Bayesian brains and the Rényi divergence

Noor Sajid[1, †], **Francesco Faccio**[2, †], **Lancelot Da Costa**[1, 3], **Thomas Parr**[1], **Jürgen Schmidhuber**[2] & **Karl Friston**[1]

[1]Wellcome Centre for Human Neuroimaging, University College London, London, UK

[2]The Swiss AI Lab IDSIA, USI, SUPSI, Lugano, Switzerland.

[3]Department of Mathematics, Imperial College London, London, UK.,

[†]Equal Contribution

Abstract

Under the Bayesian brain hypothesis, behavioural variations can be attributed to different priors over generative model parameters. This provides a formal explanation for why individuals exhibit inconsistent behavioural preferences when confronted with similar choices. For example, greedy preferences are a consequence of confident (or precise) beliefs over certain outcomes. Here, we offer an alternative account of behavioural variability using Rényi divergences, and their associated variational bounds. Rényi bounds are analogous to the variational free energy (or evidence lower bound), and can be derived under the same assumptions. Importantly, these bounds provide a formal way to establish behavioural differences through an α parameter, given fixed priors. This rests on changes in α that alter the bound (on a continuous scale), inducing different posterior estimates, and consequent variations in behaviour. Thus, it looks as if individuals have different priors, and have reached different conclusions. More specifically, $\alpha \to 0^+$ optimisation constrains the variational posterior to be positive whenever the true posterior is positive. This leads to mass-covering variational estimates and increased variability in choice behaviour. Furthermore, $\alpha \to +\infty$ optimisation constrains the variational posterior to be zero whenever the true posterior is zero. This leads to mass-seeking variational posteriors, and greedy preferences. We exemplify this formulation through simulations of the multi-armed bandit task. We note that these α parameterisations may be especially relevant, i.e., shape preferences, when the true posterior is not in the same family of distributions as the assumed (simpler) approximate density, which may be the case in many real-world scenarios. The ensuing departure from vanilla variational inference provides a potentially useful explanation for differences in behavioural preferences of biological (or artificial) agents – under the assumption that the brain performs variational Bayesian inference.

1 Introduction

The notion that the brain is Bayesian—or more appropriately, Laplacian (Stigler (1986))—and performs some form of inference has attracted enormous attention in neuroscience (Doya, Ishii, Pouget, and Rao (2007); Knill and Pouget (2004)). It takes the view that the brain embodies a model about causes of sensation, that allow for predictions about observations (Dayan, Hinton, Neal, and Zemel (1995); Hohwy (2012); Schmidhuber (1992); Schmidhuber and Heil (1995)) and future behaviour (K. Friston, FitzGerald, Rigoli, Schwartenbeck, and Pezzulo (2017); Schmidhuber (1990)). Practically, this involves the optimisation of a free energy functional (or evidence lower bound) (Bogacz (2017a); K. Friston et al. (2017); Penny (2012)), using variational inference (Blei, Kucukelbir, and McAuliffe (2017); Wainwright and Jordan (2008)), to make appropriate predictions. The free energy functional can be derived from the Kullback-Leibler (KL)-divergence (Kullback and Leibler (1951)), which measures the dissimilarity between true and approximate posterior densities. Under this formulation, behavioural variations can be attributed to altered priors over the (hyper-)parameters of a generative model, given the same (variational) free energy functional (K. Friston et al. (2014); Schwartenbeck et al. (2015)). This has been used to simulate variations in choice behaviour (FitzGerald, Schwartenbeck, Moutoussis, Dolan, and Friston (2015); K. Friston et al. (2014); K. J. Friston et al. (2015); Storck, Hochreiter, and Schmidhuber (1995)) and behavioural deficits (Sajid, Parr, Gajardo-Vidal, Price, and Friston (2020); Smith, Lane, Parr, and Friston (2019)).

Conversely, distinct behavioural profiles could be attributed to differences in the variational objective, given the same priors. In this paper, we consider this alternative account of phenotypic variations in choice behaviour using Rényi divergences (S.-i. Amari (2012); S.-i. Amari and Cichocki (2010); Phan, Abbasi-Yadkori, and Domke (2019); Rényi (1961); Van Erven and Harremos (2014)). These are a general class of divergences, indexed by an α

parameter, of which the KL-divergence is a special case. It is perfectly reasonable to diverge from this special case since variational inference does not commit to the KL-divergence (Wainwright and Jordan (2008)) (indeed, previous work has developed divergence-based lower bounds that give tighter bounds e.g., (Barber and van de Laar (1999)), yet these may be more difficult to optimise despite being better approximations). Broadly speaking, variational inference is the process of approximating a posterior probability through application of variational methods. This means finding the function (here, an approximate posterior), out of a pre-defined family of functions, that extremizes an objective functional. In variational inference, the key is choosing the objective such that the extreme value corresponds to the best approximation. Rényi divergences can be used to derive a (generalised) variational inference objective called the Rényi-bound (Li and Turner (2017)). The Rényi-bound is analogous to the variational free energy functional and provides a formal way to establish phenotypic differences despite consistent priors. This is accomplished by changes, on a continuous scale, **that give rise** to different posterior estimates, and consequent behavioural variations (Minka (2005)). Thus, changing the functional form of the bound will make it will look as if individuals have different priors i.e., have reached different conclusions from the same observations due to the distinct optimisation objective.

It is important to determine whether this formulation introduces fundamentally new differences in behaviour that cannot be accounted for by altering priors under a standard variational objective. Conversely, it may be possible to relate changes in prior beliefs to changes in the variational objective. We investigate this for a simple Gaussian system by examining the relationship between different parameterisations of the Rényi bound under fixed priors and the variational free energy under different hyper-priors. It turns out that there is no clear correspondence in most cases. This suggests that differences in behaviour caused by changes in the divergence supplement standard accounts of behavioural differences under changes of priors.

The Rényi divergences depend on an α parameter that controls the strength of the bound[1] and induces different posterior estimates. Consequently, the resulting system behaviour may vary, and point towards different priors that could have altered the variational posterior form. For this, we assume that systems (or agents) sample their actions based upon posterior beliefs, and those posterior beliefs depend on the form of the Rényi bound α parameter. This furnishes a natural explanation for observed behavioural variation. To make the link to behaviour, we assume actions are selected – based on variational estimates – that maximise the Sharpe ratio (Sharpe (1994)) i.e., a variance-adjusted return. Accordingly, evaluation of behavioural differences rests upon a separation between estimation of posterior beliefs over particular (hidden) states and the action selection criterion. That is, actions are selected given posterior estimates about states. This is contrary to other Bayesian sequential decision-making schemes -— such as active inference (Da Costa et al. (2020); K. Friston et al. (2017)) -— where actions are sampled from posterior beliefs about action sequences (i.e., policies). This effectively separates action and perception into state estimation and planning as inference[2]. However, we will use a simplification of action selection—using the Sharpe Ratio—to focus on inferences about hidden states under different values. We reserve further details for later sections.

Intuitively, under the Rényi bound, high α values lead to mass-seeking approximate[3] posteriors i.e., greedy preferences for a particular outcome. This happens because the variational posterior is constrained to be zero whenever the true posterior is zero. Conversely,

[1]Here, strength of bound refers the closeness with which the variational functional bounds the (negative) log evidence.

[2]Note that heuristics like the Sharpe Ratio are unnecessary in active inference (Da Costa et al. (2020); K. Friston et al. (2017)) – that automatically accommodates uncertainty of this sort —- however, it is a useful heuristic because it foregrounds the role of posterior uncertainty in action selection.

[3]We use approximate and variational posterior interchangeably throughout.

$\alpha \to 0^+$ can result in mass-covering approximate posteriors —- resulting in a greater range of actions for which there are plausible outcomes consistent with prior preferences. In this case, the variational posterior is constrained to be positive whenever the true posterior is positive. Hence, variable individual preferences could be attributed to differences in the variational optimisation objective. This contrasts with standard accounts of behavioural differences, where the precision of some fixed priors is used to explain divergent behaviour profiles under the same variational objective. In what follows, we present, and validate, this generalised kind of variational inference that can explain the implicit preferences of biological and artificial agents, under the assumption that the brain performs variational Bayesian inference.

The paper is structured as follows. First, we provide a primer on standard variational inference using the KL-divergence (section 2). Section 3 introduces Rényi divergences and the derivation for the Rényi bound using the same assumptions as the standard variational objective. We then consider what (if any) sort of correspondence exists between the Rényi bound and the variational free energy functional —- i.e., the evidence lower bound —- under different priors (section 4). In section 5, we validate the approach through numerical simulations of the multi-armed bandit (Auer, Cesa-Bianchi, and Fischer (2002); Lattimore and Szepesvári (2020)) paradigm with multi-modal observation distribution. Our simulations demonstrate that variational Bayesian agents, optimising a generalised variational bound (i.e., Rényi bound) can naturally account for variations in choice behaviour. We conclude with a brief discussion of future directions and the implications of our work for understanding behavioural variations.

2 Variational Inference

Variational inference is an inference scheme based on variational calculus (Parisi (1988)). It identifies the posterior distribution as the solution to an optimisation problem, allowing

otherwise intractable probability densities to be approximated (Jordan, Ghahramani, Jaakkola, and Saul (1999); Wainwright and Jordan (2008)). For this, we define a family of approximate densities over the hidden variables of the generative model (Beal (2003); Blei et al. (2017)). From this, we can use gradient descent to find the member of that variational family that minimises a divergence to the true conditional posterior. This variational density then serves as a proxy for the true density. This formulation underwrites practical applications that characterise the brain as performing Bayesian inference including predictive coding (Millidge, Tschantz, and Buckley (2020); Perrykkad and Hohwy (2020); Schmidhuber and Heil (1995); Spratling (2017); Whittington and Bogacz (2017)), and active inference (Da Costa et al. (2020); K. Friston et al. (2017); Sajid, Ball, Parr, and Friston (2021); Storck et al. (1995); Tschantz, Seth, and Buckley (2020)).

2.1 KL-divergence and the standard variational objective

To derive the standard variational objective -— known as the variational free energy, or negative evidence lower bound (ELBO) —- we consider a simple system with two random variables. These are $s \in \mathcal{S}$ denoting hidden states of the system (e.g., it rained last night) and $o \in \mathcal{O}$ the observations (e.g., the grass is wet). The joint density over these variables:

$$p(s, o) = p(o|s)p(s) \tag{1}$$

where, $p(s)$ is the prior density over states and $p(o|s)$ is the likelihood, is called the generative model. Then, the inference problem is to compute the posterior – i.e., the conditional density – of the states given the outcomes:

$$p(s|o) = \frac{p(o,s)}{p(o)}. \tag{2}$$

This quantity contains the evidence, $p(o)$, that can be calculated by marginalising out the states from the joint density. However, the evidence is notoriously difficult to compute, which makes the posterior intractable in practical applications. This problem can be finessed with variational inference[4]. For this, we introduce a variational density, $q(\cdot)$ that can be easily integrated. The following equations illustrate how we can derive the quantities of interest. We assume that both $p(s|o)$ and $q(s)$ are non-zero:

$$\log p(o) = \log p(o) + \int_{\mathcal{S}} \log \frac{p(s|o)}{p(s|o)} \, \mathrm{d}s \tag{3}$$

$$= \int_{\mathcal{S}} q(s) \log p(o) \, \mathrm{d}s + \int_{\mathcal{S}} q(s) \log \frac{p(s|o)}{p(s|o)} \, \mathrm{d}s = \int_{\mathcal{S}} q(s) \log \frac{p(s,o)}{p(s|o)} \, \mathrm{d}s \tag{4}$$

$$= \int_{\mathcal{S}} q(s) \log \frac{q(s)}{q(s)} \, \mathrm{d}s + \int_{\mathcal{S}} q(s) \log p(s,o) \, \mathrm{d}s + \int_{\mathcal{S}} q(s) \log \frac{1}{p(s|o)} \, \mathrm{d}s \tag{5}$$

$$= \underbrace{\int_{\mathcal{S}} q(s) \log \frac{1}{q(s)} \, \mathrm{d}s + \int_{\mathcal{S}} q(s) \log p(s,o) \, \mathrm{d}s}_{\text{ELBO}} + \underbrace{\int_{\mathcal{S}} q(s) \log \frac{q(s)}{p(s|o)} \, \mathrm{d}s}_{\text{KL Divergence}} \tag{6}$$

The first two summands of the last equality are the evidence lower bound (Welbourne, Woollams, Crisp, and Lambon-Ralph (2011)), and the last summand presents the KL-divergence between the approximate and true posterior. If $q(\cdot)$ and $p(\cdot)$ are of the same exponential family, then their KL divergence can be computed using the formula provided in

[4]There are other methods to estimate the posterior that include sampling-based, or hybrid approaches e.g., Markov Chain Monte Carlo (MCMC). However, variational inference is considerably faster than sampling, by employing simpler variational posteriors, which lead to a simpler optimisation procedure (Wainwright and Jordan (2008)).

(Huzurbazar (1955)). Our variational objective of interest is the free energy functional ($\mathcal{F}$) which upper bounds the negative log evidence. Therefore, we rewrite the last equality:

$$
\begin{aligned}
-\log p(o) &= -\left[\int_{\mathcal{S}} q(s)\log\frac{1}{q(s)}\,\mathrm{d}s + \int_{\mathcal{S}} q(s)\log p(s,o)\,\mathrm{d}s + \int_{\mathcal{S}} q(s)\log\frac{q(s)}{p(s|o)}\,\mathrm{d}s\right] && (7)\\
&= \int_{\mathcal{S}} q(s)\log q(s)\,\mathrm{d}s - \int_{\mathcal{S}} q(s)\log p(s,o)\,\mathrm{d}s - \int_{\mathcal{S}} q(s)\log\frac{q(s)}{p(s|o)}\,\mathrm{d}s && (8)\\
&\leq \int_{\mathcal{S}} q(s)\log q(s)\,\mathrm{d}s - \int_{\mathcal{S}} q(s)\log p(s,o)\,\mathrm{d}s && (9)\\
&= -\mathbb{E}_{q(s)}[\log p(s,o)] - \mathbb{H}[q(s)] && (10)\\
&= \underbrace{D_{KL}\big[q(s)||p(s)\big]}_{\text{complexity}} - \underbrace{\mathbb{E}_{q(s)}\big[\log p(o|s)\big]}_{\text{accuracy}} && (11)\\
&= -D_{KL}[q(s)||p(s,o)] = \mathcal{F} && (12)
\end{aligned}
$$

The second last line is the commonly presented decomposition of the variational free energy summands: complexity and accuracy (K. Friston et al. (2017); Sajid, Ball, et al. (2021)). The accuracy term represents how well observed data can be predicted, while complexity is a regularisation term. The variational free energy objective favours accurate explanations for sensory observations that are maximally consistent with prior beliefs. Additionally, the last equality defines the variational free energy in terms of a KL-divergence between $q(s)$ and $p(o,s)$. This may seem different to those used to dealing with variational free energy to see it defined in terms of a KL divergence. Since, usually this notation is reserved for arguments that are both normalised (Bishop (2006)). However, here the normalisation factors over $p(\cdot)$ become an additive constant in the KL-divergence, which has no effect on the gradients used in optimisation or inference. Contrariwise, the normalising constant of $q(\cdot)$ needs to be the same across the variational family.

In this setting, illustrations of behavioural variations i.e., differences in variational posterior

estimations can result from different priors over the (hyper-)parameters[5] of the generative model (Storck et al. (1995)), e.g., change in precision over the likelihood function (K. Friston et al. (2014)). We reserve description of hyper-priors and their impact on belief updating for section 4.

3 Rényi divergences, and their variational bound

We are interested in defining a (general) variational objective that can account for behavioural variations alternate to a change of priors. For this, we can replace the KL divergence by a general divergence objective, i.e., a non-negative function $D[\cdot||\cdot]$ that satisfies $D[q(s)||p(s|o)] = 0$ if and only if $q(s) = p(s|o)$ for all $s \in \mathcal{S}$[6]. For our purposes, we focus on Rényi divergences, a general class of divergences that includes the KL-divergence. Explicitly, we can derive the KL-divergence from the Rényi divergence as $\alpha \to 1$, e.g., using L'Hôpital's rule, or the minimum description length as $\alpha \to \infty$ (Table 1). This has the advantage of being computationally tractable, and satisfies many additional properties (S.-i. Amari (2012); Rényi (1961); Van Erven and Harremos (2014)). Rényi-divergences are defined as (Li and Turner (2017); Rényi (1961)):

$$D_\alpha\big[q(s)||p(s|o)\big] := \frac{1}{\alpha - 1} \log \int_{\mathcal{S}} q(s)^\alpha p(s|o)^{1-\alpha} \, \mathrm{d}s \tag{13}$$

[5]Note that introducing hyper-priors (or precision priors) is standard part of the Bayesian machinery (Gelman, Carlin, Stern, and Rubin (1995)). Intuitively, this involves scaling the variance over the distribution of interest to make it more or less precise (or confident). For example, a Gaussian distribution can become relatively flat (i.e., less precise) or a Dirac delta function (i.e., infinitely precise) in the limits of high and low variance, respectively.

[6]Technically, this equality holds up to a set of measure zero.

where $\alpha \in \mathbb{R}^+ \setminus \{1\}$. An analogous definition holds for the discrete case, by replacing the densities with probabilities and the integral by a sum (Rényi (1961)). This family of divergences can provide different posterior estimates as the minimum of the divergence with respect to q varies smoothly with α. These differences are possible only when the true posterior, e.g., some multi-modal distribution, is not in the same family of distributions as the approximate posterior, e.g., a Gaussian distribution. Note that other (non-Rényi) divergences in the literature are also parameterized by α, which can lead to confusion: the I divergence, Amari's α-divergence and the Tsallis divergence. All of these divergences are equivalent in that their values are related by simple formulas, see appendix A. This allows the results presented in this paper to be generalised to these divergence families using the relationships in appendix A.

3.1 Rényi bound

The accompanying variational bound for Rényi divergences can be derived using the same procedures as for deriving the evidence lower bound (Eq. 3). This gives us the Rényi bound introduced in (Li and Turner (2017)):

$$p(o) = \frac{p(o,s)}{p(s|o)} \implies \tag{14}$$

$$p(o)^{1-\alpha}p(s|o)^{1-\alpha} = p(o,s)^{1-\alpha} \tag{15}$$

$$\int_{\mathcal{S}} q(s)^{\alpha}p(o)^{1-\alpha}p(s|o)^{1-\alpha}\,\mathrm{d}s = \int_{\mathcal{S}} q(s)^{\alpha}p(o,s)^{1-\alpha}\,\mathrm{d}s \tag{16}$$

$$\log \int_{\mathcal{S}} q(s)^{\alpha}p(o)^{1-\alpha}p(s|o)^{1-\alpha}\,\mathrm{d}s = \log \int_{\mathcal{S}} q(s)^{\alpha}p(o,s)^{1-\alpha}\,\mathrm{d}s \tag{17}$$

$$\log p(o)^{1-\alpha} + \log \int_{\mathcal{S}} q(s)^{\alpha}p(s|o)^{1-\alpha}\,\mathrm{d}s = \log \int_{\mathcal{S}} q(s)^{\alpha}p(o,s)^{1-\alpha}\,\mathrm{d}s \tag{18}$$

$$\log p(o)^{1-\alpha} = \log \int_{\mathcal{S}} q(s)^{\alpha}p(o,s)^{1-\alpha}\,\mathrm{d}s - \log \int_{\mathcal{S}} q(s)^{\alpha}p(s|o)^{1-\alpha}\,\mathrm{d}s \tag{19}$$

$$\log p(o) = \underbrace{\frac{1}{1-\alpha}\log \int_{\mathcal{S}} q(s)^{\alpha}p(o,s)^{1-\alpha}\,\mathrm{d}s}_{\text{Rényi Bound}} + \underbrace{\frac{1}{\alpha-1}\log \int_{\mathcal{S}} q(s)^{\alpha}p(s|o)^{1-\alpha}\,\mathrm{d}s}_{\text{Rényi Divergence}} \tag{20}$$

$$\log p(o) = -D_{\alpha}[q(s)||p(o,s)] + D_{\alpha}[q(s)||p(s|o)] \tag{21}$$

We assume that $q(s)$ and $p(s|o)$ are non-zero and $\alpha \in \mathbb{R}^{+} \setminus \{1\}$. Additionally, we are licensed to make the move from Eq.17 to Eq.18 because $p(o)$ does not depend on s. The negative Rényi bound can be regarded as being analogous to the variational free energy objective ($\mathcal{F}$) by providing an upper bound to the negative log evidence (Eq. 7):

$$-\log p(o) = \frac{1}{\alpha-1}\log \int_{\mathcal{S}} q(s)^{\alpha}p(o,s)^{1-\alpha}\,\mathrm{d}s - \frac{1}{\alpha-1}\log \int_{\mathcal{S}} q(s)^{\alpha}p(s|o)^{1-\alpha}\,\mathrm{d}s \tag{22}$$

$$\leq \frac{1}{\alpha-1}\log \int_{\mathcal{S}} q(s)^{\alpha}p(o,s)^{1-\alpha}\,\mathrm{d}s = D_{\alpha}[q(s)||p(o,s)] \tag{23}$$

Similar to the Rényi divergence, we expect variations in the estimation of the approximate posterior with α under the Rényi bound. Explicitly, when $\alpha < 1$ the variational posterior will aim to cover the entire true posterior— this is known as exclusivity (or zero-avoiding) property. Thus, $\alpha \to 0^{+}$ optimisation constrains the variational posterior to be positive whenever the true posterior is positive. Formally, for all $s : p(s,o) > 0 \Rightarrow q(s) > 0$. This

leads to mass-covering variational estimates and increased variability. Furthermore, $\alpha \to +\infty$ optimisation constrains the variational posterior to be zero whenever the true posterior is zero. Here, the variational posterior will seek to fit the true posterior at its mode— this is known as inclusivity (or zero-forcing) mode-seeking behaviour (Li and Turner (2017)) In this case, for all $s : p(s, o) = 0 \Rightarrow q(s) = 0$. This leads to mass-seeking variational posteriors. Hence, the Rényi bound should provide a formal account of behavioural differences through changes in the α parameter. That is, we would expect a natural shift in behavioural preferences as we move from small values to large positive α values, given fixed priors. Section 5 demonstrates this shift in preferences in a multi-armed bandit setting.

α	Rényi Divergence $D_\alpha[q(s)\|\|p(s\|o)]$	Rényi Bound $-D_\alpha[q(s)\|\|p(s,o)]$	Comment
$\alpha \to 1$	$\int_{\mathcal{S}} q(s) \log \frac{q(s)}{p(s\|o)} \, \mathrm{d}s$	$-D_{KL}[q(s)\|\|p(s)] + \mathbb{E}_{q(s)} \log p(o\|s)$ $-\mathbb{H}[p(s,o)] + \mathbb{E}_{p(s,o)} \log q(s)$	Kullback-Leibler (KL) divergence: $D_{KL}[q\|\|p]$ or $D_{KL}[p\|\|q]$
$\alpha = 0.5$	$-2\log(1 - Hel^2(p(s\|o), q(s)))$ $-2 \log \sqrt{p(s\|o)q(s)} \, \mathrm{d}s$	$2\log(Hel^2(p(s,o), q(s)))$ $2 \log \sqrt{p(s,o)q(s)} \, \mathrm{d}s$	Function of the Hellinger distance or the Bhattacharyya divergence. Both are symmetric in their arguments
$\alpha = 2$	$\log \left[1 + \chi^2[q(s)\|\|p(s\|o)]\right]$	$-\log \left[1 + \chi^2[q(s)\|\|p(s,o)]\right]$	Proportional to χ^2-divergence: $\chi^2(q,p) = \int_{\mathcal{S}} \frac{q^2}{p} \, \mathrm{d} - 1$
$\alpha \to \infty$	$\log max_{s \in S} \frac{q(s)}{p(s\|o)}$	$-\log max_{s \in S} \frac{q(s)}{p(s,o)}$	Minimum description length

Table 1: Examples of (normalised) Rényi divergences (Li and Turner (2017); Minka (2005); Van Erven and Harremos (2014)) for different values of α, and the accompanying Rényi bounds. We omit $\alpha \to 0$ because the limit is not a divergence. These divergences have a non-decreasing order i.e., $Hel^2(q,p) \leq D_{\frac{1}{2}}[q||p] \leq D_1[q||p] \leq D_2[q||p] \leq \chi^2(q,p)$ (Van Erven and Harremos (2014)).

4 Variational bounds, precision, and posteriors

It is important to determine whether this formulation of behaviour introduces fundamentally new differences that cannot be accounted for by altering the priors under a standard variational

objective. Thus, we compare the Rényi bound and the variational free energy on a simple system to see whether the same kinds of inferences can be produced through the Rényi bound (Eq. 15) with fixed prior beliefs but altered α value and through the standard variational objective (Eq. 3) with altered prior beliefs. If this were to be the case, we would be able to re-write the variational free energy under different precision hyper-priors as the Rényi bound, where hyper-parameters now play the role of the α parameter. If this correspondence holds true, the two variational bounds (i.e., Rényi and variational free energy) would share similar optimisation landscapes (i.e., inflection or extrema), with respect to the posterior under some different priors or α value.

Variations in these hyper-priors speak to different priors, under which agents can exhibit conservative or greedy choice behaviour. Practically, this may be a result of either (i) lending one contribution more precision through weighting the log probability under the standard variational objective, or (ii) by altering the priors by taking the log of the probability to the power of α. To illustrate this equivalence, we consider the following systems (Fig 1). First, we formulate a Gaussian-Gamma system to derive the analytical (exact) form of the variational free energy. Here, the system is Gaussian with Gamma priors over the variance that allows us to alter prior beliefs. A Gamma prior is necessary to model an unknown variance. Next, we introduce a system with a simple Gaussian parameterisation to derive the analytical form of the Rényi bound. The difference in parameterisation is required to establish whether changes in prior beliefs (or precision) are equivalent to the α parameter. In other words, this formulation allows us to ask whether one can either alter the precision prior or the α value to evince behavioural differences. If this were the case, we would expect equivalences between the two analytical bounds, given the different parameterisations.

Though the problem setting is simple, it provides an intuition of what (if any) sort of correspondence exists between the Rényi bound and the variational free energy functional using different priors.

4.1 Variational free energy for a Gaussian-Gamma system

To derive the variational free energy, we consider a simple system with two random variables: $s \in S$ denoting (hidden) states of the system, $o \in O$ the observations (Figure 1 (A)). λ_k is the precision parameter, Σ_k is the covariance and x the parameter governing the mean. The variational family is parameterised as a Gaussian. This is formalised as:

$$p(s, \lambda_p) = \mathcal{N}(s; 0, (\lambda_p \sigma_p)^{-1}) Gam(\lambda_p; \alpha_p, \beta_p) \tag{24}$$

$$p(o|s) = \mathcal{N}(sx, \Sigma_l) \tag{25}$$

$$q(s) = \mathcal{N}(\mu_q, \Sigma_q) \tag{26}$$

where $\Sigma_k = (\lambda_k \sigma_k)^{-1}$, s are scalars, o has dimension n, and x has dimensionality $n \times 1$. Here, Σ_l represents the covariance over the likelihood, and Σ_k the general covariance where $k \in (p, l, q)$. In Eq. 24, $\mu_p = 0$ and has been written as such. Additionally, Eq. 24 denotes the joint probability distribution over $p(s, \lambda_p) = p(s|\lambda_p)p(\lambda_p)$ (Bishop (2006); Murphy (2007)).

We use these quantities to derive the variational free energy (Appendix B for the derivation):

$$-D_{KL}[q(s)||p(s, o)] = \frac{1}{2} \log \left(\frac{|\Sigma_q|}{(2\pi)^n |\Sigma_p||\Sigma_l|} \right) \tag{27}$$

$$- \frac{1}{2} \left(o^T \Sigma_l^{-1} o + \mu_q^2 \Sigma_p^{-1} + \mu_q^2 x^T \Sigma_l^{-1} x - 2\mu_q x^T \Sigma_l^{-1} o \right) \tag{28}$$

$$- \frac{1}{2} \left(\Sigma_q x^T \Sigma_l^{-1} x + \Sigma_q \Sigma_p^{-1} - 1 \right) \tag{29}$$

$$- \log \frac{\lambda_p^{\alpha_p - 1} \beta_p^{\alpha_p}}{\Gamma(\alpha_p)} - \lambda_p \beta_p \tag{30}$$

Here, Eq. 30 are the additional terms introduced via the Gamma prior.

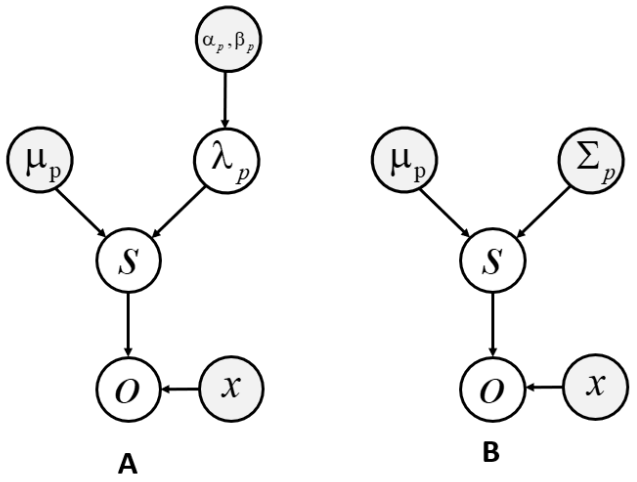


Figure 1: Graphical model for the Gaussian-Gamma (A), and Gaussian (B) system. Here white circles represent random variables, grey circles represent priors and x is the parameter governing the mean. The difference between these models is that in model (A), the precision parameters over hidden states λ_p are random variables that follow a Gamma distribution with parameters α_p, β_p, while in model (B) the precision is held fixed. Here, the scalar parameter σ_p has been deliberately omitted from the figure.

4.2 Rényi bound for a Gaussian system

Next, we consider a similar system for deriving the Rényi-bound. Unlike for the system in Section 4.1 the densities are parameterised as a Gaussian distribution (Figure 1 (B)):

$$p(s) = \mathcal{N}(0, \Sigma_p) \tag{31}$$

$$p(o|s) = \mathcal{N}(sx, \Sigma_l) \tag{32}$$

$$q(s) = \mathcal{N}(\mu_q, \Sigma_q) \tag{33}$$

where s is a scalar, o has dimension n, and x has dimensionality $n \times 1$. Additionally, $\mu_p = 0$ and has been written as such. We use these quantities to derive the Rényi bound (Appendix B for the derivation):

$$-D_\alpha[q(s)||p(s,o)] = \frac{1}{2}\log\left(\frac{|\Sigma_q|}{(2\pi)^n|\Sigma_p||\Sigma_l|}\right) \tag{34}$$

$$- \frac{\alpha}{2(\Sigma_q\Sigma_\alpha^{-1})}\left(o^T\Sigma_l^{-1}o + \mu_q^2\Sigma_p^{-1} + \mu_q^2 x^T\Sigma_l^{-1}x - 2\mu_q x^T\Sigma_l^{-1}o\right) \tag{35}$$

$$- \frac{1}{2(1-\alpha)}\log\left(1 + (1-\alpha)(\Sigma_q x^T\Sigma_l^{-1}x + \Sigma_q\Sigma_p^{-1} - 1)\right) \tag{36}$$

$$- \frac{1}{2\Sigma_\alpha^{-1}}\left((1-\alpha)\Sigma_p^{-1}o^T\Sigma_l^{-1}o\right) \tag{37}$$

where, $\Sigma_\alpha := \left((1-\alpha)(\Sigma_p^{-1} + x^T\Sigma_l^{-1}x) + \alpha\Sigma_q^{-1}\right)^{-1}$, under the assumption that Σ_α is positive-definite. Since Σ_α is a scalar, this is equivalent to satisfying the following condition: $\Sigma_\alpha \succ 0 \iff (\alpha-1)(\Sigma_p^{-1} + x^T\Sigma_l^{-1}x)\Sigma_q < \alpha$. Importantly, if $\alpha \leq 1$ the condition is always true for any choice of Σ_q. However, for $\alpha > 1$ we must impose $\Sigma_q < \frac{\alpha}{\alpha-1}\frac{\Sigma_p}{1+\Sigma_p x^T\Sigma_l^{-1}x} = \frac{\alpha}{\alpha-1}\mathrm{Cov}(p(s|o))$ (Burbea (1984); Metelli, Papini, Faccio, and Restelli (2018)).

4.3 Correspondence between variational free energy & the Rényi bound

Using the derived bounds above, we examine the correspondence between the variational free energy and the Rényi bound.

First, we consider the case when $\alpha \to 1$. Here, we expect to find an exact correspondence between the variational free energy and the Rényi bound as the Rényi divergence tends towards the KL-divergence as $\alpha \to 1$. Our derivations confirm this, upon comparison of the equivalent terms for each objective. The first terms in each objective, Eq. 27 and Eq. 34 are the same. Interestingly, the second term in the Rényi bound (Eq. 35) is a scalar multiple of the second term in variational free energy (Eq. 28), where the scalar quantity $\frac{\alpha}{\Sigma_q \Sigma_\alpha^{-1}}$ tends to 1 for $\alpha \to 1$. The third term in Eq. 36, for $\alpha \to 1$, is a limit of the form $\lim_{x \to 0} \frac{1}{x} \log(1 + xw) = w$, resulting exactly in Eq. 29. Finally, the last term in the Rényi bound tends to zero as $\alpha \to 1$ (Eq. 37).

Next, we evaluate the correspondence between the variational free energy and Rényi bound when $\alpha \in \mathbb{R}^+ \setminus \{1\}$. Now, the α values scale the terms in the Rényi bound with Eq. 37 having an influence on the final bound estimate. For comparability, we introduced the Gamma prior to a simple Gaussian system. As shown in Eq. 30, this introduces additional terms that scale the free energy $\mathcal{F}$. We expect the scaling from the α parameter to have some correspondence to the precision priors in the Gaussian-Gamma system. To assess this, we plot the variational objectives as a function of their estimated sufficient statistics for this simple system (Fig. 2). The numerical simulation illustrates that optimisation of these objectives, for appropriate priors (α_p, β_p) or the α value, can lead to (extremely) different variational densities.

Interestingly, the two variational objectives do exhibit a similar optimisation landscape under specific parameterisations. For example, a striking (local) minimum of -33.14 nats is observed when α_p is approximately 1, β_p is greater than 0.8 and $\alpha < 5$. However, this is constrained to a small space of posterior μ_q estimates. Outside these posterior parameters,

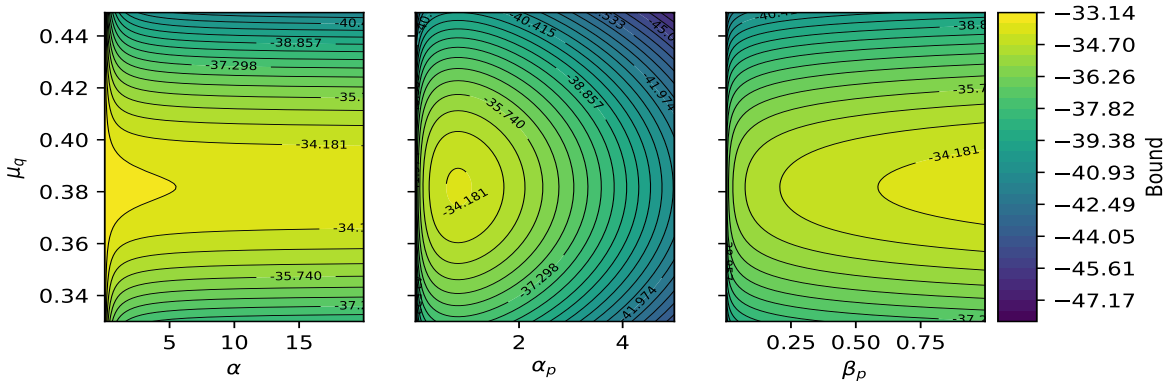
the optimisation landscape differs. Importantly, this difference becomes more acute when considering σ_q. This suggests hyper-priors may be particularly important in shaping the correspondence between the two variational objectives. However, the optimisation profile can differ under inappropriate priors (i.e., a misalignment between prior beliefs and α value), and lead to divergences in the estimated variational density (Fig. 2).

Briefly, we do not observe a direct correspondence in the optimisation landscapes (and the variational posterior) for certain priors or α value. These numerical analyses demonstrate that the Rényi divergences account for behavioural differences in a way that is formally distinct from a change in priors, through manipulation of the α parameter. Conversely the standard variational objective could require multiple alterations to the (hyper-)parameters to exhibit a similar functional form in some cases. Further investigation in more complex systems is required to quantify the correspondence (if any) between the two variational objectives.

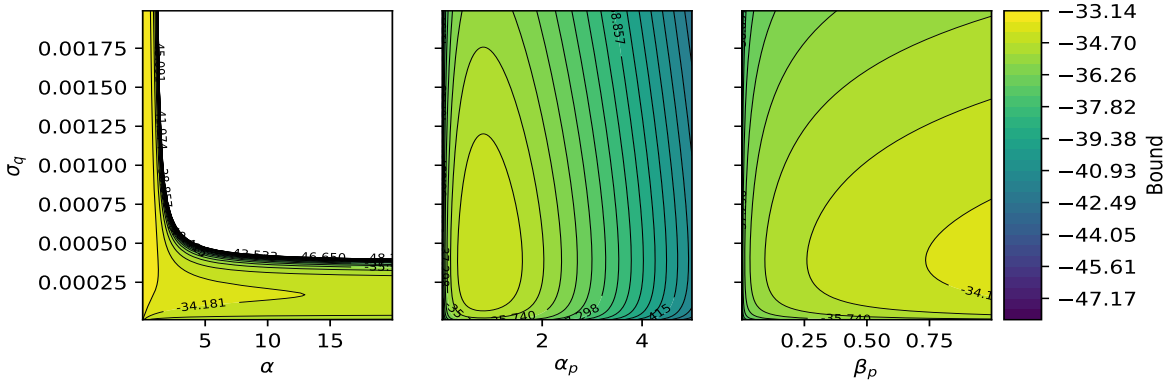
5 Multi-armed bandit simulation

In this section, we illustrate the differential preferences that arise naturally under the Rényi bound. For this, we simulated the multi-armed bandit (MAB) paradigm (Auer et al. (2002); Lattimore and Szepesvári (2020)) using 3 arms. The MAB environment was formulated as a one-state Markov Decision Process (MDP) i.e., the environment remains in the same state independently of agents' actions. At each time-step t, the agent could pull one arm and a corresponding outcome (i.e., score) R_t was observed. The agent's objective was to identify, and select, the arm with the highest Sharpe ratio (Sharpe (1994)) through its interactions with the environment across X trials.

The Sharpe ratio is a well-known financial measure for risk-adjusted return. It is an appropriate heuristic for action selection because it measures the expected return after adjusting for the variance of return distribution i.e., return to variability ratio. In particular,



(a)



(b)

Figure 2: Heat map of variational bounds as a function of estimated sufficient statistics: μ_q (a) and σ_q (b). These graphics plot the optimisation landscape for changing priors or α values. The first column plots the Rényi bound, as a function of α on the x-axis and μ_q (a) or σ_q (b) on the y-axis. Similarly, the next two columns plot the free energy, as a function of α_p (center column) or β_p (right column) on the x-axis and μ_q (a) or σ_q (b) on the y-axis. The variational bound range from -33 (yellow) to -47 nats (blue). The empty region in (b) for different α values in the Renyi bound is a consequence of the (positive definiteness) constraint imposed on Σ_q for $\alpha > 1$ restricting the possible values to be $< \frac{\alpha}{\alpha-1}\frac{\Sigma_p}{1+\sigma_p x^T \sigma_l^{-1} x}$. When not varying, hyper-parameters are fixed with $\mu_q = 0.4$, $\sigma_q = 1e-4$, $\alpha_p = 0.8$, $\beta_p = 0.8$, $\lambda_p = 0.8$, $x = \{r : r = 1.1 \times n\ ,\ \ n \in \{0, 1, \ldots 19\}\}$, $y = 0.4 \times x$, $\Sigma_l = \mathbb{I}_{20}$.

given the expected return of an arm $R = \mathbb{E}[R_t]$, the Sharpe ratio is defined as $SR := \frac{\mathbb{E}[R_t]}{\mathbb{V}[R_t]}$ – where $\mathbb{V}[R_t]$ is the variance of return distribution for a specific arm. This heuristic was chosen because it nicely illustrates how changes in α influence the sufficient statistics of the variational posterior, and ensuing behaviour. Practically, this means we sample from the posterior distribution for each state (i.e., arm) and select actions that maximise the Sharpe ratio. The Sharpe ratio affords an action selection criterion that accommodates posterior uncertainty about hidden states, which underwrites choice behaviour. For example, posterior estimates for some (sub-optimal) arms may have high variance, meaning the expected reward is obtained with less certainty. If actions were selected to sample from the arm with the highest reward, then sub-optimal arms with uncertain payoff may be selected with unduly high probability. The Sharpe Ratio precludes this, penalising arms with high posterior uncertainty.

We modelled each arm with a fixed multi-modal distribution (a mixture of Gaussians) unknown to the agent; characterising this as stationary stochastic bandit setting. Explicitly, this entailed the following parameterisation for each arm:

$$p(s) = \sum_{i}^{2} \omega_i \mathcal{N}(\mu_i, \Sigma_i) \tag{38}$$

$$p(o|s) = \mathcal{N}(s, 1.0) \tag{39}$$

$$q(s) = N(\mu_q, \Sigma_q) \tag{40}$$

$$\sum_{i}^{2} \omega_i = 1, \omega_i > 0 \tag{41}$$

where, s denotes the hidden state over the arm distribution and o the observed return (R) from an arm. The variational density $q(s)$ was constrained as a Gaussian with an arbitrary mean and variance, under a mean-field assumption[7]. However, due to the multi-modal prior,

[7]That is a fully factorised variational distribution. For further details see (Minka (2005);

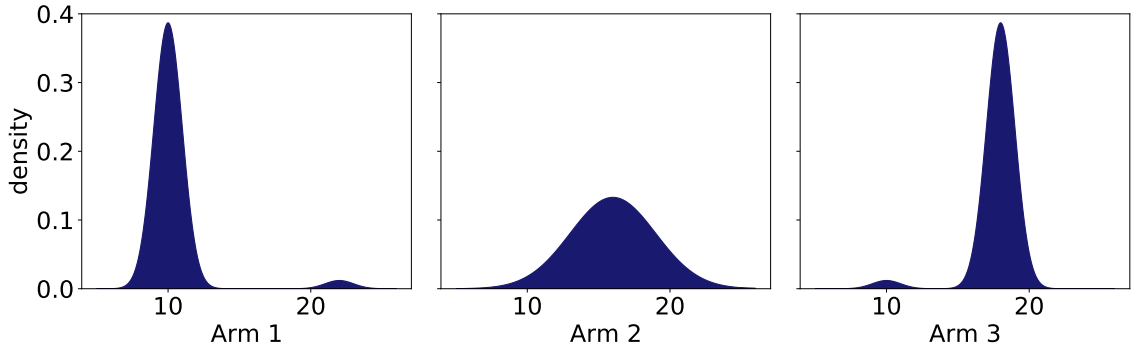
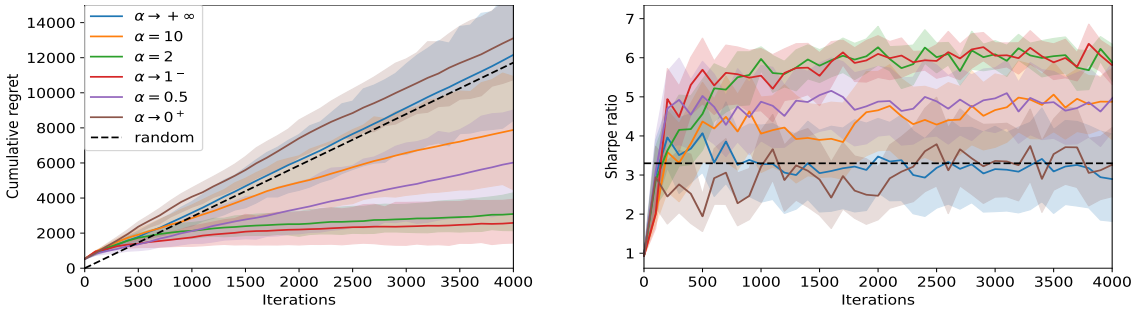


Figure 3: Score distribution for each arm. The figures plot the score distributions for each arm. The x-axis is the $s \sim q(s)$ and y-axis the score density. Arm 1 has a multi-modal distribution of $\mu_1^1 = 10$ ($\Sigma_1^1 = 1$) and $\mu_2^1 = 22$ ($\Sigma_2^1 = 1$) with $\omega_1^1 = 0.97$ and $\omega_2^1 = 0.03$, respectively. Arm 2 has a Gaussian distribution with $\mu_1^2 = 16$ ($\Sigma_1^2 = 3$), and Arm 3 has a multi-modal distribution of $\mu_1^3 = 18$ ($\Sigma_1^3 = 1$) and $\mu_2^3 = 10$ ($\Sigma_2^3 = 1$) with $\omega_1^3 = 0.97$ and $\omega_2^3 = 0.03$, respectively.

the true posterior could take a complex form that might not be in the variational family of distributions. This introduces differences in posteriors that are evident under different Rényi bounds. In Fig. 3, we show the true distribution for each arm that is unknown to the agent. The Sharpe ratio for arm 1 was $SR = 2.03$; arm 2 was $SR = 1.76$; and arm 3 was $SR = 6.20$. Thus, arm 3 was the best choice in our paradigm as the arm with the maximal Sharpe ratio. Accordingly, we measured performance using accumulated regret, $\mathcal{R}$, defined as: $\mathcal{R} = \sum_{t=1}^{X}(SR^* - SR_t)$. Here, SR^* is the maximal Sharpe ratio from arm 3, and SR_t the Sharpe ratio for the arm pulled at iteration t.

Optimising the Rényi bound under different α values led to varying posterior estimates and accompanying behavioural differences manifested by distinct arm choices. To show this,

Parr, Sajid, and Friston (2020); Sajid, Convertino, and Friston (2021))



(a) (b)

Figure 4: Regret (a) and Sharpe ratio (b) under the Rényi bound. (a) The line plot illustrates the cumulative regret across the 4000 iterations for each agent optimising a particular Rényi bound. The x-axis denotes the iteration and y-axis the accompanying cumulative regret. (b) The line plot illustrates the average achieved Sharpe ratio of an agent across the 4000 iterations, for each particular Rényi bound. The x-axis denotes the iteration and y-axis the Sharpe ratio. Here, blue is for agents optimising Rényi Bound for $\alpha \to +\infty$, orange for $\alpha = 10$, green for $\alpha = 2$, red for $\alpha \to +1^{-}$, purple for $\alpha = 0.5$ and brown for $\alpha \to 0^{+}$. Dashed black line represents regret under a random policy (i.e., any arm). Each agent was simulated 20 times (95% confidence interval). In our simulations, the agents with $\alpha \to +1^{-}$ and $\alpha = 2$ obtained the best performance.

we simulated 6 agents optimising the Rényi bound for distinct α values: $\to +\infty, 10, 2, \to 1^{-}, 0.5, \to 0^{+}$ – across 4000 iterations, repeated 20 times for each agent. Throughout, the agents selected an arm according to the following strategy. At each iteration, the Sharpe ratio (Sharpe (1994)) was calculated for each arm by dividing a sampled point from the estimated posterior with its variance. The arm with the highest Sharpe ratio was pulled. Formally, we sample one $s_i \sim q(\cdot|\mu_q^i, \Sigma_q^i)$ for each arm i and pull arm:

$$i^* = \arg\max_i \frac{s_i}{\Sigma_q^i} \,, \tag{42}$$

where Σ_q^i is the variance of the variational posterior for arm i. In this setting, we sampled from the posterior to calculate the Sharpe ratio instead of using the parameter μ_q optimised under each bound. This avoided premature convergence to sub-optimal policies that selected the greedy arm – and therefore encouraged exploration.

In contrast with Section 4.2, for these simulations we do not compute the analytical expression for the Rényi bound. Instead, at each iteration we used 300 Monte-Carlo samples to estimate the gradient of the bound which would otherwise be intractable for a multi-modal distribution. Practically, we employed sampling to estimate the gradient updates. This necessitates a stochastic gradient descent method, where – at each iteration -— the Monte-Carlo samples were used to calculate the posterior estimate (as introduced in (Li and Turner (2017))). For this, we used ADAM (as implemented in pytorch (Paszke et al. (2019))) as the optimizer because it is known to adequately escape local minima during optimization. However, other optimisation strategies could be used here (e.g., Momentum or RMSProp (Soydaner (2020))). Additionally, for each arm there was a separate memory buffer and optimisation process. The agent learnt the score distribution through the memory buffer that stored the previous 1000 observations. At each iteration the observations in memory were used to optimise the variational posterior estimate. We then selected the appropriate arm by

sampling the variational posterior estimate, at each iteration, for each arm and using it to compute a sample estimate of the Sharpe Ratio. This provided an adequate trade-off between exploration and exploitation. Appendix C provides further experimental details.

The only variable varying across simulations was the α parameter. To assess the performance of each α we plot the accumulated regret, and the accompanying Sharpe ratio in Fig. 4. We observe that optimising $\alpha \to +1^{-}; 2$ leads to the lowest cumulative regret and a high Sharpe ratio. Conversely, optimising $\alpha \to 0^{+}; \to +\infty$; leads to the highest cumulative regret and lowest Sharpe ratio.

To investigate this further, we plot the variational bounds for arm 1 under different α parameters (Fig. 5). Recall from Fig. 3 that if the variational posterior fits the right-hand-side mode, this results in sub-optimal arm selection and the highest regret. This is because the agent would wrongly infer a high Sharpe ratio for this particular arm — while it is in fact low — increasing the probability that it was selected. We can explain the high regret of agents with $\alpha \to +\infty; \to 0^{+}$ from the property of their variational bound: For agents optimising $\alpha \to +\infty$, the approximate posterior fit the right-hand-side mode of the distribution due to its lower variance (i.e., mode-seeking behaviour). Conversely, agents with $\alpha \to 0^{+}$ would exhibit mass-covering, high-variance posterior estimates. In contrast, agents optimising $\alpha \to 1^{-}; 0.5$ covered the left-hand-side mode and thus estimated a lower Sharpe ratio for this particular arm, which decreased the probability of it being selected (Fig 5).

These numerical experiments suggest that if agents sample their actions from posterior beliefs about what they are sampling, and those posterior beliefs depend on the form of the Rényi bound α parameterisation, then there is a natural space and explanation for behavioural variations. In short, the shape of the posterior — that underwrites ensuing behaviour — depends sensitively on the functional form of the variational bound.

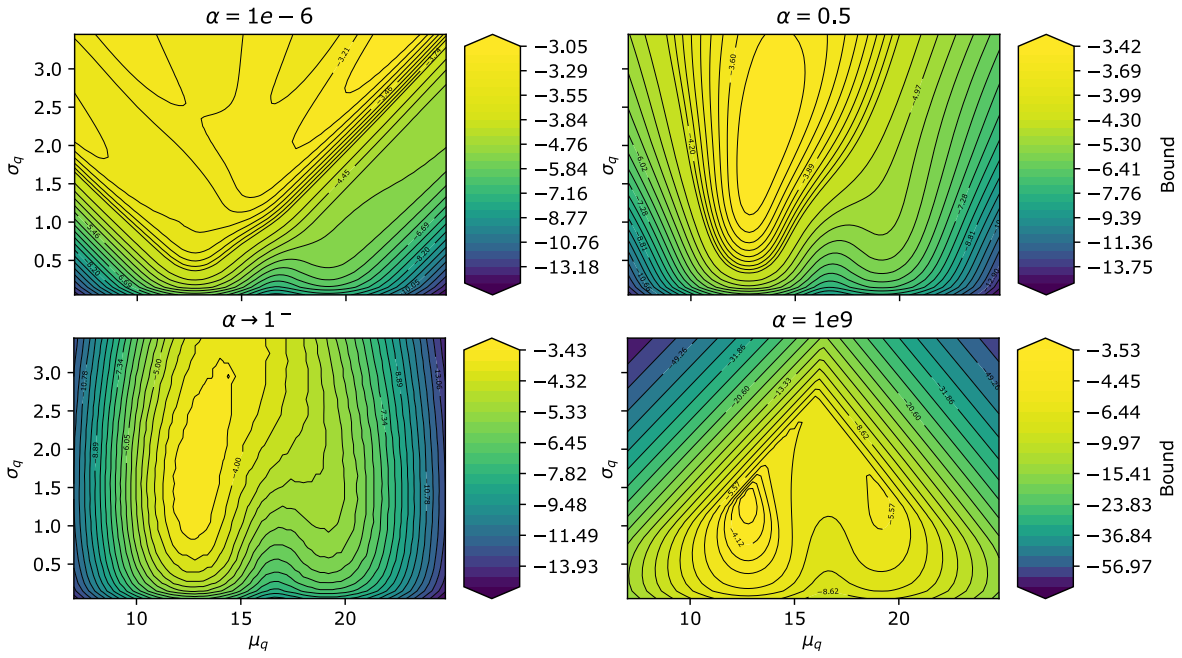


Figure 5: The Rényi bound as a function of the variational posterior. The contour plots show the optimisation landscape for each α. For $\alpha = 1e9$ we observe two optima; for small α $(1e-6)$ the optimal solution exhibits high variance.

6 Discussion

This paper accounts for behavioural variations among agents using Rényi divergences and their associated variational bounds. These divergences are Rényi relative entropies[8], and satisfy similar properties as the KL divergence (Rényi (1961); Van Erven and Harremos (2014)). Rényi divergences depend on an α parameter that controls the strength of the bound and induces different posterior estimates about the state of the world. In turn, different beliefs about the world lead to differences in behaviour. This provides a natural explanation as to why some people are more risk averse than others. For this alternative account to hold, we assumed throughout that agents sample their actions from posterior beliefs about the world, and those posterior beliefs depend on the form of the Rényi bound's α parameter. Yet, note that a similar account is possible if actions depended upon an expected free energy functional (K. Friston et al. (2017); Han, Doya, and Tani (2021); Parr and Friston (2019); van de Laar, Senoz, Özçelikkale, and Wymeersch (2021)), intrinsic reward (Schmidhuber (1991, 2006); Storck et al. (1995); Sun, Gomez, and Schmidhuber (2011)) or any class of objective functions that incorporates beliefs about the environment.

This space of Rényi bounds can provide different posterior estimates (and consequent behavioural variations) that vary smoothly with α. As illustrated, in the bi-modal scenario under our Rényi divergence definition, large positive α values will approximate the mode with the largest mass. This happens because $\alpha \geq 1$ forces the approximate posterior to be small (i.e., $q(\cdot) = 0$), whenever the true posterior is small (i.e., zero-forcing). This causes parts of the true posterior to be excluded, i.e., those parts with small total mass. Thus, the estimated variational posterior might be underestimated. Conversely, with small α values, the approximation tries to cover the entire distribution, eventually forming an upper bound when

[8]The Rényi entropy provides a parametric family of measures of information (Rényi (1961))

$\alpha \to 1$ (Table 1). This happens because $\alpha \to 1$ forces the approximate posterior to be positive (i.e., $q(\cdot) > 0$) whenever the true posterior is positive (i.e., zero-avoiding). This implies that all parts of the true posterior are included, and the variational posterior may be overestimated.

Crucially, Rényi divergences account for posterior differences in a way that is formally distinct from a change in prior beliefs. This stems from the ability to disentangle different preference modes by varying the bound's α parameter. Explicitly, we demonstrate that the Rényi bounds influences the posterior estimate over particular states (i.e., inference procedure). However, by selecting actions based on these inferences, the Rényi parameterisation shapes the preferences of the model. We observe this in our simple multi-armed bandit setting where large α values seek to fit the posterior modes that lead to greater consistency in preferences over which arm to select. Conversely, small α values try to cover the posterior distribution that led to greater flexibility over the choice of arm.

This contrasts with formal explanations based upon adjusting the precision or form of the prior under a variational bound based upon the KL divergence (i.e., $\alpha = 1$). Under active inference Da Costa et al. (2020); K. Friston et al. (2017), multiple behavioural deficits have been illustrated by manipulation of the precision over the priors (Parr and Friston (2017); Sajid et al. (2020)). Although there has been some focus upon priors and on the form of the variational posterior (Schwöbel, Kiebel, and Marković (2018)), relatively little attention has been paid to the nature of the bound itself in determining behaviour.

6.1 Implications for the Bayesian brain hypothesis

Our work is predicated on the idea that the brain is Bayesian and performs some sort of variational inference to infer its environment from its sensations. Practically, this entails the optimisation of a variational functional to make appropriate predictions. However, there are no unique functional forms for implementing such systems, and what variables account for differences in observed behaviour. On the basis of the above, we appeal to Rényi bounds,

in addition to altered priors, to model behavioural variations. By committing to the Rényi bound, we provide an alternative perspective on how variant (or sub-optimal) behaviour can be modelled. This leads to a conceptual reversal of the standard variational free energy schemes, including predictive processing, etc (Bogacz (2017b); Buckley, Kim, McGregor, and Seth (2017)). That is, we can illustrate behavioural variations to be due to different variational objectives given particular priors, instead of different priors given the variational free energy. This has implications for how we model implementations of variational inference in the brain. That is, do we model sub-optimal inferences using altered generative models or alternative variational bounds? This turns out to be significant in light of our numerical analysis (section 4.3) that show no formal correspondence between these formulations.

In a deep temporal system like the brain, one might ask if different cortical hierarchies might be performing inference under different variational objectives. It might be possible that variational objectives for lower levels to be modulated by higher levels through priors over α values – a procedure of meta-inference. This is analogous to including precision priors over model parameters that have been associated with different neuromodulatory systems e.g., state transition precision with noradrenergic and sensory precision with cholinergic systems (Fountas, Sajid, Mediano, and Friston (2020); Parr and Friston (2017)). Consequently, this temporal separation of α parameterisations may provide an interesting research avenue for understanding the role of neuromodulatory systems and how they facilitate particular behaviours (Angela and Dayan (2002, 2005)).

6.2 Generalised variational inference

The Rényi bound provides a generalised variational inference objective derived from the Rényi divergence. This is because Rényi divergences comprise the KL divergence as a special case (Minka (2005)). These divergences allow us to naturally account for multiple behavioural preferences, directly via the optimisation objective, without changing prior beliefs. Other

variational objectives can be derived from other general families of divergences such as f-divergences, Wasserstein distances (Ambrogioni et al. (2018); Dieng, Tran, Ranganath, Paisley, and Blei (2016); Regli and Silva (2018)), etc., which can improve the statistical properties of the variational bounds for particular applications (Wan, Li, and Hovakimyan (2020); Zhang, Bird, Habib, Xu, and Barber (2019)). Future work could generalise the arguments presented here and examine how these different divergences shape behaviour when planning as inference.

6.3 Limitations and future directions

We do not observe a direct correspondence between the Rényi bound and the variational free energy under particular priors. However, our evaluations are based on a restricted Gaussian system. Therefore, future work should investigate this in more complex systems to show what sorts of prior modifications are critical in establishing similar optimisation landscapes for different variational bounds, in order to understand the relationship between the two. This will entail further exploring the association between the variational posterior and β or α value.

Implementations of the Rényi bound are constrained by sampling biases and interesting differences in optimisation landscape. Indeed, when α is extremely large, even if the approximate posterior distribution belongs to the same family as the true posterior the optimisation becomes very difficult, causing the bound to be too conservative and introduce convergence issues. However, it must be noted that instances of this are due to the numerics of optimising the Rényi bound, rather than a failure of the bound itself. Practically, this means that careful consideration needs to be given to both the learning rate and stopping procedures during the optimisation of the Rényi bound.

Our work includes implicit constraints on the form of the variational posterior. We have assumed a mean-field approximation in our simulations. However, this does not necessarily have to be the case. Interestingly, richer parameterisations of the variational posterior might

negate the impact of the α values. Specifically, we noted that if the true posterior is in the same family of distributions as the variational posterior, then changing the α value does not impact the shape of the variational posterior — and consequently the system's behaviour. However, complex parameterisations are computationally expensive and can still be inappropriate. Therefore, this departure from vanilla variational inference provides a useful explanation for different behaviours that biological (or artificial) agents might adopt — under the assumption that the brain performs variational Bayesian inference. Orthogonal to this, an interesting future direction is investigating the connections between the variational posterior form and how it may impact the variational bound. This has direct consequences for the types of message passing schemes that might be implemented in the brain (Minka (2005); Parr, Markovic, Kiebel, and Friston (2019)).

We illustrate that the Rényi divergences, and their associated bounds, provide a complementary (but alternate) formulation to manipulation of priors for evaluating behavioural variations. Empirically, this poses an interesting question: are observed differences in choice behaviour a consequence of α values (i.e., optimisation objective difference) or specific priors—when the variational family is not in the same family of distributions as the true posterior? Formally, Rényi bound with $\alpha \to 0$ values provide a more graceful way of accounting for uncertainty or 'keeping options open' whilst making inferences about hidden states. We leave further links to human choice behaviour for future work.

7 Conclusion

We offer an account of behavioural variations using Rényi divergences and their associated variational bounds that complement usual formulations in terms of different prior beliefs. We show how different Rényi bounds induce behavioural differences for a fixed generative model that are formally distinct from a change of priors. This is accomplished by changes in an

α parameter that alters the bound's strength, inducing different inferences and consequent behavioural variations. Crucially, the inferences produced in this way do not seem to be accounted for by a change in priors under the standard variational objective. We emphasise that the Rényi bounds are analogous to the variational free energy (or evidence lower bound) and can be derived using the same assumptions. This formulation is illustrated through numerical analysis and demonstrates that $\alpha > 1$ values give rise to mode-seeking behaviours and $\alpha < 1$ values to mode-covering behaviours when priors are held constant.

Software note The code required to reproduce the simulations and figures is available here: https://github.com/ucbtns/renyibounds

Acknowledgements NS is funded by Medical Research Council (MR/S502522/1). FF is funded by the ERC Advanced Grant (no: 742870) and by the Swiss National Supercomputing Centre (CSCS, project: s1090). LD is supported by the Fonds National de la Recherche, Luxembourg (Project code: 13568875). This publication is based on work partially supported by the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1). KJF is funded by the Wellcome Trust (Ref: 203147/Z/16/Z and 205103/Z/16/Z).

Conflicts of Interest The authors declare no conflict of interest

References

Amari, S. (2009). α-divergence is unique, belonging to both f-divergence and bregman divergence classes. *IEEE Transactions on Information Theory*, *55*, 4925-4931.

Amari, S.-i. (2012). *Differential-geometrical methods in statistics* (Vol. 28) [Book]. Springer Science & Business Media.

Amari, S.-i., & Cichocki, A. (2010). Information geometry of divergence functions [Journal Article]. *Bulletin of the polish academy of sciences. Technical sciences*, *58*(1), 183-195.

Ambrogioni, L., Güçlü, U., Güçlütürk, Y., Hinne, M., Maris, E., & van Gerven, M. A. (2018). Wasserstein variational inference [Journal Article]. *arXiv preprint arXiv:1805.11284*.

Angela, J. Y., & Dayan, P. (2002). Acetylcholine in cortical inference [Journal Article]. *Neural Networks*, *15*(4-6), 719-730.

Angela, J. Y., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention [Journal Article]. *Neuron*, *46*(4), 681-692.

Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, *47*(2), 235–256.

Ay, N., & Gibilisco, P. (2016). *Information geometry and its applications* [Book]. Springer.

Barber, D., & van de Laar, P. (1999). Variational cumulant expansions for intractable distributions [Journal Article]. *Journal of Artificial Intelligence Research*, *10*, 435-455.

Beal, M. J. (2003). Variational algorithms for approximate bayesian inference [Journal Article]. *PhD. Thesis, University College London*.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians [Journal Article]. *Journal of the American Statistical Association*, *112*(518), 859-877.

Bogacz, R. (2017a). A tutorial on the free-energy framework for modelling perception and learning [Journal Article]. *Journal of Mathematical Psychology*, *76*, 198–211. Retrieved from `http://dx.doi.org/10.1016/j.jmp.2015.11.003` doi: doi:10.1016/j.jmp.2015.11.003

Bogacz, R. (2017b, February). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, *76*, 198–211. doi: doi:10.1016/j.jmp.2015.11.003

Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017, December). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, *81*, 55–79. doi: doi:10.1016/j.jmp.2017.09.004

Burbea, J. (1984). *Informative geometry of probability spaces* (Tech. Rep.). PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS.

Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). Active inference on discrete state-spaces: a synthesis [Journal Article]. *arXiv preprint arXiv:2001.07203*.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural computation*, *7*(5), 889–904.

Dieng, A. B., Tran, D., Ranganath, R., Paisley, J., & Blei, D. M. (2016). Variational inference via χ-upper bound minimization [Journal Article]. *arXiv preprint arXiv:1611.00328*.

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. MIT press.

FitzGerald, T. H., Schwartenbeck, P., Moutoussis, M., Dolan, R. J., & Friston, K. (2015). Active inference, evidence accumulation, and the urn task [Journal Article]. *Neural Comput*, *27*(2), 306-28. Retrieved from `http://www.ncbi.nlm.nih.gov/pubmed/25514108` doi: doi:10.1162/NECO_a_00699

Fountas, Z., Sajid, N., Mediano, P. A., & Friston, K. (2020). Deep active inference agents using monte-carlo methods [Journal Article]. *arXiv preprint arXiv:2006.04176*.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory [Journal Article]. *Neural Comput*, *29*(1), 1-49. Retrieved from `https://www.ncbi.nlm.nih.gov/pubmed/27870614` doi: doi:10.1162/NECO_a_00912

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: dopamine and decision-making [Journal Article].

Philos Trans R Soc Lond B Biol Sci, *369*(1655). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25267823 doi: doi:10.1098/rstb.2013.0481

Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value [Journal Article]. *Cognitive Neuroscience*, *6*(4), 187–224. Retrieved from http://dx.doi.org/10.1080/17588928.2015.1020053 doi: doi:10.1080/17588928.2015.1020053

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Han, D., Doya, K., & Tani, J. (2021). Goal-directed planning by reinforcement learning and active inference. *arXiv preprint arXiv:2106.09938*.

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain [Journal Article]. *Frontiers in Psychology*, *3*(APR), 1–14. doi: doi:10.3389/fpsyg.2012.00096

Huzurbazar, V. S. (1955). Exact forms of some invariants for distributions admitting sufficient statistics [Journal Article]. *Biometrika*, *42*(3/4), 533-537. Retrieved from http://www.jstor.org/stable/2333405 doi: doi:10.2307/2333405

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models [Journal Article]. *Machine learning*, *37*(2), 183-233.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Knill, D. C., & Pouget, A. (2004). The bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712–719.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 79–86.

Lattimore, T., & Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

Li, Y., & Turner, R. E. (2017). Rényi divergence variational inference [Conference Proceed-

ings]. In *Advances in neural information processing systems* (p. 1073-1081).

Metelli, A. M., Papini, M., Faccio, F., & Restelli, M. (2018). Policy optimization via importance sampling. *arXiv preprint arXiv:1809.06098*.

Millidge, B., Tschantz, A., & Buckley, C. L. (2020). Predictive coding approximates backprop along arbitrary computation graphs [Journal Article]. *arXiv preprint arXiv:2006.04182*.

Minka, T. (2005). *Divergence measures and message passing* (Report).

Murphy, K. P. (2007). Conjugate bayesian analysis of the gaussian distribution. *def*, *1*($2\sigma 2$), 16.

Nielsen, F., & Nock, R. (2011). On rényi and tsallis entropies and divergences for exponential families [Journal Article]. *arXiv preprint arXiv:1105.3259*.

Parisi, G. (1988). *Statistical field theory* [Book]. Basic Books. Retrieved from `https://books.google.co.uk/books?id=OF8sAAAAYAAJ`

Parr, T., & Friston, K. J. (2017). Uncertainty, epistemics and active inference [Journal Article]. *Journal of the Royal Society Interface*, *14*(136), 20170376. Retrieved from `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5721148/` doi: doi:10.1098/rsif.2017.0376

Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference [Journal Article]. *Biological Cybernetics*, *113*(5-6), 495-513. Retrieved from `https://dx.doi.org/10.1007/s00422-019-00805-w` doi: doi:10.1007/s00422-019-00805-w

Parr, T., Markovic, D., Kiebel, S. J., & Friston, K. J. (2019). Neuronal message passing using mean-field, bethe, and marginal approximations [Journal Article]. *Scientific Reports*, *9*(1), 1889. Retrieved from `https://doi.org/10.1038/s41598-018-38246-3` doi: doi:10.1038/s41598-018-38246-3

Parr, T., Sajid, N., & Friston, K. J. (2020). Modules or mean-fields? *Entropy*, *22*(5), 552.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chin-

tala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32* (pp. 8024–8035). Curran Associates, Inc. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Penny, W. (2012). Bayesian models of brain and behaviour [Journal Article]. *ISRN Biomathematics*, *2012*, 785791. Retrieved from https://doi.org/10.5402/2012/785791 doi: doi:10.5402/2012/785791

Perrykkad, K., & Hohwy, J. (2020). Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action. *New Ideas in Psychology*, *56*, 100750.

Phan, M., Abbasi-Yadkori, Y., & Domke, J. (2019). Thompson sampling with approximate inference. *arXiv preprint arXiv:1908.04970*.

Regli, J.-B., & Silva, R. (2018). Alpha-beta divergence for variational inference [Journal Article]. *arXiv preprint arXiv:1805.01045*.

Rényi, A. (1961). On measures of entropy and information [Conference Proceedings]. In *Proceedings of the fourth berkeley symposium on mathematical statistics and probability, volume 1: Contributions to the theory of statistics.* The Regents of the University of California.

Sajid, N., Ball, P. J., Parr, T., & Friston, K. J. (2021). Active inference: demystified and compared [Journal Article]. *Neural Computation*, 1-39.

Sajid, N., Convertino, L., & Friston, K. (2021). Cancer niches and their kikuchi free energy. *Entropy*, *23*(5), 609.

Sajid, N., Parr, T., Gajardo-Vidal, A., Price, C. J., & Friston, K. J. (2020). Paradoxical lesions, plasticity and active inference [Journal Article]. *Brain Communications*. Retrieved from https://doi.org/10.1093/braincomms/fcaa164 doi:

doi:10.1093/braincomms/fcaa164

Schmidhuber, J. (1990). *Making the world differentiable: On using fully recurrent self-supervised neural networks for dynamic reinforcement learning and planning in non-stationary environments* (Tech. Rep. Nos. FKI-126-90, https://people.idsia.ch/~juergen/FKI-126-90ocr.pdf). Institut für Informatik, Technische Universität München.

Schmidhuber, J. (1991). Curious model-building control systems [Journal Article]. *In Proc. International Joint Conference on Neural Networks, Singapore. IEEE*, *2*, 1458–1463.

Schmidhuber, J. (1992). Learning complex, extended sequences using the principle of history compression. *Neural Computation*, *4*(2), 234-242.

Schmidhuber, J. (2006). Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts [Journal Article]. *Connection Science*, *18*(2), 173-187. Retrieved from <GotoISI>://WOS:000239500800006 doi: doi:10.1080/09540090600768658

Schmidhuber, J., & Heil, S. (1995). Predictive coding with neural nets: Application to text compression. In *Advances in neural information processing systems* (pp. 1047–1054).

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., Wurst, F., Kronbichler, M., & Friston, K. (2015). Optimal inference with suboptimal models: addiction and active bayesian inference [Journal Article]. *Med Hypotheses*, *84*(2), 109-17. doi: doi:10.1016/j.mehy.2014.12.007

Schwöbel, S., Kiebel, S., & Marković, D. (2018). Active inference, belief propagation, and the bethe approximation [Journal Article]. *Neural computation*, 1-38.

Sharpe, W. F. (1994). The sharpe ratio. *Journal of portfolio management*, *21*(1), 49–58.

Smith, R., Lane, R. D., Parr, T., & Friston, K. J. (2019). Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their potential clinical relevance [Journal Article]. *Neuroscience & Biobehavioral Reviews*,

107, 473-491.

Soydaner, D. (2020). A comparison of optimization algorithms for deep learning. *International Journal of Pattern Recognition and Artificial Intelligence*, *34*(13), 2052013.

Spratling, M. W. (2017). A review of predictive coding algorithms [Journal Article]. *Brain and Cognition*, *112*, 92–97. Retrieved from http://dx.doi.org/10.1016/j.bandc.2015.11.003 doi: doi:10.1016/j.bandc.2015.11.003

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press.

Storck, J., Hochreiter, S., & Schmidhuber, J. (1995). Reinforcement driven information acquisition in non-deterministic environments. In *Proceedings of the international conference on artificial neural networks, paris* (Vol. 2, p. 159-164). EC2 & Cie.

Sun, Y., Gomez, F., & Schmidhuber, J. (2011). Planning to be surprised: Optimal bayesian exploration in dynamic environments [Book Section]. In J. Schmidhuber, K. R. Thórisson, & M. Looks (Eds.), *Artificial general intelligence: 4th international conference, agi 2011, mountain view, ca, usa, august 3-6, 2011. proceedings* (p. 41-51). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-642-22887-2_5 doi: doi:10.1007/978-3-642-22887-2_5

Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). Learning action-oriented models through active inference [Journal Article]. *PLOS Computational Biology*, *16*(4), e1007805. Retrieved from https://doi.org/10.1371/journal.pcbi.1007805 doi: doi:10.1371/journal.pcbi.1007805

van de Laar, T., Senoz, I., Özçelikkale, A., & Wymeersch, H. (2021). Chance-constrained active inference [Journal Article]. *arXiv preprint arXiv:2102.08792*.

Van Erven, T., & Harremos, P. (2014). Rényi divergence and kullback-leibler divergence [Journal Article]. *IEEE Transactions on Information Theory*, *60*(7), 3797-3820.

Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and

variational inference [Journal Article]. *Foundations and Trends® in Machine Learning*, *1*(1-2), 1-305.

Wan, N., Li, D., & Hovakimyan, N. (2020). f-divergence variational inference [Journal Article]. *Advances in Neural Information Processing Systems*, *33*.

Welbourne, S. R., Woollams, A. M., Crisp, J., & Lambon-Ralph, M. A. (2011). The role of plasticity-related functional reorganization in the explanation of central dyslexias [Journal Article]. *Cognitive Neuropsychology*.

Whittington, J. C. R., & Bogacz, R. (2017). An approximation of the error backpropagation algorithm in a predictive coding network with local hebbian synaptic plasticity [Journal Article]. *Neural Comput*, *29*(5), 1229-1262. doi: doi:10.1162/NECO_a_00949

Zhang, M., Bird, T., Habib, R., Xu, T., & Barber, D. (2019). Variational f-divergence minimization [Journal Article]. *arXiv preprint arXiv:1907.11891*.