to-End Pixel-Based Deep Active Inference for Body Perception and Action. In Proceedings of the 2020 Joint IEEE 10th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), Valparaiso, Chile, 26–30 October 2020. [CrossRef]

103. Ghosh, P.; Sajjadi, M.S.M.; Vergari, A.; Black, M.; Schölkopf, B. From Variational to Deterministic Autoencoders. *arXiv* **2020**, arXiv:1903.12436.

104. Friston, K.; Schwartenbeck, P.; Fitzgerald, T.; Moutoussis, M.; Behrens, T.; Dolan, R. The anatomy of choice: Active inference and agency. *Front. Hum. Neurosci.* **2013**, *7*, 598. [CrossRef]

105. Parr, T.; Benrimoh, D.A.; Vincent, P.; Friston, K.J. Precision and False Perceptual Inference. *Front. Integr. Neurosci.* **2018**, *12*, 39. [CrossRef]

106. Parr, T.; Friston, K.J. Uncertainty, epistemics and active inference. *J. R. Soc. Interface* **2017**, *14*, 20170376. [CrossRef]

107. Higgins, I.; Matthey, L.; Pal, A.; Burgess, C.P.; Glorot, X.; Botvinick, M.M.; Mohamed, S.; Lerchner, A. Beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. In Proceedings of the ICLR Conference, Toulon, France, 24–26 April 2017.

108. Razavi, A.; van den Oord, A.; Poole, B.; Vinyals, O. Preventing Posterior Collapse with delta-VAEs. *arXiv* **2019**, arXiv:1901.03416.

109. Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; Wierstra, D. Weight Uncertainty in Neural Networks. *arXiv* **2015**, arXiv:1505.05424.

110. Gal, Y.; Ghahramani, Z. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *arXiv* **2016**, arXiv:1506.02142.

111. Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles. *arXiv* **2017**, arXiv:1612.01474.

112. Pathak, D.; Gandhi, D.; Gupta, A. Self-Supervised Exploration via Disagreement. *arXiv* **2019**, arXiv:1906.04161.

113. Sekar, R.; Rybkin, O.; Daniilidis, K.; Abbeel, P.; Hafner, D.; Pathak, D. Planning to Explore via Self-Supervised World Models. In Proceedings of the ICML Conference, Virtual Conference, 12–18 July 2020.

114. Tschantz, A.; Millidge, B.; Seth, A.K.; Buckley, C.L. Reinforcement Learning through Active Inference. *arXiv* **2020**, arXiv:2002.12636.

115. van den Oord, A.; Li, Y.; Vinyals, O. Representation Learning with Contrastive Predictive Coding. *arXiv* **2019**, arXiv:1807.03748.

116. Caron, M.; Misra, I.; Mairal, J.; Goyal, P.; Bojanowski, P.; Joulin, A. Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. *arXiv* **2021**, arXiv:2006.09882.

117. Grill, J.B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.H.; Buchatskaya, E.; Doersch, C.; Pires, B.A.; Guo, Z.D.; Azar, M.G.; et al. Bootstrap your own latent: A new approach to self-supervised Learning. *arXiv* **2020**, arXiv:2006.07733.

118. Chen, X.; He, K. Exploring Simple Siamese Representation Learning. *arXiv* **2020**, arXiv:2011.10566.
119. Zbontar, J.; Jing, L.; Misra, I.; LeCun, Y.; Deny, S. Barlow Twins: Self-Supervised Learning via Redundancy Reduction. *arXiv* **2021**, arXiv:2103.03230.
120. Chen, Z.; Bei, Y.; Rudin, C. Concept whitening for interpretable image recognition. *Nat. Mach. Intell.* **2020**, *2*, 772–782. [CrossRef]
121. Schwarzer, M.; Anand, A.; Goel, R.; Hjelm, R.D.; Courville, A.; Bachman, P. Data-Efficient Reinforcement Learning with Self-Predictive Representations. *arXiv* **2021**, arXiv:2007.05929.
122. Ma, X.; Chen, S.; Hsu, D.; Lee, W.S. Contrastive Variational Model-Based Reinforcement Learning for Complex Observations. In Proceedings of the 4th Conference on Robot Learning, Virtual Conference, 16–18 November 2020.
123. Mazzaglia, P.; Verbelen, T.; Dhoedt, B. Contrastive Active Inference. In Proceedings of the Advances in Neural Information Processing Systems, Virtual Conference, 6–14 December 2021.
124. Çatal, O.; Wauthier, S.; De Boom, C.; Verbelen, T.; Dhoedt, B. Learning Generative State Space Models for Active Inference. *Front. Comput. Neurosci.* **2020**, *14*, 103. [CrossRef]
125. Friston, K.J.; Rosch, R.; Parr, T.; Price, C.; Bowman, H. Deep temporal models and active inference. *Neurosci. Biobehav. Rev.* **2017**, *77*, 388–402. [CrossRef]
126. Millidge, B. Deep Active Inference as Variational Policy Gradients. *arXiv* **2019**, arXiv:1907.03876.
127. Saxena, V.; Ba, J.; Hafner, D. Clockwork Variational Autoencoders. *arXiv* **2021**, arXiv:2102.09532.
128. Wu, B.; Nair, S.; Martin-Martin, R.; Fei-Fei, L.; Finn, C. Greedy Hierarchical Variational Autoencoders for Large-Scale Video Prediction. *arXiv* **2021**, arXiv:2103.04174.
129. Tschantz, A.; Baltieri, M.; Seth, A.K.; Buckley, C.L. Scaling Active Inference. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 19–24 July 2020; pp. 1–8. [CrossRef]
130. Kaiser, L.; Babaeizadeh, M.; Milos, P.; Osinski, B.; Campbell, R.H.; Czechowski, K.; Erhan, D.; Finn, C.; Kozakowski, P.; Levine, S.; et al. Model-Based Reinforcement Learning for Atari. *arXiv* **2020**, arXiv:1903.00374.
131. Srinivas, A.; Laskin, M.; Abbeel, P. CURL: Contrastive Unsupervised Representations for Reinforcement Learning. *arXiv* **2020**, arXiv:2004.04136.
132. Pezzulo, G.; Rigoli, F.; Friston, K.J. Hierarchical Active Inference: A Theory of Motivated Control. *Trends Cogn. Sci.* **2018**, *22*, 294–306. [CrossRef] [PubMed]
133. Zakharov, A.; Guo, Q.; Fountas, Z. Variational Predictive Routing with Nested Subjective Timescales. *arXiv* **2021**, arXiv:2110.11236.
134. Wauthier, S.T.; Çatal, O.; De Boom, C.; Verbelen, T.; Dhoedt, B. Sleep: Model Reduction in Deep Active Inference. In *Active Inference*; Verbelen, T., Lanillos, P., Buckley, C.L., De Boom, C., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 72–83.
135. Pezzulo, G.; Rigoli, F.; Friston, K. Active Inference, homeostatic regulation and adaptive behavioural control. *Prog. Neurobiol.* **2015**, *134*, 17–35. [CrossRef] [PubMed]
136. Millidge, B.; Tschantz, A.; Buckley, C.L. Whence the Expected Free Energy? *arXiv* **2020**, arXiv:2004.08128.
137. Andrychowicz, M.; Wolski, F.; Ray, A.; Schneider, J.; Fong, R.; Welinder, P.; McGrew, B.; Tobin, J.; Pieter Abbeel, O.; Zaremba, W. Hindsight Experience Replay. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
138. Warde-Farley, D.; de Wiele, T.V.; Kulkarni, T.D.; Ionescu, C.; Hansen, S.; Mnih, V. Unsupervised Control through Non-Parametric Discriminative Rewards. In Proceedings of the 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, 6–9 May 2019.
139. Mendonca, R.; Rybkin, O.; Daniilidis, K.; Hafner, D.; Pathak, D. Discovering and Achieving Goals via World Models. *arXiv* **2021**, arXiv:2110.09514.
140. Lee, L.; Eysenbach, B.; Parisotto, E.; Xing, E.; Levine, S.; Salakhutdinov, R. Efficient Exploration via State Marginal Matching. *arXiv* **2020**, arXiv:1906.05274.
141. Levine, S. Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review. *arXiv* **2018**, arXiv:1805.00909.
142. Millidge, B.; Tschantz, A.; Seth, A.K.; Buckley, C.L. On the Relationship Between Active Inference and Control as Inference. *arXiv* **2020**, arXiv:2006.12964.
143. Sajid, N.; Ball, P.J.; Parr, T.; Friston, K.J. Active Inference: Demystified and Compared. *Neural Comput.* **2021**, *33*, 674–712. [CrossRef]
144. Clark, J.; Amodei, D. *Faulty Reward Functions in the Wild*; OpenAI: San Francisco, CA, USA, 2016.
145. Ziebart, B.D.; Maas, A.L.; Bagnell, J.A.; Dey, A.K. Maximum entropy inverse reinforcement learning. In Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, Chicago, IL, USA, 13–17 July 2008; Volume 8, pp. 1433–1438.
146. Abbeel, P.; Ng, A.Y. Apprenticeship Learning via Inverse Reinforcement Learning. In Proceedings of the Twenty-First International Conference on Machine Learning, ICML '04, Banff, AB, Canada, 4–8 July 2004; Association for Computing Machinery: New York, NY, USA, 2004; p. 1. [CrossRef]
147. Shyam, P.; Jaśkowski, W.; Gomez, F. Model-Based Active Exploration. *arXiv* **2019**, arXiv:1810.12162.
148. Achiam, J.; Sastry, S. Surprise-Based Intrinsic Motivation for Deep Reinforcement Learning. *arXiv* **2017**, arXiv:1703.01732.
149. Burda, Y.; Edwards, H.; Pathak, D.; Storkey, A.J.; Darrell, T.; Efros, A.A. Large-Scale Study of Curiosity-Driven Learning. In Proceedings of the 7th International Conference on Learning Representations, ICLR, New Orleans, LA, USA, 6–9 May 2019.

150. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347.

151. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; Dy, J., Krause, A., Eds.; PMLR: Brookline, MA, USA, 2018; Volume 80, pp. 1861–1870.

152. Eysenbach, B.; Levine, S. Maximum Entropy RL (Provably) Solves Some Robust RL Problems. *arXiv* **2021**, arXiv:2103.06257.

153. Silver, D.; Hubert, T.; Schrittwieser, J.; Antonoglou, I.; Lai, M.; Guez, A.; Lanctot, M.; Sifre, L.; Kumaran, D.; Graepel, T.; et al. Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. *arXiv* **2017**, arXiv:1712.01815.

154. Maisto, D.; Gregoretti, F.; Friston, K.; Pezzulo, G. Active Tree Search in Large POMDPs. *arXiv* **2021**, arXiv:2103.13860.

155. Clavera, I.; Fu, V.; Abbeel, P. Model-Augmented Actor-Critic: Backpropagating through Paths. *arXiv* **2020**, arXiv:2005.08068.

156. Pardo, F.; Tavakoli, A.; Levdik, V.; Kormushev, P. Time Limits in Reinforcement Learning. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; Dy, J., Krause, A., Eds.; PMLR: Brookline, MA, USA, 2018; Volume 80, pp. 4045–4054.

157. Mhaskar, H.; Liao, Q.; Poggio, T. When and Why Are Deep Networks Better than Shallow Ones? In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI'17, San Francisco, CA, USA, 4–9 February 2017; AAAI Press: Palo Alto, CA, USA, 2017; pp. 2343–2349.

158. Novak, R.; Bahri, Y.; Abolafia, D.A.; Pennington, J.; Sohl-Dickstein, J. Sensitivity and Generalization in Neural Networks: An Empirical Study. *arXiv* **2018**, arXiv:1802.08760.

159. Colbrook, M.J.; Antun, V.; Hansen, A.C. Can stable and accurate neural networks be computed?—On the barriers of deep learning and Smale's 18th problem. *arXiv* **2021**, arXiv:2101.08286.

160. Ben-David, S.; Hrubeš, P.; Moran, S.; Shpilka, A.; Yehudayoff, A. Learnability can be undecidable. *Nat. Mach. Intell.* **2019**, *1*, 44–48. [CrossRef]