entropy

MDPI

Article

Intrinsic Motivation as Constrained Entropy Maximization

Alex B. Kiefer [1,2]

[1] VERSES, Los Angeles, CA 90016, USA; alex.kiefer@monash.edu
[2] Monash Centre for Consciousness and Contemplative Studies, Monash University, Melbourne, VIC 3800, Australia

Abstract: "Intrinsic motivation" refers to the capacity for intelligent systems to be motivated endogenously, i.e., by features of agential architecture itself rather than by learned associations between action and reward. This paper views active inference, empowerment, and other formal accounts of intrinsic motivation as variations on the theme of constrained maximum entropy inference, providing a general perspective on intrinsic motivation complementary to existing frameworks. The connection between free energy and empowerment noted in previous literature is further explored, and it is argued that the maximum-occupancy approach in practice incorporates an implicit model-evidence constraint.

Keywords: intrinsic motivation; active inference; empowerment; entropy

1. Introduction

In psychology, "intrinsic motivation" refers to the tendency for intelligent creatures to be motivated to act even in the absence of externally specified goals or learned reward contingencies [1]. A central thread in accounts of intrinsic motivation—including foundational work linked to "self-determination theory" [2]—focuses on innate capacities for exploration, learning, and growth [3]. Intrinsic motivation in a broader sense, however, subsumes these, as well as "built-in" mechanisms favoring survival and well-being.

The paradigm of intrinsic motivation has increasingly gained traction in modern machine learning, where it is operationalized as the idea that policies for action may be optimized based on structural features of agents and agent–environment interactions, in contrast to traditional approaches like reinforcement learning, in which policies are optimized based on ad hoc reward functions.

An early, and increasingly influential, formal account of intrinsic motivation in machine learning is based on *empowerment*, defined as the capacity of the channel linking agents' actions (actuator states) to sensory feedback (observations) [4,5]. This information-theoretic construct is loosely related to the broader notion of empowerment in psychology, as the capacity for autonomous, contextually impactful action [6]. One interpretation of the empowerment objective is that empowered agents "keep their options open", as wide action-conditioned channel capacity entails that agents are able to realize a variety of states (for which observations are a proxy).

The *active inference* framework [7] shares similar motivations, and provides a Bayesian method for combining a general form of intrinsic motivation (i.e., curiosity or "epistemic drive") [8] with agent-specific prior distributions over states or outcomes [9], which model homeostatic set points and can function like explicit rewards. The key idea is that agents minimize variational free energy (VFE), a proxy for surprise, where the expected (variational) free energy (EFE), as discussed below, guides policy selection by supplying an empirical prior over policies, given observations.

Academic Editor: Danny JJ Wang

Received: 8 February 2025
Revised: 15 March 2025
Accepted: 26 March 2025
Published: 31 March 2025

Citation: Kiefer, A.B. Intrinsic Motivation as Constrained Entropy Maximization. *Entropy* **2025**, *27*, 372. https://doi.org/10.3390/e27040372

More recently, the objective of *maximum path occupancy* has been proposed as a framework for intrinsic motivation [10]. On this account, agents are motivated to maximize future action–state path occupancy, which can be measured in terms of both the entropy of the action distribution and the entropy of the ensuing state distribution, given an initial state. This somewhat more radical perspective explicitly inverts the perhaps natural assumption that drives for exploration and curiosity have evolved as a means to achieving rewarding states (such as the consumption of food), and, in effect, views the latter as instrumentally valuable in enabling future exploration, i.e., avoiding absorbing states that afford little or no action variability (e.g., death).

There are many other formal treatments of intrinsic motivation in the literature on machine learning, some closely related to those just discussed, such as pioneering work on artificial curiosity (see e.g., [11,12]) and treatments in terms of Bayesian surprise [13,14]. More broadly, "intrinsic motivation" as defined here encompasses many instances of behavior in the absence of explicit reward that can be explained in terms of intrinsic cybernetic drives or closed-loop control systems, e.g., as in perceptual control theory (PCT) [15], which describes behaviors that function to regulate the variability of sensory input [16–18]. These accounts are closely related to both empowerment and active inference and are touched on below. That said, the focus here is mainly on the relationship between active inference and empowerment and on the relationship of both to maximum occupancy, which has recently been proposed explicitly as an alternative to these.

While ref. [19] conducts a comparative empirical study of these three frameworks for intrinsic motivation on a toy problem and ref. [20] considers how active inference may be formally related to broader schemes for intrinsic motivation, comparatively little work exists on the formal and conceptual relations among these frameworks. Here, I highlight the fact that all three can be understood as variations on the theme of constrained entropy maximization, a principle with deep connections to the free energy principle and active inference [21]. I explore the connection between empowerment and active inference [22] by casting the empowerment objective itself explicitly as a form of variational inference. I also argue that the ability of occupancy-maximizing agents to exhibit apparently goal-directed behavior depends on a "survival instinct" or model-evidence constraint implicit in the factorization of the overall system into actions and states. These considerations frame entropy maximization under local constraints as the kernel of intelligence and agency, with particular facets of this process such as empowerment, perception, curiosity, and the "will to live" as corollaries.

Section 2 below unpacks the three frameworks for intrinsic motivation mentioned above (empowerment, active inference, and maximum occupancy) in some detail, both formally and in terms of conceptual motivation, and articulates their ties to constrained entropy maximization. Section 3 looks closely at some connections among these theories, then distills a few general conclusions.

2. Three Formal Accounts of Intrinsic Motivation

2.1. Empowerment

The *empowerment* objective for intrinsic motivation, originally proposed in [4], is defined as the capacity of the information channel linking an agent's actions to its observation of the effects of those actions. That is, given a space of possible observations O_T at future timestep T and a sequence of actions $A_{t:T}$ from the present timestep t to the future, there is some distribution $P(O_T|A_{t:T})$ capturing the probabilistic dependence of the future observation on the actions taken, and the empowerment $\mathcal{E}$ of an agent is measured as the capacity C of the information transmission channel defined by this distribution:

$$\mathcal{E}_t = C\Big(P(O_T|A_{t:T})\Big) = \max_{P(A)} I(A_{t:T}; O_T) \tag{1}$$

In this case, C is defined as the maximum mutual information between actions and future observations, $I(A_{t:T;O_T})$, when the conditional distribution $P(O_T|A_{t:T})$ is held fixed and the distribution over actions, $P(A)$, is allowed to vary.

It is worth taking a moment to unpack this, as a detailed understanding will be useful for comparisons below. The mutual information is standardly defined, for two random variables X and Y, as a double sum equivalent to the KL divergence from the joint density $P(X, Y)$ to the product of the marginals over X and Y:

$$\begin{aligned} I &= \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P(x,y) \log\Big(\frac{P(x,y)}{P(x)P(y)}\Big) \\ &= D_{KL}\Big(P(X,Y)||P(X)P(Y)\Big) \end{aligned} \tag{2}$$

Intuitively, this expression measures how different the actual joint distribution is from what it would be were the two variables independent, i.e., how much information the variables carry about one another. While this measure is symmetric (i.e., the same for X and Y), it can be broken down in terms of conditional probabilities in either direction. Since the joint density can be factorized into a prior and a conditional density, i.e., $P(X, Y) = P(X)P(Y|X) = P(Y)P(X|Y)$, the mutual information can also be expressed as an expected KL divergence from a conditional density $P(Y|X)$ to the marginal over Y:

$$\begin{aligned} I &= \sum_{y\in\mathcal{Y}} \sum_{x\in\mathcal{X}} P(x)P(y|x) \log\Big(\frac{P(x)P(y|x)}{P(x)P(y)}\Big) && \text{Factorize joint distribution} \\ &= \sum_{x\in\mathcal{X}} P(x) \Bigg[\sum_{y\in\mathcal{Y}} P(y|x) \log\Big(\frac{P(y|x)}{P(y)}\Big)\Bigg] && \text{Cancel out } P(x)\text{s and rearrange} \\ &= \mathbb{E}_{P(X)}\Bigg[D_{KL}\Big(P(Y|X)||P(Y)\Big)\Bigg] \end{aligned} \tag{3}$$

Given a fixed $P(Y|X)$ (channel), the *channel capacity* $C\Big[P(Y|X)\Big]$ is then the maximum value that this mutual information can take, given a free choice of $P(X)$.

The empowerment objective is just this channel capacity, with respect to the channel linking actions over timesteps $t \ldots T$ with observations at T (for simplicity, the discussion here focuses on the original formulation in [4], but obviously many variations on this theme are possible, e.g., using different time indices (as explored in [23]) or swapping out observations for latent states). Intuitively, the mutual information term (i.e., information gain expected under the action distribution) measures both the *controllability* of outcomes (the influence of action selection on such outcomes) and the *variety* of achievable outcomes (i.e., "keeping one's options open") [4]. This combination of controllability and variety is characteristic of constrained entropy maximization, a common theme in many frameworks for intrinsic motivation [10,23,24], and is related to Ashby's "law of requisite variety" [25].

The "variety" aspect of empowerment can be made more explicit by considering the relation of mutual information to entropy. Any mutual information $I(X; Y)$ can be expressed in terms of entropy in several ways:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned} \tag{4}$$

Thus, empowerment can be seen as maximizing the entropy of the action distribution $H\big(P(A)\big)$ while ensuring that actions are "rational" in the sense of being reliably related to observations, i.e., minimizing $H\big(P(A|O)\big)$. At the same time, it can be viewed as maximizing the variety of observations O while ensuring that they remain controllable, i.e., minimizing $H\big(P(O|A)\big)$. Here, the entropy-*minimizing* terms may be read as "energetic" (negative log probability) constraints by analogy to thermodynamics, i.e., departures from maximum-entropy inference that encode prior knowledge in Jaynes's framework [26].

The empowerment objective may be read as a signal guiding model *evolution* or selection, as in the work just cited (i.e., choosing a generative model of actions and outcomes $P(A)P(O|A)$). Given a fixed model, agents may also choose policies (actions) so as to maximize the time-dependent empowerment $\mathcal{E}_t$ by seeking the position in the state space of the overall system (where external states are implicitly represented here by observations) in which the channel capacity is highest, since $P(O_T|A_{t:T})$ implicitly depends on the states at $t \dots T$.

Before moving on to consider other treatments of intrinsic motivation, we note that in ref. [23] it is shown (in the setting of continuous state spaces) that generalizing the empowerment objective just discussed, by varying the length of action and observation sequences and the time interval separating actions from target observations, allows one to recover various extant descriptions of control in dynamical systems. Saliently for present purposes, a generalized empowerment objective in which actions are taken only at the first time-step corresponds to a "kicked" (controlled) version of Causal Entropic Forcing [24], a framework that models intelligent behavior in terms of entropy maximization.

2.2. Active Inference and Expected Free Energy

Advances in cognitive (neuro)science over the past decade or so have seen the rise to prominence of the idea that most (if not all) intelligent action can be understood in terms of Bayesian inference [27]. This paradigm encompasses quite specific models of neuronal information processing such as predictive coding [28,29], which has been invoked to explain perceptual inference [30], as well as more abstract and general frameworks, mostly saliently the free energy principle [31,32], an account of self-organization in terms of variational Bayesian inference, and active inference [7,33], which derives a scheme for action (i.e., planning or policy selection) from the assumption that agents select actions that are expected to minimize variational free energy in the future. Accounts of motor control in terms of high-precision kinesthetic predictions [34] are closely related to active inference, but here the latter term is reserved to denote the idea that policies for action are selected on the basis of expected (variational) free energy.

Here, the variational free energy $\mathbf{F}$ is an upper bound on the surprise $-\ln P(o|m)$ of sensory observations o relative to a model m (or, equivalently, the negative free energy is a lower bound on the marginal likelihood or Bayesian model evidence):

$$\mathbf{F} = \underbrace{\mathbb{E}_{Q(s)}\big[-\log P(s,o)\big]}_{\text{Energy}} - \underbrace{H\big[Q(s)\big]}_{\text{Entropy}} \tag{5}$$

$$= \underbrace{-\ln P(o)}_{\text{Surprise}} + \underbrace{D_{KL}\Big(P(s)||Q(s)\Big)}_{\text{Divergence}} \tag{6}$$

Perceptual inference (i.e., inference of the latent states s that cause observations) can be performed in this scheme by optimizing the variational posterior $Q(s)$ so as to minimize the free energy functional $\mathbf{F}$:

$$Q^*(s) = \underset{Q(s)}{\operatorname{argmin}}\, \mathbf{F} \tag{7}$$

In this inferential process, the entropy of the posterior is maximized under the energy (model evidence) constraint, in accordance with the principle of constrained maximum-entropy inference [21,26,31,35].

Agents governed by active inference build on this form of perceptual inference to implement a specific form of planning as inference [36], "reasoning backward" from preferred outcomes (cast in this context as observations that furnish evidence for a prior generative model [37]) to the policies (i.e., possible sequences of discrete actions or control states $[u_0, \ldots, u_T]$ up to some finite planning horizon T) most likely to bring them about. In brief, this involves inferring a (variational) posterior distribution $Q(\pi)$ over policies π in which the probability assigned to each policy is proportional to its associated (negative) expected free energy (EFE, denoted $\mathbf{G}$ in equations).

The EFE associated with a policy $\mathbf{G}_\pi$ is defined as the variational free energy that the agent expects to accumulate under that policy [8]. This expectation is taken by conditioning on the variational state posterior $Q(s_0)$, rolling the generative model out into the future using the transition model $P(s_{t+1}|s_t, u)$ to obtain a policy-conditioned variational prior $Q(s_t|\pi)$ over states, and then using the likelihood $P(o|s)$ to predict observations:

$$Q(o_t|\pi) = P(o_t|s_t)Q(s_t|\pi) \tag{8}$$

$$= \underbrace{P(o_t|s_t)}_{\text{Likelihood}} \underbrace{Q(s_0) \prod_{\tau=1}^{t} P(s_\tau|s_{\tau-1}, u_\tau)}_{\text{Variational prior}} \tag{9}$$

Thanks to the conditional independence relations in this factorization, $\mathbf{G}_\pi$ can be computed as a sum $\sum_{t=0}^{T} \mathbf{G}_\pi^t$ over timestep-specific terms $\mathbf{G}_\pi^t$:

$$\mathbf{G}_\pi^t = \underbrace{\mathbb{E}_{Q(s_t,o_t|\pi)}\Big[\log Q(s_t|\pi) - \log P(s_t, o_t|\pi)\Big]}_{\text{Policy-conditioned VFE}}$$

$$\approx \mathbb{E}_{Q(s_t,o_t|\pi)}\Big[\log Q(s_t|\pi) - \log Q(s_t|o_t, \pi) - \log P(o_t)\Big]$$

$$= \underbrace{-\mathbb{E}_{Q(s_t,o_t|\pi)}\Big[-\log P(o_t)\Big]}_{\text{Expected utility}} \underbrace{- D_{KL}\Big[Q(s_t|o_t, \pi)||Q(s_t|\pi)\Big]}_{\text{Information gain}} \tag{10}$$

$$= \underbrace{D_{KL}\Big(Q(o_t|\pi)||P(o_t)\Big)}_{\text{Risk}} + \underbrace{\mathbb{E}_{Q(s_t|\pi)}\Big[H(P(o_t|s_t))\Big]}_{\text{Ambiguity}} \tag{11}$$

Given the posterior over policies, actions or control states u are sampled at each timestep based on a Bayesian model average of the policies:

$$Q(\pi) = \sigma\Big(-\mathbf{G}\Big) \qquad \text{Posterior over policies} \tag{12}$$

$$u_t \sim Q(u_t) = \sum_\pi P(u_t|\pi)Q(\pi) \qquad \text{Control state sampled from marginal} \tag{13}$$

where $P(u_t = u_i|\pi_j)$ is a mapping whose value is 1 if policy j begins with control state u_i and 0 otherwise, $\mathbf{G}$ is the vector of EFE terms per policy, and σ is the softmax function.

In accord with the EFE's role as a prior belief about which policies should be (i.e., will be, in a planning-as-inference scheme) pursued, the generative model that figures in policy inference includes a state-independent distribution $P(o)$ over outcomes (observations o) encoding what the agent "prefers" to see. While $P(o)$ is sometimes specified independently of the predictive (generative) model of the world as an ad hoc "preference" or reward distribution [20,38], from a more principled perspective it may be cast as the marginal

likelihood of observations, and models the characteristic attracting set of states that homeostatic systems must remain within in order to persist [31]. In ref. [9] for example, an EFE objective is formulated solely in terms of the difference between $P(o, s)$ and $P(o, s|a)$, i.e., the preference model is the same generative model used for prediction, with actions marginalized out.

$P(o)$ can be thought of as a kind of intrinsic motivation in the broader sense since it is "built in" to the agent rather than learned (though it is worth noting that the distinction between intrinsic and extrinsic motivation would then cross-cut that between intrinsic (i.e., epistemically-based) and extrinsic (utility-based) *reward* or value commonly used in the active inference literature [8] and elsewhere). Crucially, however, the EFE also implements the model-independent inductive bias that actions will minimize variational free energy in the future and, thus, subserves a more general form of intrinsic motivation.

In Equation (11) (see [33], Appendix A for a derivation), *risk* is simply a measure of expected negative reward, which in this context measures how different predicted outcomes are from those expected *a priori* (i.e., preferred). The entropy of the likelihood mapping from states to observations expected under a given policy ("Ambiguity") quantifies how uncertain the agent will be about outcomes if that policy is pursued. Thus, minimizing expected free energy (Equations (12) and (13)) encourages agents to choose policies (actions) that render outcomes predictable, subject to the constraint that risk is minimized. Interestingly, this divergence minimization objective [39] recapitulates core insights of perceptual control theory [15] and of cybernetic approaches to intelligence more broadly, while recasting these insights in the terms of information theory [40], thus providing a link to statistical physics (cf. the inversion of perceptual generative models as constrained maximum-entropy inference).

It will be useful in what follows to consider one more breakdown of the EFE that is not usually discussed in the literature. The EFE at time t can be written as a Helmholtz free energy, showing that at each timestep the policy-conditioned empirical prior over states $Q(s_t|\pi)$) maximizes the entropy of the state distribution, under an evidence constraint:

$$\mathbf{G}_\pi^t = \underbrace{\mathbb{E}_{Q(s_t, o_t|\pi)}\Big[-\log P(s_t, o_t|\pi)\Big]}_{\text{Energy}} \underbrace{- H\Big(Q(s_t|\pi)\Big)}_{\text{Entropy}} \tag{14}$$

The preceding exposition omits several features of active inference models that are inessential for present purposes, such as the baseline policy or "habit" prior and the temperature parameter used in action selection, as well as the variational free energy of policies, which may be combined with the EFE in policy inference (please see [7,33,41] for further details). Perhaps more importantly, I focus here on the most well-known variant of active inference which appeals to EFE. Work using alternative an "generalized free energy" objective [42], touched on below, shows how similar results can be achieved by minimizing VFE in a model that treats future observations as latent states.

2.3. Maximum Occupancy

The Maximum Occupancy Principle (MOP) [10] carries the theme of intrinsic motivation to its logical conclusion, proposing that a traditional picture of rational agency, in which curiosity and other intrinsic drives have evolved in order to serve reward maximization, should be inverted: we can instead understand rewarding states as a means to the end of continuing to live, i.e., to explore (thus maximally occupy) the action–state path space.

Formally, the occupancy objective is defined in terms of a state-conditioned policy distribution $\pi(A|S)$ and transition dynamics $P(S'|S, A)$, which can be alternately sampled

from to generate action–state paths τ. The reward function $R(\tau)$ for a given trajectory is then specified as:

$$R(\tau) = -\sum_{t=0}^{\infty} \gamma^t \log \left[\pi^{\alpha}(a_t|s_t) P^{\beta}(s_{t+1}|s_t, a_t) \right] \tag{15}$$

where γ^t is the standard temporal reward discount in reinforcement learning and α and β are weights modulating the influence of action and state path occupancy. Agents select policies so as to maximize the expected reward or "value" of states s, $V_{\pi}(s)$:

$$\begin{aligned} V_{\pi}(s) &= \mathbb{E}_{\pi(A|S)P(S'|S,A)} \left[R(\tau)|s_0 = s \right] \\ &= \mathbb{E}_{\pi(A|S)P(S'|S,A)} \left[\sum_{t=0}^{\infty} \gamma^t \Big(\alpha H(A|s_t) + \beta H(S'|s_t, a_t) \Big) |s_0 = s \right] \end{aligned} \tag{16}$$

Here, the realization of τ depends on the initial state s, and $H(A|s_t)$ and $H(S'|s_t, a_t)$ denote the conditional entropy of the action distribution given the current state, and of the distribution of the next state given the current state and action, respectively. Thus, agents that maximize $V_{\pi}(s)$ maximize an expectation over the (summed step-wise conditional) entropy of both action and state paths, subject to the weights and initial condition.

In [19], empirical studies are presented in which MOP agents aggressively explore state and action space while still exhibiting apparently goal-directed behavior. The former is perhaps to be expected, given the purely intrinsic, surprisal-maximizing reward function, thanks to which agents will directly seek out improbable actions that lead to improbable states. Presumably, the ability of MOP agents to behave in goal-oriented ways despite the absence of explicit tasks, rewards, or even preference distributions is underwritten by the imperative to maximize longer-term path occupancy, which balances the tendency to greedily maximize entropy at each timestep. This implicit constraint on short-term entropy maximization, in the service of increasing entropy in the long run, is evocative of the argument in [43] according to which the structured, relatively low-entropy states characteristic of complex forms of life are favored for their ability to accelerate the dissipation of free energy within the broader universe. Notably, in [44] similar emergent task-oriented behavior is demonstrated in agents governed by an empowerment objective in the presence of absorbing states.

3. A Unified View of Intrinsic Motivation

This section begins by analyzing the relationship between active inference and empowerment, then considers the maximum-occupancy perspective in relation to both of these. It then concludes with a discussion of some themes common across these frameworks, and a synthesis that allows us to resolve some apparent dichotomies from a multi-scale or scale-free perspective.

3.1. Empowerment and Active Inference

Maximizing the empowerment objective is closely related to minimizing expected free energy. Most straightforwardly, in the absence of a constraint (expected utility term), the expected free energy described above reduces to the negative information gain $D_{KL}\Big(Q(s_t|o_t, \pi)||Q(s_t|\pi)\Big)$, so that minimizing EFE maximizes the mutual information between states and observations [45]. Although this mutual information can be interpreted as maximizing the entropy of observations (constrained by their controllability), optimization of Q in variational inference is always constrained by the generative model P so there is

no conflict with the imperative to maximize model evidence. That said, in a multi-scale setting, one may also consider learning the parameters of P as discussed below.

While the original empowerment objective [4] leaves the mediation of the action-sensation channel $P(O_T|A_{t:T})$ by hidden states implicit, the active inference objective makes this explicit: in choosing actions, agents effectively choose the transition dynamics for controllable states (in typical implementations, discrete actions index slices of transition tensors), such that they are rendered informative about observations. Thus, effectively, states are (probabilistically) chosen so as to maximize the mutual information between actions and observations, as in the empowerment objective.

In [22] (Appendix), it is claimed that "empowerment is a special case of active inference, when we can ignore risk (i.e., when all policies are equally risky)". We can run a similar argument by considering the empowerment objective described in [4] as part of a variational inference process. In terms of the notation used for active inference, the goal would be to maximize $I_t(\pi; o_T)$, where, as above, π is a sequence of control states $[u_0, u_1, \ldots, u_T]$. This objective can be expressed in terms of the entropies of posterior observation and policy distributions, and also as a KL divergence:

$$\begin{aligned} I_t(\pi; o_T) &= H\Big(Q(\pi)\Big) - H\Big(Q(o_T|\pi)\Big) \\ &= D_{KL}\Big(Q(\pi, o_T)||Q(\pi)Q(o_T)\Big) \end{aligned} \tag{17}$$

The divergence simply states that agents maximizing empowerment should select policies that provide information about the target observation, which in this context amounts to the former affording control over the latter. The subscript in I_t indicates that, like the original empowerment objective $\mathcal{E}_t$, this term is implicitly time-dependent. More specifically, in the present setting, the variational posteriors Q at t depend on the observation o_t via the state posterior $Q(s_t)$.

Interestingly, defining a conditional "energy" term as the negative log probability of the observation at T given policies, the expression of the mutual information in terms of entropies can be written in a form analogous to a free energy $F_t(\pi, o_T)$ simply by flipping the sign and rearranging terms:

$$F_t(\pi, o_T) = -\Big[H\Big(Q(\pi)\Big) - H\Big(Q(o_T|\pi)\Big)\Big] \tag{18}$$

$$= \underbrace{\mathbb{E}_{Q(o_T|\pi)}\Big[-\log Q(o_T|\pi)\Big]}_{\text{"Energy"}} - \underbrace{H\Big(Q(\pi)\Big)}_{\text{Entropy}}$$

$$= \underbrace{\mathbb{E}_{Q(o_T|\pi)}\Big[-\log P(o_T|s_T)\Big]}_{\text{Energy of final observation}} + \underbrace{H\Big(Q(s_T|\pi)\Big)}_{\text{Conditional state entropy}} - \underbrace{H\Big(Q(\pi)\Big)}_{\text{Policy entropy}} \tag{19}$$

$$= \underbrace{\mathbb{E}_{Q(s_T|\pi)}\Big[H\Big(P(o_T|s_T)\Big)\Big]}_{\text{Ambiguity}} - \underbrace{I_t(\pi; s_T)}_{\text{State information gain}} \tag{20}$$

$$Q(s_T|\pi) = \sum_{s_0 \in S} \cdots \sum_{s_{T-1} \in S} \Big[\prod_{t=0}^{T-1} P(s_{t+1}|s_t, \pi) Q(s_0)\Big] \tag{21}$$

Maximizing $I_t(\pi; o_T)$ is then equivalent to minimizing this energy. The second line lacks the form of a proper (variational) free energy because the "energy" term is just the entropy of a variational density $Q(o_T|\pi)$, rather than a joint probability (generative model) $P(o, s)$. However, $Q(o_T|\pi)$ factors into several terms some of which are distributions of the generative model. Taking this into account, we arrive at the expression in the penultimate row, which is similar to a Helmholtz free energy with an additional entropy term to be *min-*

imized: under this objective, agents will seek low-energy (predictable) observations, while maximizing the entropy of policies ("keeping options open") and also seeking policies that minimize the entropy of the final state, i.e., seeking paths that result in controllable states.

Finally (last line), the expected energy (negative log probability under the generative model) of o_T is equivalent to the ambiguity term in the EFE mentioned above (with respect to the final observation in a trajectory), while the two entropy terms can be combined into a state information gain term (in ref. [22], a similar formulation of empowerment in terms of free energy is reached by considering action–state empowerment in the context of the generalized free energy functional [42]). Thus, from the empowerment objective alone (and ignoring additional "preference" constraints), we can derive drives for both *epistemic value* (minimizing ambiguity) and *control* (maximizing state information gain).

Active inference agents are thus "empowered" in that they maximize the entropy of future state distributions, under the constraint that these states or the ensuing observations be controllable. Crucially, in active inference, agents are also constrained to maximize *model evidence* (or its tractable lower bound, variational free energy) [37]. In fact, the latter (approximately maximizing model evidence) is the central concept in the FEP and active inference, where (constrained) entropy maximization falls out of variational free energy minimization, and specifically exploratory behavior emerges thanks to the distribution-matching (KL-divergence) term in the EFE objective [39]. It should be stressed that the addition of preference-maximizing constraints to empowerment in active inference can be expected to yield significantly different behavior: to the extent that one commits to a specific goal, one ceases to "keep options open".

3.2. Constrained Maximum Occupancy

Prima facie, it is difficult to square the maximum occupancy objective with those just considered in precise terms, since its objective involves only maximizing (expected) entropy, without constraints. In fact, the MOP objective described above is general enough to encode an approximation to empowerment, if the β term is set to a negative value [10], which encourages agents to choose actions that *minimize* the entropy of the state transition distribution, while still maximizing the entropy of the distribution over actions. This is clearly closely related to the empowerment objectives discussed above once the distinction between hidden states and observations is accommodated (i.e., it results in agents that "keep options open", ensuring a variety of controllable states and thus observations). However, while of practical interest, this really amounts to a departure from the spirit of MOP.

It is argued in ref. [10], on both conceptual and experimental grounds, that MOP agents exhibit more robust exploratory behavior and variety in policy selection than agents governed by empowerment or EFE objectives. The experiments reported in that work and in ref. [19], however, involve full observation of the state space, so that the ambiguity term in the EFE does no work (and more generally, the usual motivations for the FEP and active inference, in which agents are assumed to infer unknown states of the environment, do not apply). Moreover, the experiments reported in ref. [10] use a setting of $\beta = 0$ by default, thus effectively maximizing the entropy of only the action distribution. For these reasons, the ensuing discussion focuses on the conceptual arguments surrounding entropy maximization and the role of constraints rather than on these experimental results.

On conceptual grounds, the MOP objective may (it is argued in [10]) be expected to produce a greater variety of actions than active inference for two reasons: (a) the EFE objective contains an explicit "preference" term which MOP lacks, and which biases action in favor of certain outcomes (thus reducing the entropy of action-state paths); and (b) while the EFE objective maximizes the entropy of the state-transition distribution at each timestep (cf. Equation (14)), it contains no term to maximize the entropy of the action distribution.

The maximization of action (policy) entropy does seem to fall out of the empowerment framework. Thus, given the equivalences outlined above, the same should be true of active inference. The authors of [10] argue that the EFE deterministically selects a single policy. However, in the context of a full variational inference treatment (i.e., planning-as-inference), the entropy of the *policy* distribution should also be maximized (under relevant constraints).

Conceptually, π is a latent variable, and *ceteris paribus* its entropy should be maximized during variational inference just as the entropy of $Q(s)$, the variational density over hidden causes, is maximized. This is captured formally in work on active inference exploring a formulation of expected variational free energy that is, in some ways, more parsimonious than the EFE, called the *generalized free energy* [42]. As shown in [22], this objective can (as is usual in variational inference) be written as a Helmholtz free energy, where in this case, the energy term is the expected EFE under the policy posterior, and the entropy of the policy distribution is explicitly maximized as free energy is minimized:

$$\underbrace{F\Big[Q(s,\pi)\Big]}_{\text{Generalized free energy}} = \underbrace{\mathbb{E}_{Q(\pi)}\Big[\mathbf{G}_\pi\Big]}_{\text{Expected EFE}} - \underbrace{H\Big(Q(\pi)\Big)}_{\text{Policy entropy}} \tag{22}$$

It is the generalized free energy which in [22] is shown to be equivalent to empowerment constrained by risk. Relatedly, the "free energy of empowerment" $F_t(\pi, o_T)$ defined above also contains this policy entropy term. Thus, the main difference between active inference (viewed broadly so as to include maximum-entropy policy inference) and MOP seems to be the presence or absence of explicit model evidence constraints.

The core concept in MOP is that maximizing path occupancy is an "intrinsic" value, from which reward is derivative. The core claim of the FEP and active inference (which we have seen to entail empowerment) is that maximizing *model evidence* is an "intrinsic" value, and that rewards, as well as information-seeking behavior, derive from this imperative. At first glance, these frameworks may appear difficult to reconcile since the former maximizes surprisal while the latter minimizes it (at least with respect to sensory observations).

One of the central claims of [10] is that intelligent, goal-directed action emerges naturally from the MOP objective, in the presence of absorbing states together with the means to (foreseeably) avoid them given certain courses of action. It may be wondered whether pure MOP agents would be as successful in less predictable environments in which risk-aversion may be more important, but independently of this, there are deep reasons to suppose that MOP agents would not produce richly intelligent behavior without an implicit model-evidence constraint.

Occupancy-maximizing agents seek control only in order to remain alive, a goal that is argued to flow elegantly from the desire to maximize entropy in the distant future. However, this argument assumes that being dead corresponds to an "absorbing" state, which in the experiments is modeled as entailing zero entropy for the rest of time. In a more physically realistic model, dying would correspond to a breakdown of the agent–environment boundary, and so to a state that is much *higher* in entropy, even if it is lower in *agent-caused* entropy (with the dissolution of individual agents corresponding to an unconstrained maximum-entropy state, or in physical terms, thermal equilibrium). Relatedly, the "survival instinct" is encoded in active inference agents in the fact that departures from homeostatic set points (defined by the generative model or "preference distribution") score high in free energy and so are aversive.

Thus, identifying a lack of action availability with a low-entropy state is plausible only in toy scenarios in which the entropy increase induced within the overall system by the dissolution of the agent is ignored. Other things being equal, death ought to be *attractive* to MOP agents unless they possess an *a priori* distinction between agent and environment,

i.e., a "sense of self". The upshot is that the implicit constraint enabling the emergence of goal-directed behavior in such agents is, in the general case, not simply long-term entropy maximization but also the existence of an agent with a repertoire of actions, encoded in the very partitioning of the space into action and state variables. Effectively, this amounts to a version of the "controllability" constraints that appear explicitly in active inference and empowerment, as the agent must exert control sufficient to enable homeostasis (i.e., the maintenance of internal states against dissipative forces). We may note that the ability to predict the entropy of future states so as to compute state value—however this is implemented—also corresponds to some local disequilibrium and thus imposes a de facto constraint on the entropy of the agent's internal states.

3.3. Model Evidence and the Will to Live

Despite the arguments just given, the inversion of traditional assumptions about the relationship between exploration and reward highlighted by the MOP is appealing, as entropy maximization (albeit under constraints) appears to be an essential feature of intelligence and life [21,24,46,47], more constant across distinct forms of life than any particular reward-seeking behavior. The idea that future path occupancy, as measured by entropy [10], is tantamount to remaining alive is one way of understanding the place of entropy maximization at the heart of accounts of intrinsic motivation.

We have seen, however, that in order to reproduce the goal-oriented behavior characteristic of complex biological intelligence, it is necessary to maximize entropy under the constraint that the agent's existence, operationalized as conditional independence between internal and external states [31] (which appears in simple models as an action–state partition), is maintained. Taking a page from Schopenhauer [48], intrinsic motivation may then be cast as simply the "will to live", i.e., to persist as a living (moving, changing) thing, a basal impulse that takes different particular forms depending on local constraints (generative models). These constraints shape the primary motivational force of entropy production, such that conditional independence structures are maintained.

In simpler models of intelligence, the relevant partitioning of the entire (agent–environment) system is assumed to be fixed, but in more sophisticated treatments such as multi-scale or scale-free active inference [49,50], model structure itself may evolve, typically at slower timescales. We may then view the life of an agent at any given instant as seeking not only observational evidence for the currently parameterized model, but also evidence for the parameters themselves, as well as for hyperparameters (or priors over parameters, including structural priors). This structural evolution can be understood in terms of Bayesian model selection [45].

From this perspective, there is no deep contradiction between scale-free self-evidencing (i.e., the seeking of model evidence) [37] and maximum occupancy. Once constraints (parameters and model structure) are themselves treated as random variables, the process of self-evidencing is seen to be data- or observation-driven through and through, and it appears to be a property of our universe (insofar as it is accurately modeled as a closed system) that the entropy of data-generating processes as a whole can only increase. From this perspective, maximum-entropy inference is a ubiquitous self-fulfilling prophecy in virtue of which the universe evolves toward thermal equilibrium (the conceptual distinction between thermodynamic and merely information-theoretic or variational free energy [35,51] need not concern us here, as the entropy of observations is sufficient to drive this process). Thus, all agents indeed maximize occupancy on the longest timescale, though in a rather selfless way, i.e., they gather evidence for a maximum-entropy model of the universe at large, in which boundaries between agents (Markov blankets) and their corresponding energetic constraints have disappeared.

The idea that entropy is maximized "for its own sake" does not, of course, preclude interpretations of this phenomenon in terms of epistemic value [8], curiosity [12], and so on, in various contexts. What the preceding discussion does suggest is that exploratory behavior is by no means "merely" an evolved mechanism for securing outcomes high in utility, but is at least as fundamental an aspect of agency as the latter tendency and plays an irreducible role in psychological functioning [52], with the two plausibly participating in a dance of circular causality. The presence of both goal-seeking and information-seeking drives in the expected free energy functional, regardless of the particular generative model, points to this same conclusion [53].

4. Conclusions

Seeking common themes across contemporary accounts of intrinsic motivation has surfaced the inevitability of constrained entropy maximization as a core principle describing motivation in biological systems. This insight is hardly novel at a fundamental level, as entropy maximization has long been recognized as a crucial principle both in physics generally [26] and for the physics of life and intelligence specifically [24,31,43], and has played an explicit role in several accounts of intrinsic motivation [12,23]. Here, the goal has been primarily to explore in detail how three accounts of intrinsic motivation that have previously been juxtaposed in the literature [19] may nonetheless be understood as variants of this general perspective.

Funding: This research is supported by VERSES.

Data Availability Statement: Data is contained within the article.

Acknowledgments: The author would like to thank, in particular, Karl Friston, Jacqueline Hynes, and Dalton Sakthivadivel for conversations directly relevant to this work, as well as Mahault Albarracin, Riddhi J. Pitliya, Maxwell Ramstead, Tim Verbelen, and Ran Wei for related discussions. Thanks also to three anonymous peer reviewers for very useful comments.

Conflicts of Interest: The author declares that this study received funding from the company VERSES, where the author was employed. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Domenico, S.I.D.; Ryan, R.M. The Emerging Neuroscience of Intrinsic Motivation: A New Frontier in Self-Determination Research. *Front. Hum. Neurosci.* **2017**, *11*, 145.
2. Ryan, R.M.; Deci, E.L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* **2000**, *55*, 68–78.
3. Oudeyer, P.Y.; Gottlieb, J.; Lopes, M.C. Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Prog. Brain Res.* **2016**, *229*, 257–284. [PubMed]
4. Klyubin, A.S.; Polani, D.; Nehaniv, C.L. Empowerment: A universal agent-centric measure of control. In Proceedings of the 2005 IEEE Congress on Evolutionary Computation, Edinburgh, UK, 2–5 September 2005; Volume 1, pp. 128–135.
5. Salge, C.; Glackin, C.; Polani, D. Empowerment—An Introduction. *arXiv* **2013**, arXiv:1310.1863.
6. Rappaport, J. Terms of empowerment/exemplars of prevention: Toward a theory for community psychology. *Am. J. Community Psychol.* **1987**, *15*, 121–148.
7. Friston, K.J.; FitzGerald, T.H.B.; Rigoli, F.; Schwartenbeck, P.; Pezzulo, G. Active Inference: A Process Theory. *Neural Comput.* **2017**, *29*, 1–49.
8. Friston, K.J.; Rigoli, F.; Ognibene, D.; Mathys, C.D.; FitzGerald, T.H.B.; Pezzulo, G. Active inference and epistemic value. *Cogn. Neurosci.* **2015**, *6*, 187–214.
9. Da Costa, L.; Tenka, S.; Zhao, D.; Sajid, N. Active Inference as a Model of Agency. *arXiv* **2024**, arXiv:2401.12917.
10. Ramirez-Ruiz, J.; Grytskyy, D.; Mastrogiuseppe, C.; Habib, Y.; Moreno-Bote, R. Complex behavior from intrinsic motivation to occupy future action-state path space. *Nat. Commun.* **2024**, *15*, 6368. [CrossRef]

11. Schmidhuber, J. *Adaptive Confidence and Adaptive Curiosity*; Technical Report FKI-149-91; TU Munich: München, Germany 1991; pp. 1–9.
12. Schmidhuber, J. Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010). *IEEE Trans. Auton. Ment. Dev.* **2010**, *2*, 230–247.
13. Itti, L.; Baldi, P. Bayesian surprise attracts human attention. *Vis. Res.* **2009**, *49*, 1295–1306.
14. Mazzaglia, P.; Çatal, O.; Verbelen, T.; Dhoedt, B. Curiosity-Driven Exploration via Latent Bayesian Surprise. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 22 February–1 March 2021.
15. Powers, W.T. Feedback: Beyond Behaviorism. *Science* **1973**, *179*, 351–356. [CrossRef] [PubMed]
16. Barter, J.W.; Yin, H.H. Achieving natural behavior in a robot using neurally inspired hierarchical perceptual control. *iScience* **2021**, *24*, 102948. [CrossRef]
17. Pellis, S.M.; Bell, H.C. Chapter 5 - Unraveling the dynamics of dyadic interactions: Perceptual control in animal contests. In *The Interdisciplinary Handbook of Perceptual Control Theory*; Mansell, W., Ed.; Academic Press: Cambridge, MA, USA, 2020; pp. 75–99. [CrossRef]
18. Mansell, W. The perceptual control model of psychopathology. *Curr. Opin. Psychol.* **2021**, *41*, 15–20. . [CrossRef]
19. Moreno-Bote, R.; Ramírez-Ruiz, J. Empowerment, Free Energy Principle and Maximum Occupancy Principle Compared. In Proceedings of the NeurIPS 2023 Workshop: Information-Theoretic Principles in Cognitive Systems, New Orleans, LA, USA, 10–16 December 2023. Available online: https://openreview.net/forum?id=OcHrsQox0Z (accessed on 7 February 2025).
20. Biehl, M.; Guckelsberger, C.; Salge, C.; Smith, S.C.; Polani, D. Expanding the Active Inference Landscape: More Intrinsic Motivations in the Perception-Action Loop. *Front. Neurorobot.* **2018**, *12*, 45. [CrossRef]
21. Sakthivadivel, D.A.R. Towards a Geometry and Analysis for Bayesian Mechanics. *arXiv* **2022**, arXiv:2204.11900.
22. Friston, K.; Da Costa, L.; Hafner, D.; Hesp, C.; Parr, T. Sophisticated Inference. *Neural Comput.* **2021**, *33*, 713–763. [CrossRef]
23. Tiomkin, S.; Nemenman, I.; Polani, D.; Tishby, N. Intrinsic Motivation in Dynamical Control Systems. *PRX Life* **2024**, *2*, 033009. [CrossRef]
24. Wissner-Gross, A.D.; Freer, C.E. Causal Entropic Forces. *Phys. Rev. Lett.* **2013**, *110*, 168702. [CrossRef]
25. Ashby, W.R. Requisite Variety and Its Implications for the Control of Complex Systems. In *Facets of Systems Science*; Springer: Boston, MA, USA, 1991; pp. 405–417. [CrossRef]
26. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [CrossRef]
27. Hohwy, J. *The Predictive Mind*; Oxford University Press: Oxford, UK, 2013.
28. Rao, R.P.N.; Ballard, D.H. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **1999**, *2*, 79–87. [PubMed]
29. Salvatori, T.; Song, Y.; Yordanov, Y.; Millidge, B.; Sha, L.; Emde, C.; Xu, Z.; Bogacz, R.; Lukasiewicz, T. A Stable, Fast, and Fully Automatic Learning Algorithm for Predictive Coding Networks. In Proceedings of the Twelfth International Conference on Learning Representations, Vienna, Austria, 7–11 May 2024.
30. von Helmholtz, H.; Southall, J.P.C. *Helmholtz's Treatise on Physiological Optics*; Dover Publications: New York, NY, USA, 1962.
31. Friston, K.J. A free energy principle for a particular physics. *arXiv* **2019**, arXiv:1906.10184.
32. Friston, K.; Da Costa, L.; Sakthivadivel, D.A.; Heins, C.; Pavliotis, G.A.; Ramstead, M.; Parr, T. Path integrals, particular kinds, and strange things. *Phys. Life Rev.* **2023**, *47*, 35–62. [CrossRef] [PubMed]
33. Smith, R.; Friston, K.J.; Whyte, C.J. A step-by-step tutorial on active inference and its application to empirical data. *J. Math. Psychol.* **2022**, *107*, 102632. [CrossRef]
34. Brown, H.R.; Friston, K.J.; Bestmann, S. Active Inference, Attention, and Motor Preparation. *Front. Psychol.* **2011**, *2*, 218. [CrossRef]
35. Kiefer, A. Psychophysical Identity and Free Energy. *J. R. Soc. Interface* **2020**, *17*, 20200370.
36. Botvinick, M.; Toussaint, M. Planning as inference. *Trends Cogn. Sci.* **2012**, *16*, 485–488. [CrossRef]
37. Hohwy, J. The Self-Evidencing Brain. *Noûs* **2014**, *50*, 259–285. [CrossRef]
38. Millidge, B.; Tschantz, A.; Buckley, C.L. Whence the Expected Free Energy? *Neural Comput.* **2021**, *33*, 447–482. [CrossRef]
39. Millidge, B.; Tschantz, A.; Seth, A.K.; Buckley, C.L. Understanding the origin of information-seeking exploration in probabilistic objectives for control. *arXiv* **2021**, arXiv:2103.06859.
40. Parr, T.; Pezzulo, G.; Friston, K. *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*; MIT Press: Cambridge, MA, USA, 2022.
41. Heins, C.; Millidge, B.; Demekas, D.; Klein, B.; Friston, K.; Couzin, I.D.; Tschantz, A. pymdp: A Python library for active inference in discrete state spaces. *J. Open Source Softw.* **2022**, *7*, 4098. [CrossRef]
42. Parr, T.; Friston, K.J. Generalised free energy and active inference. *Biol. Cybern.* **2018**, *113*, 495–513. [CrossRef] [PubMed]
43. Ueltzhöffer, K. On the thermodynamics of prediction under dissipative adaptation. *arXiv* **2020**, arXiv:2009.04006.
44. Ringstrom, T.J. Reward is not Necessary: How to Create a Modular & Compositional Self-Preserving Agent for Life-Long Learning. *arXiv* **2023**, arXiv:2211.10851.

45. Friston, K.J.; Da Costa, L.; Tschantz, A.; Kiefer, A.; Salvatori, T.; Neacsu, V.; Koudahl, M.; Heins, C.; Sajid, N.; Markovic, D.; et al. Supervised structure learning. *Biol. Psychol.* **2024**, *193*, 108891. [CrossRef]
46. England, J.L. Statistical physics of self-replication. *J. Chem. Phys.* **2013**, *139*, 121923. [CrossRef]
47. Costa, L.D. Probabilistic Principles for Biophysics and Neuroscience: Entropy Production, Bayesian Mechanics & the Free-Energy Principle. *arXiv* **2024**, arXiv:2410.11735.
48. Schopenhauer, A.; Payne, E.F.J. *The World as Will and Representation*; Dover Publications: New York, NY, USA, 1958.
49. Hesp, C.; Ramstead, M.; Constant, A.; Badcock, P.; Kirchhoff, M.; Friston, K. A Multi-scale View of the Emergent Complexity of Life: A Free-Energy Proposal. In *Evolution, Development and Complexity*; Georgiev, G.Y., Smart, J.M., Flores Martinez, C.L., Price, M.E., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 195–227.
50. Friston, K.; Heins, C.; Verbelen, T.; Costa, L.D.; Salvatori, T.; Markovic, D.; Tschantz, A.; Koudahl, M.; Buckley, C.; Parr, T. From pixels to planning: Scale-free active inference. *arXiv* **2024**, arXiv:2407.20292.
51. Fields, C.; Goldstein, A.; Sandved-Smith, L. Making the Thermodynamic Cost of Active Inference Explicit. *Entropy* **2024**, *26*, 622. [CrossRef]
52. Arocha, J.F. Scientific realism and the issue of variability in behavior. *Theory Psychol.* **2021**, *31*, 375–398. [CrossRef]
53. Smith, R.; Ramstead, M.J.D.; Kiefer, A. Active Inference Models Do Not Contradict Folk Psychology. *Synthese* **2022**, *200*, 81. [CrossRef]