

Addressing the Subsumption Thesis: A Formal Bridge between Microeconomics and Active Inference

Anonymous¹[0]

No Institute Given

Abstract. As a unified theory of sentient behaviour, active inference is formally intertwined with multiple normative theories of optimal behaviour. Specifically, we address what we call the subsumption thesis: The claim that expected utility from economics, as an account of agency, is subsumed by active inference. To investigate this claim, we present multiple examples that challenge the subsumption thesis. To formally compare these two accounts of agency, we analyze the objective functions for MDPs and POMDPs. By imposing information-theoretic rationality bounds (ITBR) on the expected utility agent, we find that the resultant agency is equivalent to that of active inference in MDPs, but slightly different in POMDPs. Rather than being strictly resolved, the subsumption thesis motivates the construction of a formal bridge between active inference and expected utility. This highlights the necessary formal assumptions and frameworks to make these disparate accounts of agency commensurable.

Keywords: Active Inference · Expected Utility · Information-Theoretic Bounded Rationality · Microeconomics

1 Introduction

Since the middle of the previous century, expected utility has formed the bedrock of the agency underwriting microeconomics. With early implementations dating back to Bernoulli in 1713 [38], expected utility has undergone many augmentations in order to reflect realistic deliberate decision processes. The comprehensive start of this lineage can be traced to the classic utility theorem [24]; [8], with earlier applications found in [31]. Subsequent accounts include Bayesian Decision Theory [35]; [6], Bounded Rationality [37]; [26], Prospect Theory [19], and many more flavours. The algorithmic implementation of expected utility theory is found in the Reinforcement Learning literature [3]. While seemingly disparate, practically all expected utility accounts of agency depict an agent making decisions in a probabilistic setting to attain optimal reward – to pursue utility [5].

Coming from the completely different background of neuroscience, Active Inference a comparatively new account of agency [12], positioning itself as “a unifying perspective on action and perception [...] richer than the common optimization objectives used in other formal frameworks (e.g., economic theory

and reinforcement learning)” [29, pg. 1;4]. Here the agent seeks to minimize information-theoretic surprisal expressed as free energy (See Definition 3, 4). Active inference allows for a realistic modeling of the very neuronal processes underwriting biological agency [27].

Given the breadth of successful applications [9] combined with its strong fundamental first principles [13], some proponents of active inference posited what we call the Subsumption thesis: Expected utility theory as seen in economics is subsumed by active inference – it is an edge case. A formulation in the same vein posits: “Active inference [...] englobes the principles of expected utility theory [...] it is theoretically possible to rewrite any RL algorithm [...] as an active inference algorithm” [11]. So how does the subsumption thesis hold up in the given examples? Is it possible to formally delineate how expected utility and active inference differ? This paper then establishes a firm connection between microeconomics and active inference, which has scarcely been explored before [17].

To formally compare the two accounts of agency, we require a commensurable space for agent-environment interactions: MDPs and POMDPs (Definition 1 and 2). These agent-environment frameworks are the bread and butter of expected utility applications [4], [3], [20]. Active inference agency has more recently also been specified for the same frameworks [28], [9], [10], [11]. As such (PO)MDPs provide a theoretical arena for the subsumption thesis to be evaluated.

What exactly is at stake that motivates this inquiry into the subsumption thesis? Firstly, expected utility and active inference rest upon different first principles to substantiate their respective account of agency [11]. Analysis of the formal relationship between these two accounts could provide insights into how the first principles of one account might be a specification the other’s first principles. Secondly, this inquiry will shed light on how each account handles the exploration-exploitation dilemma [7]: How should an agent prioritise between exploring an environment versus exploiting what they already know about the environment for utility? Finally, if active inference truly subsumed expected utility, then the ramifications for welfarist economics would be enormous: Currently, the formal mainstream understanding of welfare which informs economic policy [30] is based on aggregating individual agents acting according to expected utility [23, pg. 45] [33]. The subsumption thesis challenges foundations of ‘optimal’ economic policy if expected utility only captures a sliver of ‘optimal’ behaviour.

To investigate the subsumption thesis, the rest of the paper is structured as follows. In section 2. the agent-environment frameworks are defined alongside the relevant accounts of agency and basic concepts in microeconomics. In section 3., some examples are investigated which challenge the subsumption thesis. In section 4., the formal bridge between expected utility and active inference is established via Information Theoretic Bounded Rationality (ITBR) [26]. Finally section 5. provides some concluding and summarizing remarks.

2 Preliminary Definitions and Microeconomics

2.1 Agent-Environment Frameworks

A finite Markov Decision Process (MDP) is a mathematical model that specifies the elements involved in agent-environment interaction and development [3]. This formalization of sequential decision making towards reward maximization originates in dynamic programming, and currently enjoys much popularity in model-based Reinforcement Learning (RL). Although potentially reductive, employing MDPs and POMDPs allows for formal commensurability between different accounts of agency.

Definition 1 (Finite Horizon MDP). *An MDP is defined according to the following given tuple: $(\mathbb{S}, \mathbb{A}, P(s'|a, s), R(s', a), \gamma = 1, \mathbb{T})$*

- \mathbb{S} is a finite set of states.
- \mathbb{A} is a finite set of actions.
- $P(s'|a, s)$ is the transition probability of posterior state s' occurring upon the agent's selection of action a in the prior state s .
- $R(s', a) \in \mathbb{R}^+$ is the reward function taking as arguments the agent's action and resulting state. For our purposes, the action taken will be irrelevant to the resulting reward: $R(s', a) = R(s')$.
- γ denotes the discount factor of future rewards. This is set to 1 as this parameter is not commonly used in the cited active inference literature.
- $\mathbb{T} = \{1, 2, \dots, t, \dots, \tau, \dots, T\}$ is a finite set for discrete time periods whereby $t < \tau$ and the horizon is T .

Note that time period subscripts e.g. s_τ are sometimes omitted when unnecessary.

In a single-step decision problem, an expected reward-maximizing agent would evaluate the optimal action a^* as follows:

$$a_t^* = \arg \max_{a \in \mathbb{A}} E_{P(s_\tau|a_t, s_t)} R(s_\tau) \quad (1)$$

Further, a Partially Observable Markov Decision Process (POMDP) generalizes an MDP by introducing observations o that contain incomplete information about the latent state s of the environment [20, 3]. The agent can only infer latent states via observations. Thus, POMDPs are ideal for modeling action-perception cycles [13] with the cyclical causal graphical model $a \rightarrow s \rightarrow o \dots$

Definition 2 (Finite Horizon POMDP). *A finite horizon POMDP further adds two elements to the previously given MDP tuple: $(\mathbb{O}, P(o|s))$*

- \mathbb{O} is a finite set of observations.
- $P(o|s)$ is the probability of observation o occurring to the agent given the state s .

2.2 Active Inference Agency

With the environment-agent frameworks established, we can proceed to define how an active inference agent approaches a (PO)MDP. While fundamental and interesting, the Variational Free Energy objective crucial to perception in active inference will not be examined here; Inference on latent states is assumed to occur through exact Bayesian inference [10, pg. 16]. The central objective function for agency in active inference is the Expected Free Energy (EFE), the formulation of which for (PO)MDPs we will take from [9],[11],[10],[28]. Essentially, the agent takes the action trajectory $\pi = \{a_\tau, \dots, a_T\}$ that minimizes the cumulative expected free energy G , which is roughly the sum of the single-step EFEs G_τ . By inferring the resultant EFE of policies through $Q(\cdot)$, the optimal trajectory π^* corresponds to the most likely trajectory – the path of least action. [13]. Formally:

$$\begin{aligned} \pi^* &= \arg \min_{\pi} G(\pi) \\ G(\pi) &\approx \sum_{\tau}^T G_\tau(\pi) \\ G_\tau(\pi) &= G(a_t) \end{aligned} \tag{2}$$

We can then define the EFE for single-step for MDPs and POMDPs. Note that this could also be scaled up to trajectories/vectors of the relevant elements e.g $s_{t:T}$. For simplicity we will look at single-step formulations for the remainder of the paper.

Definition 3: (EFE on MDPs). *For an agent in an MDP with preference distribution $P(s|C)$, the Expected Free Energy of an action for some given current state s_t is defined as follows:*

$$\begin{aligned} G_\tau(a_t) &= D_{KL}[P(s_\tau|a_t, s_t) || P(s|C)] \\ &= - \underbrace{\mathfrak{H}[P(s_\tau|a_t, s_t)]}_{\text{Entropy of future states}} - \underbrace{E_{P(s_\tau|a_t)}[\log P(s|C)]}_{\text{Expected Surprise}} \end{aligned} \tag{3}$$

As seen in the rearranged objective function of the second line, the agent seeks to keep future options open while meeting preferences; The entropy of future possible states is to be maximized while the information-theoretic surprisal according to the preference distribution is to be minimized. The conditionalisation on C specifies a parameterized preference distribution [28].

Definition 4: (EFE in POMDPs). *For an agent in a POMDP with preference distribution $P(s|C)$, $P(o|C)$, the Expected Free Energy of an action for some given current state s_t is defined as follows:*

$$\begin{aligned}
G_\tau(a_t) &= \underbrace{E_{P(s_\tau|o_t, a_t)}[\mathfrak{H}[P(o_\tau|s_\tau)]]}_{\text{Ambiguity}} + \underbrace{D_{KL}[P(s_\tau|a_t, o_t)|P(s|C)]}_{\text{Risk}} \\
&= - \underbrace{E_{P(o_\tau|a_t)}[D_{KL}[P(s_\tau|o_\tau)||P(s_\tau|a_t)]]}_{\text{Intrinsic Value}} - \underbrace{E_{P(o_\tau, s_\tau|a_t)}[\log P(o|C)]}_{\text{Extrinsic Value}}
\end{aligned} \tag{4}$$

With some auxiliary assumptions [11, pg. 10] which are admissible for our purposes, the two formulations of EFE in a POMDP are equivalent, and both contain a curiosity inducing term and an exploitation term [15]. The first formulation motivates the agent to minimize the expected entropy of observations given unknown states and to minimize the divergence between actual states and preferred states. The second formulation motivates the agent to maximize the expected informational value of observations while also maximizing the expected log probability of preferred observations – note how the underbrace does not include the minus.

2.3 Microeconomics

As this paper investigates an intersection between fields which are generally not in direct contact, a brief introduction to risk attitudes and lotteries in microeconomics is provided. The origin of these studies can be traced back to the gambling houses of the 18th century; As early as 1713, Bernoulli employed marginally decreasing utility functions to resolve the famous St. Petersburg Paradox [38]. This paradox asks what amount a rational agent would be willing to pay to enter lottery with an infinite expected value. To answer this question, we utilize lotteries [8] and risk-attitudes [1] from microeconomics:

Definition 5: (Lottery). *A (monetary) lottery is a probability distribution over outcomes x that are the argument of the utility function. Therefore, a lottery L can be modelled as an integrable random variable defined by the probability space triplet consisting of a sample space, sigma algebra, and probability measure: $(\Omega, \mathfrak{F}, \mu)$*

A decision maker then evaluates their preference over a set of lotteries according to their utility function $U(x) \in \mathbb{R}^+$ where $x \in \mathfrak{X}$. The expected utility of each lottery then induces a preference ordering over lotteries. For example, the strong preference relation $L_1 \succ L_2$ means that lottery L_1 is more preferable to lottery L_2 . Classically, this ordering is in line with the von Neumann-Morgenstern axioms of completeness, transitivity, continuity, and independence [24]. Any such preference ordering is also maintained for any positive affine transformation of $U(x)$ [8]; [24]. By juxtaposing the expected utility $E[U(L)]$ of a lottery against the utility of the expectation of the same lottery $U(E[L])$, risk aversion can be defined.

Definition 6: (Risk Aversion). *An agent with some utility function $U(\cdot)$ is considered risk averse if for some lottery the following preference relation holds: $U(E[L]) \succ U(L)$*

This preference relation occurs if an agent’s utility function is concave, i.e the marginal utility is decreasing. A risk loving agent conversely acts according to a convex utility function, and a risk neutrality is associated with a linear utility function. Accordingly, Bernoulli used lotteries and a log-utility function to resolve the St. Petersburg paradox, the solution of which is relegated to appendix A for readers unfamiliar with the problem – The pertinent point is that concave utility functions on set rewards are extensively studied in economics.

3 Subsumption Examples

Equipped with an understanding of marginal utility and lotteries, we can now tackle two manifest exhibits of the subsumption thesis by proponents of active inference. Further, an illustrative MDP demonstrates the divergence in behaviour between active inference and expected utility. The results of the simulated behaviour are directly taken from the discussed papers. These exhibits then motivate the bridging in section 4 later.

The first [34] and second [11] exhibit both concern agency in a classical T-maze: A simple forked pathway in which the agent can either go left or right (See Figure 1 below). This environment is also called “Light” in POMDP literature [20]. The agent-environment dynamics are modeled using a POMDP; Unbeknownst to the agent, the reward is either in the left or right arm. The agent can also go down to observe a cue indicating the definite location of the reward. Going down the ‘wrong’ arm of the fork leads to a punishment equal to the negative reward, say -1 . The performance of the agency is evaluated by the reward attainment of the agent within a two period horizon. At this point however, the setup of the first and second exhibit diverge crucially.

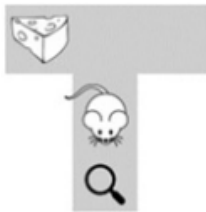


Fig. 1. An agent in a T-Maze with unknown context. Illustration from [34]

In the first exhibit [34], the right and left fork are absorbing states – the agent cannot leave them upon entry. As such, the agent cannot correct going down the wrong arm in the first period by the second period. Given this setup, the expected utility agent performs very poorly, while the active inference is cue-seeking and therefore performs optimally [34, pg. 138]. The expected utility agent performs so poorly because supposedly “the agent does not care about

the information inferred and is indifferent about going to the cue location or remaining at the central location”[34, pg. 137]. This appears reductive, as an expected utility agent facing two lotteries will behave the same as the active inference agent. Consider the risky lottery L_1 which is the result of a gambling and non-information seeking strategy. Contrast this lottery with L_2 , which is the degenerate lottery of investigating the cue first and going to the reward in the second period. Assuming even just a linear utility function $U(R) = R(s)$, then $U(L_1) = 0.5 \cdot 1 + 0.5 \cdot -1 = 0$ and $U(L_2) = 1 \cdot 1$. Clearly, the expected utility agent holds a preference which motivates cue-seeking behaviour: $L_2 \succ L_1$.

Regarding the second exhibit [11], there is a slight difference in the setup. The arms of the fork are no longer absorbing states, which allows for mistake correction and a cumulative reward of 2 over two periods. Now, the focus of [11] isn’t anymore on performance comparison but instead achieving the desiderata of risk-aversion and information sensitivity [11, pg. 10]. While the agency according to active inference meets the desiderata, the expected utility agent does not. However, risk aversion and the resulting information sensitivity can easily be induced by using a concave utility function. Consider again the risky lottery L_1 and a cue-seeking lottery L_2 . Assuming a utility function taking the reward as argument $U(R) = R(s)^c$ where $c \in \mathbb{R}^+$, then $U(L_1) = 0.5 \cdot 0 + 0.5 \cdot 2^c$ and $U(L_2) = 1^c$. Accordingly if $c < 1$, then $L_2 \succ L_1$, and only if $c = 1$, then indeed the agent is indifferent $L_1 \sim L_2$. As is evident, it is the risk-neutral agent who does not meet the desiderata.

Finally, consider the following single-step MDP created for illustrative purposes. A paraglider stands at the foot of two steep mountains s_1, s_2 separated by a chasm s_3 and must decide which one to climb. While still risky, the path up mountain 1 is far more secure than the path up mountain 3. However, mountain 2 is taller than mountain 1 and therefore allows for a more enjoyable flight. This decision process can aptly be modeled in an MDP (See Figure 2 below). Note that the subscript here does not relate to the period. Taking a_1 gives $\{P(s_1|a_1), P(s_2|a_1), P(s_3|a_1)\} = \{0.6, 0, 0.4\}$, and a_2 gives $\{P(s_1|a_2), P(s_2|a_2), P(s_3|a_2)\} = \{0, 0.4, 0.6\}$. The height in kilometers gives the reward function $\{R(s_1), R(s_2), R(s_3)\} = \{1, 1.5, 0\}$. With the MDP sufficiently specified, we can compare the agency of an active inference agent and an expected utility agent. See Appendix B for details on the resulting expected utility and free energy.

The active inference agent is indifferent between the two actions as both actions result in the same EFE – equation (3). For expected utility however, only the linear utility function agent is indifferent; The risk averse agent prefers the safer mountain and the risk loving agent prefers the riskier mountain due to the concavity or convexity of the utility function respectively. As such, this simple but valid MDP provides a setup in which specified expected utility may better meet the desiderata than the active inference agent.

It should be clear by now that wrapping a utility function around the rewards is a well-studied and principled approach which differs from simply including “ad-hoc exploration bonuses in the reward function” [11, pg. 2]. Introducing non-linearity over the rewards seems to lead to an impasse in the comparison

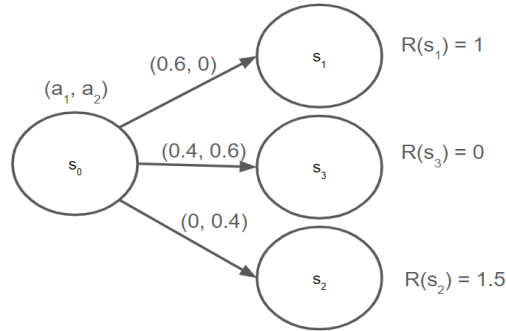


Fig. 2. ‘Paraglider’ MDP with states, actions, transition probabilities, and rewards

between expected utility. The most direct case for comparing and subsuming expected utility [10] only considers a linear utility function ($U(\cdot) = R(\cdot)$) for the expected utility agent. Even if non-linearity for expected utility were considered in [10], it appears unclear to us as to how the resulting agency – more specifically the induced careful and explorative aspect – could be compared in a generalized manner.

To resolve this issue of incommensurability, we would like to draw attention to physical and biological constraints on agents which have motivated active inference. For example, tractability is a central concern for active inference as evidenced by the appeal to variational Bayes. Luckily, there already exists an account of agency which imbues expected utility with constraints: ITBR [26],[16]. The connection between ITBR and active inference in an MDP has briefly been explored before [25]. We seek to now clearly establish this conceptual bridge between microeconomics and active inference for both MDPs and POMDPs.

4 From expected utility to active inference via ITBR

4.1 In MDPs

Let us first establish the bridge between expected utility and active inference in an MDP. Essentially, both objective functions can both be transformed into the “Divergence Objective” [21]:

$$a^* = \arg \min_{a \in \mathbb{A}} D_{KL}[P(s_\tau | a_t) || P^*(s)] \quad (5)$$

Where a^* is the optimal action and $P^*(s)$ is a preference distribution over states, for example, a softmax or Gibbs distribution. Note the immediate similarity to the EFE objective function for MDPs (3) – here conditionalisation on the current state s_t is omitted for brevity as we consider a single step.

To get there from expected utility, we can consider the following Lagrangian

constraints on the utility objective function [16, pg. 3]. Let $P(\cdot)$ be the prior distribution over relevant elements of the MDP, and $Q(\cdot)$ the posterior distribution after limited search or ‘bounded deliberation’ – see [26] for details. The deliberation bound is given as an information-theoretic quantity– hence the name information-theoretic bounded rationality. Let $J, K \in \mathbb{R}^+$ be in terms of whatever information-theoretic unit base is chosen:

$$\begin{aligned} D_{KL}[Q(a)||P(a)] &\leq J \\ D_{KL}[Q(s_\tau|a_t)||P(s_\tau|a_t)] &\leq K \end{aligned} \quad (6)$$

The first constraint of equation 6 can be interpreted as a bound on the search for the optimal action. The second constraint of (6) similarly means that the agent is uncertain about the transition probabilities in the MDP. These constraints give us the following ITBR free energy objective function [25]:

$$F_{ITBR}(Q) = \sum_a Q(a) \sum_s Q(s|a) U(s, a) - \frac{1}{\alpha} \log \frac{Q(a)}{P(a)} - \frac{1}{\beta} \log \frac{Q(s|a)}{P(s|a)} \quad (7)$$

This function is to be maximized ($Q^*(a), Q^*(s|a)$) with given parameters $\alpha, \beta \in \mathbb{R}^+$. The objective function (7) can be rearranged:

$$= \sum_a Q(a) \left(-\frac{1}{\alpha} \log \frac{Q(a)}{P(a)} - \frac{1}{\beta} D_{KL}[Q(s|a)||P(s|a)] - \frac{1}{\beta} \log \underbrace{\left(\sum_s P(s|a) e^{\beta U(s,a)} \right)}_{Z_\beta(a): \text{Normalising constant}} \right) \quad (8)$$

Where $Z_\beta(a)$ is the partition function dependent on a . Now, the parameters are set to be in line with the MDP setting for active inference: $\alpha \rightarrow \infty$ because there is no cost of action selection, and $\beta = 1$ as there is no ‘malevolent’ model uncertainty due to the given MDP transition probabilities. Finally, we introduce the preference distribution over states $P^*(s)$. Like in active inference, we assume that the preference distribution over states is independent of the action taken to get there. Further, we take the preference distribution to be the a priori known optimal argument of objective function (7) given by the Gibbs distribution:

$$P^*(s|a) = \frac{P(s|a) \cdot e^{\beta U(s,a)}}{Z_\beta(a)} \rightarrow P^*(s) \quad (9)$$

Inputting these values for parameters and preference distribution (9) into objective function (8) gives us the previous negative divergence objective (5) plus some constant irrelevant for maximization.

$$= -D_{KL}[Q(s_\tau|a_t)||P^*(s)] + \text{constant} \quad (10)$$

Where the transition equation does not have to be deliberated as it is given by the MDP tuple.

As we are looking for the maximizing argument a_t for the negative KL divergence, this is of course equivalent to the minimizing argument for the positive KL divergence. As such, the optimal action a^* in an MDP is the same for the ITBR specification (10), the divergence objective (5), and active inference (3). Next, let us bridge expected utility to active inference in a POMDP.

4.2 In POMDPs

Analogous to the MDP setting, we can transform the ITBR objective to get to the divergence objective function for POMDPs. Luckily, this divergence objective has been formulated before as the “Free Energy of the Expected Future” (FEEF): [22, pg. 10]. Again, this objective function motivates a minimal posterior divergence from a preference distribution, now jointly over states and observations:

$$a^* = \arg \min_{a \in \mathbb{A}} D_{KL}[P(o_\tau, s_\tau | a_t) || P^*(o, s)] \quad (11)$$

To attain this expression, we can formulate a new ITBR objective in the POMDP framework [16] and transform it analog to the MDP case before. We can again first consider the information-theoretic bounds $V, K \in \mathbb{R}^+$:

$$\begin{aligned} D_{KL}[Q(a) || P(a)] &\leq K \\ D_{KL}[Q(s_\tau, o_\tau | a_t) || P(s_\tau, o_\tau | a_t)] &\leq V \end{aligned} \quad (12)$$

The causal connection between observation, action, and latent state is again given by the POMDP tuple of definition 2. The expectation in the objective function (13) is essentially the action-averaged generative model fundamental to active inference [32] $P(o, s, a) = \sum_a P(o, s) = \sum_a P(o|s)P(s)$. Considering these constraints, we can express the ITBR Free energy objective function again:

$$\begin{aligned} F_{ITBR} &= \underbrace{E_{Q(a)} E_{Q(o, s|a)}}_{\text{Generative Model}} \left[U(o, s, a) - \frac{1}{\alpha} \log \frac{Q(a)}{P(a)} - \frac{1}{\beta} \log \frac{Q(o, s|a)}{P(o, s|a)} \right] \\ &= \sum_a Q(a) \left(-\frac{1}{\alpha} \log \frac{Q(a)}{P(a)} - \frac{1}{\beta} D_{KL}[Q(o, s|a) || P^*(o, s|a)] - \frac{1}{\beta} \log Z_\beta(a) \right) \end{aligned} \quad (13)$$

where similarly as before

$$\begin{aligned} Z_\beta(a) &= \sum_{o, s} P(o, s|a) e^{\beta U(o, s, a)} \\ P^*(o, s|a) &= \frac{P(o, s|a) e^{\beta U(o, s, a)}}{Z_\beta(a)} \end{aligned}$$

Setting parameters $\alpha \rightarrow \infty$, $\beta = 1$ and assumptions analog as before to attain 10, we get the resultant maximization objective, which is of course the same as the divergence minimization objective (11):

$$= -D_{KL}[Q(o_\tau, s_\tau | a_t) || P^*(o, s)] + \text{constant} \quad (14)$$

Which again intuitively motivates the agent to have the inferred posterior distribution given the action be as close as possible to the prior preference distribution over states. It is crucial to note however that this is not the same objective function as EFE in POMDPs (4)! To get from the divergence objective (11) for POMDPs to EFE (4), we can follow the steps taken in [22]; for a detailed discussion on the relationship between the divergence objective and EFE the reader is to consult [21], [22] as well. Essentially, the divergence objective can also be decomposed into an exploitative and explorative term. However, while the explorative term is equal to that of active inference, the divergence objective additionally further encourages the agent to increase posterior entropy of observations given latent states – to keep options open. Note that in the formulation below, both objective functions below (15), (4) are to be minimized.

$$\begin{aligned} -F_{ITBR} &= D_{KL}[Q(o, s|a)||P^*(o, s)] \\ &= \underbrace{E_{Q(s|a)} [D_{KL}[Q(o|s)||P^*(o)]]}_{\text{Extrinsic Value}} - \underbrace{E_{Q(o|a)} [D_{KL}[Q(s|o)||Q(s|a)]]}_{\text{Intrinsic Value}} \end{aligned} \quad (15)$$

$$\begin{aligned} G &= - \underbrace{E_{Q(o, s|a)} [\log P(o|C)]}_{\text{Extrinsic Value}} - \underbrace{E_{Q(o|a)} [D_{KL}[Q(s|o)||Q(s|a)]]}_{\text{Intrinsic Value}} \end{aligned} \quad (4)$$

Whereby the relationship between G and $-F_{ITBR}$ is as follows:

$$G - E_{Q(o|a)} \mathfrak{H}[Q(o|s)] = -F_{ITBR} \quad (16)$$

Thus comparing the decomposed divergence objective to active inference, in the pursuit of extrinsic value the boundedly rational utility agent seeks to additionally keep posterior options open compared to the active inference agent – similarly to the agency in an MDP.

4.3 The Bridge summarized

Let us reconsider the entire journey from expected utility to active inference so as to not lose sight of the forest in front of all the trees. First, simply incorporate a utility function into the reward maximizing objective function (1) to get an expected utility agent. Then, impose information-theoretic deliberation constraints on the optimization process (6). Consequently, the agent faces a Lagrangian optimization problem (7). This objective function is rearranged (8), and agent-environment framework assumptions subsequently are incorporated. This results in the divergence objective [21], which can then be compared to the active inference objective function. In an MDP, the resultant agency is the exact same (5). In a POMDP however, the objective functions differ (16).

Bar this difference, one key aspect must be elucidated for the extrinsic value terms in both MDPs and POMDPs. While the intrinsic value term is equal for the different objective functions, the two prior preference distributions $P^*(s)$ of ITBR (9) and $P(s|C)$ of active inference (3) are not necessarily the same. For

$\beta = 1$, if we consider $P(s|C)$ as a Gibbs distribution as per [10, pg. 9];[28, pg. 134], then the two preference distributions are only equal if either the utility function is linear, or if active inference admits agent-specific utility functions (17); An admission which prima-facie seems irreconcilable with the physicalist/non-subjectivist philosophy behind active inference. This larger discussion is however to be relegated to a later paper.

$$P^*(s) = \frac{e^{U(s)}}{\sum_s e^{U(s)}} \quad \text{and} \quad P(s|C) = \frac{e^{R(s)}}{\sum_s e^{R(s)}} \quad (17)$$

Where optimal behaviour in an MDP, i.e a^* , is the same for both accounts of agency only if $U(s)$ is a positive affine transformation of $R(s)$.

5 Conclusion

Having formalized the bridge from expected utility to active inference, we can reevaluate the subsumption thesis. Simple reward-oriented agency ($U(\cdot) = R(\cdot)$) can be effectively subsumed by active inference in MDPs, and if exact Bayesian inference is used, also in POMDPs [10]. However, as shown in section 2., expected utility in microeconomics uses utility functions that take rewards as arguments. As seen in section 3. then, there are various examples where the subsumption argument does not hold up; Expected utility acts the same as active inference, or under specific circumstances, may meet desiderata of agency even more. In section 4., we establish the formal bridge between expected utility and active inference. By using ITBR [26], we can directly compare the objective functions of (bounded) expected utility and active inference. Upon considering agent-environment assumptions, the divergence objective [21] is used as a reference point to compare the two accounts of agency. It is demonstrated that in an MDP, ITBR and active inference lead to the same agency [25]. In a POMDP, ITBR is equivalent to the divergence objective, which however differs from the active inference objective function [22]. While the explorative/information-seeking terms are equal, the exploitative/reward-oriented term differs: $E_{Q(o|a)} \mathfrak{H}[Q(o|s)]$ must be subtracted from the active inference objective function, and the preference distributions are not necessarily equal.

An area where expected utility cannot compete however is in the first principles which motivate agency [12], [13] [14], [2], [11]. Still, the debate on what objective function follows from the first principles is not yet sealed in this flourishing field [22]. Furthermore, computational simulations [16] and empirical studies [36] might flesh out the practical comparison between bounded expected utility and active inference; Computational efficiency has not remotely been addressed in this paper. Finally, it would be especially interesting for economics to understand how an economy could develop from multiple ITBR or active inference agents [18]. By integrating interdisciplinary approaches to agency, we aim to foster a holistic understanding of agency that enriches the roles of both human and artificial agents in society.

Acknowledgments. I would like to express my immense gratitude to my supervisor for allowing me to delve into this topic and lending his support along the way. Further, I want to thank the various researchers willing to so openly discuss the contents and concepts of the paper. Only thanks to those fruitful exchanges could these connections across varying fields even be grasped. As per the paper submission formalities requested, the exact names have been omitted to maintain anonymity.

Disclosure of Interests. The author has no competing interests to declare that are relevant to the content of this article.

References

1. Arrow, K.J.: *Essays in the Theory of Risk Bearing*. Markham Publishing Co, Chicago (1971)
2. Barp, A., Da Costa, L., França, G., Friston, K., Girolami, M., Jordan, M.I., Pavliotis, G.A.: Geometric methods for sampling, optimisation, inference and adaptive agents. *Handbook of Statistics* **46**, 21–78 (2022). <https://doi.org/10.48550/arXiv.2203.10592>, <https://doi.org/10.48550/arXiv.2203.10592>, arXiv:2203.10592v3 [stat.ML]
3. Barto, A., Sutton, R.S.: *Reinforcement Learning: An Introduction*. The MIT Press, 2nd edn. (2018)
4. Bellman, R.: A markovian decision process. *Journal of Mathematics and Mechanics* **6**, 679–684 (1957). <https://doi.org/10.1512/iumj.1957.6.56038>, <https://doi.org/10.1512/iumj.1957.6.56038>
5. Bentham, J.: *An Introduction to the Principles of Morals and Legislation*. Batoche Books, Kitchener, 2000 edn. (1781)
6. Berger, J.O.: *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics, Springer-Verlag, New York, 2nd edn. (1985). <https://doi.org/10.1007/978-1-4757-4286-2>
7. Berger-Tal, O., Nathan, J., Meron, E., Saltz, D.: The exploration-exploitation dilemma: A multidisciplinary framework. *PLOS ONE* **9**(4), e95693 (April 2014). <https://doi.org/10.1371/journal.pone.0095693>
8. Bonanno, G.: *Decision making* (2017), https://faculty.econ.ucdavis.edu/faculty/bonanno/PDF/DM_book.pdf
9. Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., Friston, K.: Active inference on discrete state-spaces: a synthesis. *Journal of Mathematical Psychology* **102447**, 36 (2021). <https://doi.org/10.1016/j.jmp.2020.102447>, <https://doi.org/10.48550/arXiv.2001.07203>, submitted on 20 Jan 2020 (v1), last revised 28 Mar 2020 (this version, v2)
10. Da Costa, L., Sajid, N., Parr, T., Friston, K., Smith, R.: Reward maximisation through discrete active inference. *arXiv preprint arXiv:2009.08111* **v4**, 18 pages (2022), <https://doi.org/10.48550/arXiv.2009.08111>
11. Da Costa, L., Tenka, S., Zhao, D., Sajid, N.: Active inference as a model of agency. *arXiv preprint arXiv:2401.12917* (2024), <https://doi.org/10.48550/arXiv.2401.12917>, accepted in RLDM2022 for the workshop 'RL as a model of agency'
12. Friston, K.: The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences* **13**(7), 293–301 (July 2009). <https://doi.org/10.1016/j.tics.2009.04.005>

13. Friston, K., Da Costa, L., Sajid, N., Heins, C., Ueltzhöffer, K., Pavliotis, G.A., Parr, T.: The free energy principle made simpler but not too simple. *Physics Reports* **1024**, 1–29 (June 2023). <https://doi.org/10.1016/j.physrep.2023.07.001>
14. Friston, K., Da Costa, L., Sakthivadivel, D.A., Heins, C., Pavliotis, G.A., Ramstead, M., Parr, T.: Path integrals, particular kinds, and strange things. *Physics of Life Reviews* **47** (2023). <https://doi.org/10.1016/j.pprev.2023.08.016>, <https://doi.org/10.48550/arXiv.2210.12761>
15. Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., Pezzulo, G.: Active inference and epistemic value. *COGNITIVE NEUROSCIENCE* **6**(4), 187–224 (2015). <https://doi.org/10.1080/17588928.2015.1020053>, <http://dx.doi.org/10.1080/17588928.2015.1020053>
16. Genewein, T., Leibfried, F., Grau-Moya, J., Braun, D.A.: Bounded rationality, abstraction, and hierarchical decision-making: An information-theoretic optimality principle. *Frontiers in Robotics and AI* **2**, 27 (2015). <https://doi.org/10.3389/frobt.2015.00027>, <https://doi.org/10.3389/frobt.2015.00027>, this article is part of the Research Topic Theory and Applications of Guided Self-Organisation in Real and Synthetic Dynamical Systems
17. Henriksen, M.: Variational free energy and economics: Optimizing with biases and bounded rationality. *Frontiers in Psychology* **11** (November 2020). <https://doi.org/10.3389/fpsyg.2020.549187>, <https://doi.org/10.3389/fpsyg.2020.549187>
18. Hyland, D., Gavenciak, T., Da Costa, L., Heins, C., Kovarik, V., Gutierrez, J., Wooldridge, M., Kulveit, J.: Multi-agent active inference. Forthcoming Manuscript in preparation
19. Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. *Econometrica* **47**(2), 263–291 (March 1979)
20. Littman, M.: A tutorial on partially observable markov decision processes. *Journal of Mathematical Psychology* **53**(2), 119–125 (2009)
21. Millidge, B., Seth, A., Buckley, C.: Understanding the origin of information-seeking exploration in probabilistic objectives for control. arXiv preprint arXiv:2103.06859 (2021), <https://doi.org/10.48550/arXiv.2103.06859>, submitted on 11 Mar 2021 (v1), last revised 24 Nov 2021 (this version, v7)
22. Millidge, B., Tschantz, A., Buckley, C.L.: Whence the expected free energy? *Neural Computation* **33**(2), 447–482 (February 2021). https://doi.org/10.1162/neco_a_01354, https://doi.org/10.1162/neco_a_01354
23. Mongin, P.: A concept of progress for normative economics. *Economics and Philosophy* **22**, 19–54 (2006). <https://doi.org/10.1017/S0266267105000696>
24. von Neumann, J., Morgenstern, O.: *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, NJ (1953)
25. Ortega, P.A., Braun, D.A.: What is epistemic value in free energy models of learning and acting? a bounded rationality perspective. *Cognitive Neuroscience* **6**(4), 215–216 (2015). <https://doi.org/10.1080/17588928.2015.1051525>, <https://doi.org/10.1080/17588928.2015.1051525>
26. Ortega, P.A., Braun, D.A., Dyer, J., Kim, K.E., Tishby, N.: Information-theoretic bounded rationality. arXiv preprint arXiv:1512.06789 (2015), <https://doi.org/10.48550/arXiv.1512.06789>, submitted on 21 Dec 2015
27. Parr, T., Markovic, D., Kiebel, S.J., Friston, K.J.: Neuronal message passing using mean-field, bethe, and marginal approximations. *Scientific Reports* **9**(1), 1–18 (2019). <https://doi.org/10.1038/s41598-019-50764-9>

28. Parr, T., Pezzulo, G., Friston, K.J.: *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. The MIT Press (2022)
29. Pezzulo, G., Parr, T., Friston, K.: Active inference as a theory of sentient behavior. *Biological Psychology* **186** (February 2024). <https://doi.org/10.1016/j.biopsycho.2023.108741>, under a Creative Commons license
30. Pigou, A.C.: *The Economics of Welfare*. Macmillan and Co., Limited, London (1920)
31. Ramsey, F.P.: Truth and probability. In: Braithwaite, R.B. (ed.) *The Foundations of Mathematics and Other Logical Essays*, chap. VII, pp. 156–198. Kegan, Paul, Trench, Trubner & Co. and Harcourt, Brace and Company, London and New York (1931), originally published in 1926
32. Ramstead, M.J., Kirchhoff, M.D., Friston, K.J.: A tale of two densities: active inference is enactive inference. *Adaptive Behavior* **28**(4), 225–239 (2020). <https://doi.org/10.1177/1059712319862774>
33. Ross, D.: *Philosophy of Economics*. Palgrave Philosophy Today, Palgrave Macmillan London, 1 edn. (2014). <https://doi.org/10.1057/9781137318756>, <https://doi.org/10.1057/9781137318756>
34. Sajid, N., Da Costa, L., Parr, T., Friston, K.: Active inference, bayesian optimal design, and expected utility. In: Cogliati Dezza, I., Schulz, E., Wu, C.M. (eds.) *The Drive for Knowledge: The Science of Human Information Seeking*, pp. 124–146. Cambridge University Press, Cambridge (2022)
35. Savage, L.J.: *The Foundations of Statistics*. Dover Publications, Inc., New York, N.Y., revised and enlarged edn. (1972), originally published by John Wiley & Sons in 1954
36. Schwartenbeck, P., FitzGerald, T.H.B., Dolan, R.J., Friston, K.J.: Evidence for surprise minimization over value maximization in choice behavior. *Scientific Reports* **5**, 16575 (2015). <https://doi.org/10.1038/srep16575>, <https://doi.org/10.1038/srep16575>
37. Simon, H.A.: *Models of Man*. John Wiley & Sons, New York (1957)
38. Szapiro, G.: *Risk, Choice, and Uncertainty: Three Centuries of Economic Decision-Making*. Columbia University Press, New York City (2020)

Appendix

A:

Resolving the St. Petersburg Paradox

Consider a lottery on the outcome of a fair coin toss. Starting at two dollars, the stake doubles with every subsequent outcome of heads. The game ends once tails comes up for the first time in the sequence. The expected payout $E[L]$ of the game is thus infinite:

$$E[L] = \sum_{i=1}^{\infty} \frac{1}{2^i} \cdot 2^i = \infty$$

How much would someone pay to participate in this game? Taking a linear utility function on the payout, the gambler should be willing to pay any amount to enter the game. Daniel Bernoulli suggested a logarithmic utility function $U(x) = \ln(x)$. Assume the cost of entry is x . Then the expected utility of the lottery has a finite value; The amount the agent at most would be willing to enter the lottery:

$$E[U(L)] = \sum_{i=1}^{\infty} \frac{1}{2^i} \cdot \ln(2^i) = 2 \cdot \ln(2)$$

Therefore the agent expects finite utility from the payout of the lottery due to the concavity of the utility function. As such, only a finite amount will be paid to enter the game.

B:

Expected Utility and Active Inference for the 'Paraglider' MDP

The single-step MDP is specified as follows. Therefore note that the subscript does not pertain to the period:

$$\begin{aligned} \mathbb{S} &= s_1, s_2, s_3 \\ \mathbb{A} &= a_1, a_2 \\ \{P(s_1|a_1), P(s_2|a_1), P(s_3|a_1)\} &= \{0.6, 0, 0.4\} \\ \{P(s_1|a_2), P(s_2|a_2), P(s_3|a_2)\} &= \{0, 0.4, 0.6\} \\ \{R(s_1), R(s_2), R(s_3)\} &= \{1, 1.5, 0\} \end{aligned}$$

Consider an expected utility agent with utility function $U(R(s)) = R(s)^c$ where $c \in \mathbb{R}^+$. As such,

$$\begin{aligned} E[U(a_1)] &= 0.6 \cdot 1^c \\ E[U(a_2)] &= 0.4 \cdot 1.5^c \\ \text{For } c < 1 &\rightarrow \arg \max_{a \in \mathbb{A}} E[U(a)] = a_1 \\ \text{For } c > 1 &\rightarrow \arg \max_{a \in \mathbb{A}} E[U(a)] = a_2 \end{aligned}$$

So a risk-averse expected utility agent will scale the smaller but safer mountain.

The active inference agent however is indifferent between the two actions. If we assume the preference distribution to be a softmax on the rewards, then we can ignore the normalizing denominator as it is constant w.r.t to action. Therefore we can write the relevant objective function as:

$$\begin{aligned}
 G(a_t) &= - \sum_s P(s_\tau | a_t) \cdot R(s_\tau) - \sum_s P(s_\tau | a_t) \log \frac{1}{P(s_\tau | a_t)} \\
 G(a_1) &= -0.6 - 0.3065 - 0.366 = G(a_2) \\
 \rightarrow \arg \min_{a \in \mathbb{A}} G(a) &= \{a_1, a_2\}
 \end{aligned}$$

Therefore the optimal action of the risk-averse expected utility agent is a subset of the optimal active inference agency.

C:

Using the Gibbs distribution optimal argument to get from ITBR to the divergence objective

In an MDP setting:

(9)

$$P^*(s|a) = \frac{P(s|a)e^{\beta U(s)}}{Z_\beta}$$

$$\frac{1}{\beta} \ln(P^*(s|a) \cdot Z_\beta) = U(s)$$

plug in for $U(s)$ into (7) reparameterized:

$$F_{ITBR} = \sum_s Q(s|a) U(s, a) - \frac{1}{\beta} \log \frac{Q(s|a)}{P(s|a)}$$

to attain

$$\begin{aligned} & \operatorname{argmax}_a \frac{1}{\beta} E_{Q(s|a)} [\ln P^*(s|a) + \ln(Z_\beta)] - \frac{1}{\beta} \ln \frac{Q(s|a)}{P(s|a)} \\ &= \operatorname{argmax}_a E_{Q(s|a)} [\ln P^*(s|a) + \ln(Z_\beta) - \ln Q(s|a) + \ln P(s|a)] \\ &= \operatorname{argmin}_a E_{Q(s|a)} [-\ln P^*(s|a) - \ln(Z_\beta) + \ln Q(s|a) - \ln P(s|a)] \\ & \quad \operatorname{argmin}_a D_{KL}[Q(s|a) || P^*(s|a)] \end{aligned}$$

→ (10)

In a POMDP setting analogue:

(13.5)

$$P^*(o, s|a) = \frac{P(o, s|a)e^{\beta U(o, s, a)}}{Z\beta}$$

rearrange and plug into (13)

$$= \operatorname{argmin}_a D_{KL}[Q(o, s|a) || P^*(o, s|a)]$$