

## Bayesian Brains and the Rényi Divergence

**Noor Sajid**

*noor.sajid.18@ucl.ac.uk*

*Wellcome Centre for Human Neuroimaging, University College London,  
London WC1N 3AR, U.K.*

**Francesco Faccio**

*francesco@idsia.ch*

*The Swiss AI Lab IDSIA, USI, SUPSI, 6962, Viganello, Lugano, Switzerland*

**Lancelot Da Costa**

*l.da-costa@imperial.ac.uk*

*Wellcome Centre for Human Neuroimaging, University College London,  
London WC1N 3AR, U.K., and Department of Mathematics, Imperial College  
London, London SW7 2AZ, U.K.*

**Thomas Parr**

*thomas.parr.12@ucl.ac.uk*

*Wellcome Centre for Human Neuroimaging, University College London,  
London WC1N 3AR, U.K.*

**Jürgen Schmidhuber**

*juergen@idsia.ch*

*The Swiss AI Lab IDSIA, USI, SUPSI, 6962, Viganello, Lugano, Switzerland*

**Karl Friston**

*k.friston@ucl.ac.uk*

*Wellcome Centre for Human Neuroimaging, University College London,  
London WC1N 3AR, U.K.*

**Under the Bayesian brain hypothesis, behavioral variations can be attributed to different priors over generative model parameters. This provides a formal explanation for why individuals exhibit inconsistent behavioral 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 behavioral variability using Rényi divergences and their associated variational bounds. Rényi bounds are analogous to the variational free**

---

Noor Sajid and Francesco Faccio contributed equally to this article.

energy (or evidence lower bound) and can be derived under the same assumptions. Importantly, these bounds provide a formal way to establish behavioral differences through an  $\alpha$  parameter, given fixed priors. This rests on changes in  $\alpha$  that alter the bound (on a continuous scale), inducing different posterior estimates and consequent variations in behavior. Thus, it looks as if individuals have different priors and have reached different conclusions. More specifically,  $\alpha \rightarrow 0^+$  optimization 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 behavior. Furthermore,  $\alpha \rightarrow +\infty$  optimization 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 multiarmed bandit task. We note that these  $\alpha$  parameterizations 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 behavioral 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, & Rao, 2007; Knill & Pouget, 2004). It takes the view that the brain embodies a model about causes of sensation that allows for predictions about observations (Dayan, Hinton, Neal, & Zemel, 1995; Hohwy, 2012; Schmidhuber, 1992; Schmidhuber & Heil, 1995) and future behavior (Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017; Schmidhuber, 1990). Practically, this involves the optimization of a free energy functional (or evidence lower bound; Bogacz, 2017a; Friston et al., 2017; Penny, 2012), using variational inference (Blei, Kucukelbir, & McAuliffe, 2017; Wainwright & Jordan, 2008), to make appropriate predictions. The free energy functional can be derived from the Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951), which measures the dissimilarity between true and approximate posterior densities. Under this formulation, behavioral variations can be attributed to altered priors over the (hyper-)parameters of a generative model, given the same (variational) free energy functional (Friston et al., 2014; Schwartenbeck et al., 2015). This has been used to simulate variations in choice behavior (FitzGerald, Schwartenbeck, Moutoussis, Dolan, & Friston, 2015; Friston et al., 2014, 2015; Storck, Hochreiter, & Schmidhuber, 1995) and behavioral

deficits (Sajid, Parr, Gajardo-Vidal, Price, & Friston, 2020; Smith, Lane, Parr, & Friston, 2019).

Conversely, distinct behavioral profiles could be attributed to differences in the variational objective, given the same priors. In this article, we consider this alternative account of phenotypic variations in choice behavior using Rényi divergences (Amari, 2012; Amari & Cichocki, 2010; Phan, Abbasi-Yadkori, & Domke, 2019; Rényi, 1961; Van Erven & Harremos, 2014). These are a general class of divergences, indexed by an  $\alpha$  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 & Jordan, 2008) (indeed, previous work has developed divergence-based lower bounds that give tighter bounds—Barber & van de Laar, 1999), yet these may be more difficult to optimize 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 predefined 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 (generalized) variational inference objective called the Rényi bound (Li & 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 behavioral variations (Minka, 2005). Thus, changing the functional form of the bound will make it look as if individuals have different priors that is, they have reached different conclusions from the same observations due to the distinct optimization objective.

It is important to determine whether this formulation introduces fundamentally new differences in behavior 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 parameterizations of the Rényi bound under fixed priors and the variational free energy under different hyperpriors. It turns out that there is no clear correspondence in most cases. This suggests that differences in behavior caused by changes in the divergence supplement standard accounts of behavioral differences under changes of priors.

The Rényi divergences depend on an  $\alpha$  parameter that controls the strength of the bound<sup>1</sup> and induces different posterior estimates. Consequently, the resulting system behavior may vary and point toward different

---

<sup>1</sup>Here, strength of bound refers the closeness with which the variational functional bounds the (negative) log evidence.

priors that could have altered the variational posterior form. For this, we assume that systems (or agents) sample their actions based on posterior beliefs, and those posterior beliefs depend on the form of the Rényi bound  $\alpha$  parameter. This furnishes a natural explanation for observed behavioral variation. To make the link to behavior, we assume actions are selected, based on variational estimates, that maximize the Sharpe ratio (Sharpe, 1994), a variance-adjusted return. Accordingly, evaluation of behavioral differences rests on 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; 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.<sup>2</sup> 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  $\alpha$  values lead to mass-seeking approximate<sup>3</sup> posteriors that is, greedy preferences for a particular outcome. This happens because the variational posterior is constrained to be zero whenever the true posterior is zero. Conversely,  $\alpha \rightarrow 0^+$  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 optimization objective. This contrasts with standard accounts of behavioral differences, where the precision of some fixed priors is used to explain divergent behavior profiles under the same variational objective. In what follows, we present, and validate, this generalized 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 article 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

<sup>2</sup>Note that heuristics like the Sharpe ratio are unnecessary in active inference (Da Costa et al., 2020; Friston et al., 2017), which automatically accommodates uncertainty of this sort; however, it is a useful heuristic because it foregrounds the role of posterior uncertainty in action selection.

<sup>3</sup>We use approximate and variational posterior interchangeably throughout.

approach through numerical simulations of the multiarmed bandit (Auer, Cesa-Bianchi, & Fischer, 2002; Lattimore & Szepesvári, 2020) paradigm with multimodal observation distribution. Our simulations demonstrate that variational Bayesian agents, optimizing a generalized variational bound (i.e., Rényi bound) can naturally account for variations in choice behavior. We conclude with a brief discussion of future directions and the implications of our work for understanding behavioral 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 optimization problem, allowing otherwise intractable probability densities to be approximated (Jordan, Ghahramani, Jaakkola, & Saul, 1999; Wainwright & 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 minimizes 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 characterize the brain as performing Bayesian inference including predictive coding (Milidge, Tschantz, & Buckley, 2020; Perrykkad & Hohwy, 2020; Schmidhuber & Heil, 1995; Spratling, 2017; Whittington & Bogacz, 2017), and active inference (Da Costa et al., 2020; Friston et al., 2017; Sajid, Ball, Parr, & Friston, 2021; Storck et al., 1995; Tschantz, Seth, & 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,

$$p(s, o) = p(o|s)p(s), \quad (2.1)$$

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:

$$p(s|o) = \frac{p(o, s)}{p(o)}. \quad (2.2)$$

This quantity contains the evidence,  $p(o)$ , that can be calculated by marginalizing 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.<sup>4</sup> 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 nonzero:

$$\log p(o) = \log p(o) + \int_S \log \frac{p(s|o)}{p(s|o)} ds \quad (2.3)$$

$$= \int_S q(s) \log p(o) ds + \int_S q(s) \log \frac{p(s|o)}{p(s|o)} ds = \int_S q(s) \log \frac{p(s, o)}{p(s|o)} ds \quad (2.4)$$

$$= \int_S q(s) \log \frac{q(s)}{q(s)} ds + \int_S q(s) \log p(s, o) ds + \int_S q(s) \log \frac{1}{p(s|o)} ds \quad (2.5)$$

$$= \underbrace{\int_S q(s) \log \frac{1}{q(s)} ds + \int_S q(s) \log p(s, o) ds}_{\text{ELBO}} + \underbrace{\int_S q(s) \log \frac{q(s)}{p(s|o)} ds}_{\text{KL Divergence}} \quad (2.6)$$

The first two summands of the last equality are the evidence lower bound (Welbourne, Woollams, Crisp, & 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 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:

$$-\log p(o) = - \left[ \int_S q(s) \log \frac{1}{q(s)} ds + \int_S q(s) \log p(s, o) ds + \int_S q(s) \log \frac{q(s)}{p(s|o)} ds \right] \quad (2.7)$$

$$= \int_S q(s) \log q(s) ds - \int_S q(s) \log p(s, o) ds - \int_S q(s) \log \frac{q(s)}{p(s|o)} ds \quad (2.8)$$

<sup>4</sup>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 optimization procedure (Wainwright & Jordan, 2008).

$$\leq \int_{\mathcal{S}} q(s) \log q(s) \, ds - \int_{\mathcal{S}} q(s) \log p(s, o) \, ds \tag{2.9}$$

$$= -\mathbb{E}_{q(s)}[\log p(s, o)] - \mathbb{H}[q(s)] \tag{2.10}$$

$$= \underbrace{D_{KL}[q(s)||p(s)]}_{\text{complexity}} - \underbrace{\mathbb{E}_{q(s)}[\log p(o|s)]}_{\text{accuracy}} \tag{2.11}$$

$$= -D_{KL}[q(s)||p(s, o)] = \mathcal{F}. \tag{2.12}$$

The second-to-last line is the commonly presented decomposition of the variational free energy summands: complexity and accuracy (Friston et al., 2017; Sajid et al., 2021). The accuracy term represents how well observed data can be predicted, while complexity is a regularization term. The variational free energy objective favors 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 this notation is usually reserved for arguments that are both normalized (Bishop, 2006). However, here the normalization factors over  $p(\cdot)$  become an additive constant in the KL-divergence, which has no effect on the gradients used in optimization or inference. Contrariwise, the normalizing constant of  $q(\cdot)$  needs to be the same across the variational family.

In this setting, illustrations of behavioral variations (i.e., differences in variational posterior estimations) can result from different priors over the (hyper-)parameters<sup>5</sup> of the generative model (Storck et al., 1995), such as change in precision over the likelihood function (Friston et al., 2014). We reserve description of hyperpriors 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 behavioral variations alternate to a change of priors. For this, we can replace the KL divergence by a general divergence objective, that is, a nonnegative function  $D[\cdot||\cdot]$  that satisfies  $D[q(s)||p(s|o)] = 0$  if and only if  $q(s) = p(s|o)$  for all  $s \in \mathcal{S}$ .<sup>6</sup> For our purposes, we focus on Rényi

---

<sup>5</sup>Note that introducing hyperpriors (or precision priors) is standard part of the Bayesian machinery (Gelman, Carlin, Stern, & 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.

<sup>6</sup>Technically, this equality holds up to a set of measure zero.

divergences, a general class of divergences that includes the KL-divergence. Explicitly, we can derive the KL-divergence from the Rényi divergence as  $\alpha \rightarrow 1$ , for example, using L'Hôpital's rule, or the minimum description length as  $\alpha \rightarrow \infty$  (see Table 1). This has the advantage of being computationally tractable and satisfies many additional properties (Amari, 2012; Rényi, 1961; Van Erven & Harremoës, 2014). Rényi divergences are defined as (Li & Turner, 2017; Rényi, 1961)

$$D_\alpha[q(s)||p(s|o)] := \frac{1}{\alpha - 1} \log \int_S q(s)^\alpha p(s|o)^{1-\alpha} ds, \quad (3.1)$$

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  $\alpha$ . These differences are possible only when the true posterior (e.g., some multimodal distribution) is not in the same family of distributions as the approximate posterior, such as a gaussian distribution. Note that other (non-Rényi) divergences in the literature are also parameterized by  $\alpha$ , which can lead to confusion: the I divergence, Amari's  $\alpha$ -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 article to be generalized 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 (see equation 2.3). This gives us the Rényi bound introduced in Li & Turner (2017):

$$p(o) = \frac{p(o, s)}{p(s|o)} \implies \quad (3.2)$$

$$p(o)^{1-\alpha} p(s|o)^{1-\alpha} = p(o, s)^{1-\alpha} \quad (3.3)$$

$$\int_S q(s)^\alpha p(o)^{1-\alpha} p(s|o)^{1-\alpha} ds = \int_S q(s)^\alpha p(o, s)^{1-\alpha} ds \quad (3.4)$$

$$\log \int_S q(s)^\alpha p(o)^{1-\alpha} p(s|o)^{1-\alpha} ds = \log \int_S q(s)^\alpha p(o, s)^{1-\alpha} ds \quad (3.5)$$

$$\log p(o)^{1-\alpha} + \log \int_S q(s)^\alpha p(s|o)^{1-\alpha} ds = \log \int_S q(s)^\alpha p(o, s)^{1-\alpha} ds \quad (3.6)$$

$$\log p(o)^{1-\alpha} = \log \int_S q(s)^\alpha p(o, s)^{1-\alpha} ds - \log \int_S q(s)^\alpha p(s|o)^{1-\alpha} ds \quad (3.7)$$



Table 1: Examples of (Normalized) Rényi Divergences (Li & Turner, 2017; Minka, 2005; Van Erven & Harremos, 2014) for Different Values of  $\alpha$ , and the Accompanying Rényi Bounds.

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

Notes: We omit  $\alpha \rightarrow 0$  because the limit is not a divergence. These divergences have a nondecreasing order:  $Hel^2(q, p) \leq D_{\frac{1}{2}}[q||p] \leq D_1[q||p] \leq D_2[q||p] \leq \chi^2(q, p)$  (Van Erven & Harremos, 2014).

$$\log p(o) = \underbrace{\frac{1}{1-\alpha} \log \int_S q(s)^\alpha p(o, s)^{1-\alpha} ds}_{\text{Rényi Bound}} + \underbrace{\frac{1}{\alpha-1} \log \int_S q(s)^\alpha p(s|o)^{1-\alpha} ds}_{\text{Rényi Divergence}} \quad (3.8)$$

$$\log p(o) = -D_\alpha[q(s)||p(o, s)] + D_\alpha[q(s)||p(s|o)]. \quad (3.9)$$

We assume that  $q(s)$  and  $p(s|o)$  are nonzero and  $\alpha \in \mathbb{R}^+ \setminus \{1\}$ . Additionally, we are licensed to make the move from equations 3.5 to 3.6 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 (see equation 2.7):

$$-\log p(o) = \frac{1}{\alpha-1} \log \int_S q(s)^\alpha p(o, s)^{1-\alpha} ds - \frac{1}{\alpha-1} \log \int_S q(s)^\alpha p(s|o)^{1-\alpha} ds \quad (3.10)$$

$$\leq \frac{1}{\alpha-1} \log \int_S q(s)^\alpha p(o, s)^{1-\alpha} ds = D_\alpha[q(s)||p(o, s)]. \quad (3.11)$$

Similar to the Rényi divergence, we expect variations in the estimation of the approximate posterior with  $\alpha$  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,  $\alpha \rightarrow 0^+$  optimization 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,  $\alpha \rightarrow +\infty$  optimization 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 behavior (Li & 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 behavioral differences through changes in the  $\alpha$  parameter. That is, we would expect a natural shift in behavioral preferences as we move from small values to large, positive  $\alpha$  values, given fixed priors. Section 5 demonstrates this shift in preferences in a multiarmed bandit setting.

#### 4 Variational Bounds, Precision, and Posteriors

It is important to determine whether this formulation of behavior 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

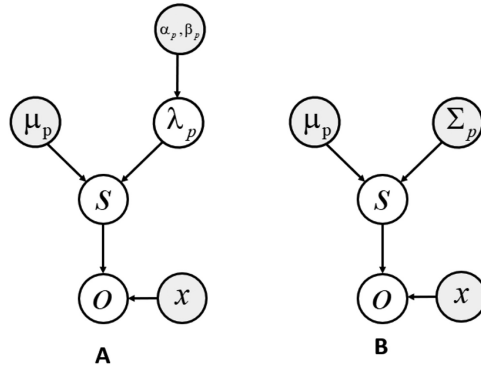


Figure 1: Graphical model for the gaussian-gamma (A) and gaussian (B) system. White circles represent random variables, gray circles represent priors, and  $x$  is the parameter governing the mean. The difference between these models is that in model A, the precision parameters over hidden states  $\lambda_p$  are random variables that follow a gamma distribution with parameters  $\alpha_p, \beta_p$ , while in model B, the precision is held fixed. Here, the scalar parameter  $\sigma_p$  has been deliberately omitted from the figure.

bound (see equation 3.3) with fixed prior beliefs but altered  $\alpha$  value and through the standard variational objective (see equation 2.3) with altered prior beliefs. If this were to be the case, we would be able to rewrite the variational free energy under different precision hyperpriors as the Rényi bound, where hyperparameters now play the role of the  $\alpha$  parameter. If this correspondence holds true, the two variational bounds (i.e., Rényi and variational free energy) would share similar optimization landscapes (i.e., inflection or extrema), with respect to the posterior under some different priors or  $\alpha$  value.

Variations in these hyperpriors speak to different priors, under which agents can exhibit conservative or greedy choice behavior. Practically, this may be a result of either lending one contribution more precision through weighting the log probability under the standard variational objective or altering the priors by taking the log of the probability to the power of  $\alpha$ . To illustrate this equivalence, we consider the following systems (see Figure 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 parameterization to derive the analytical form of the Rényi bound. The difference in parameterization is required to establish whether changes in prior beliefs (or precision) are equivalent to the  $\alpha$  parameter. In other words, this formulation allows us to ask whether

one can either alter the precision prior or the  $\alpha$  value to evince behavioral differences. If this were the case, we would expect equivalences between the two analytical bounds, given the different parameterizations.

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 and  $o \in O$  the observations (see Figure 1A).  $\lambda_k$  is the precision parameter,  $\Sigma_k$  is the covariance, and  $x$  the parameter governing the mean. The variational family is parameterized as a gaussian. This is formalized as

$$p(s, \lambda_p) = \mathcal{N}(s; 0, (\lambda_p \sigma_p)^{-1}) \text{Gam}(\lambda_p; \alpha_p, \beta_p), \quad (4.1)$$

$$p(o|s) = \mathcal{N}(so, \Sigma_l), \quad (4.2)$$

$$q(s) = \mathcal{N}(\mu_q, \Sigma_q), \quad (4.3)$$

where  $\Sigma_p = (\lambda_p \sigma_p)^{-1}$ ,  $s$  are scalars,  $o$  has dimension  $n$ , and  $x$  has dimensionality  $n \times 1$ . Here,  $\Sigma_l$  represents the covariance over the likelihood and  $\Sigma_k$  the covariance where  $k \in (p, l, q)$ . In equation 4.1,  $\mu_p = 0$  and has been written as such. Additionally, equation 4.1 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 (see appendix B for the derivation):

$$\begin{aligned} & -D_{KL}[q(s)||p(s, o)] \\ &= \frac{1}{2} \log \left( \frac{|\Sigma_q|}{(2\pi)^n |\Sigma_p| |\Sigma_l|} \right) \end{aligned} \quad (4.4)$$

$$- \frac{1}{2} \left( o^T \Sigma_l^{-1} o + \mu_q^2 \Sigma_p^{-1} + \mu_q^2 x^T \Sigma_l^{-1} x - 2\mu_q x^T \Sigma_l^{-1} o \right) \quad (4.5)$$

$$- \frac{1}{2} \left( \Sigma_q x^T \Sigma_l^{-1} x + \Sigma_q \Sigma_p^{-1} - 1 \right) \quad (4.6)$$

$$- \log \frac{\lambda_p^{\alpha_p - 1} \beta_p^{\alpha_p}}{\Gamma(\alpha_p)} - \lambda_p \beta_p. \quad (4.7)$$

For additional terms introduced via the gamma prior, see equation 4.7.

**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 parameterized as a gaussian distribution (see Figure 1B),

$$p(s) = \mathcal{N}(0, \Sigma_p), \tag{4.8}$$

$$p(o|s) = \mathcal{N}(sx, \Sigma_l), \tag{4.9}$$

$$q(s) = \mathcal{N}(\mu_q, \Sigma_q), \tag{4.10}$$

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 (see appendix B for the derivation):

$$\begin{aligned} & -D_\alpha[q(s)||p(s, o)] \\ &= \frac{1}{2} \log \left( \frac{|\Sigma_q|}{(2\pi)^n |\Sigma_p| |\Sigma_l|} \right) \end{aligned} \tag{4.11}$$

$$- \frac{\alpha}{2(\Sigma_q \Sigma_\alpha^{-1})} \left( o^T \Sigma_l^{-1} o + \mu_q^2 \Sigma_p^{-1} + \mu_q^2 x^T \Sigma_l^{-1} x - 2\mu_q x^T \Sigma_l^{-1} o \right) \tag{4.12}$$

$$- \frac{1}{2(1-\alpha)} \log \left( 1 + (1-\alpha) \left( \Sigma_q x^T \Sigma_l^{-1} x + \Sigma_q \Sigma_p^{-1} - 1 \right) \right) \tag{4.13}$$

$$- \frac{1}{2\Sigma_\alpha^{-1}} \left( (1-\alpha) \Sigma_p^{-1} o^T \Sigma_l^{-1} o \right), \tag{4.14}$$

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

**4.3 Correspondence between Variational Free Energy and 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  $\alpha \rightarrow 1$ . Here, we expect to find an exact correspondence between the variational free energy and the Rényi bound as the Rényi divergence tends toward the KL-divergence as  $\alpha \rightarrow 1$ . Our derivations confirm this, upon comparison of the equivalent terms for each objective. The first terms in each objective, equations 4.4 and 4.11, are the same. Interestingly, the second term in the Rényi bound, equation 4.12, is a scalar multiple of the second term in variational free energy (see equation 4.5), where the scalar quantity  $\frac{\alpha}{\Sigma_q \Sigma_\alpha^{-1}}$  tends to 1 for  $\alpha \rightarrow 1$ . The third term in equation 4.13, for  $\alpha \rightarrow 1$ , is a limit of the form  $\lim_{x \rightarrow 0} \frac{1}{x} \log(1 + xw) = w$ ,

resulting exactly in equation 4.6. Finally, the last term in the Rényi bound tends to zero as  $\alpha \rightarrow 1$ , equation 4.14.

Next, we evaluate the correspondence between the variational free energy and Rényi bound when  $\alpha \in \mathbb{R}^+ \setminus \{1\}$ . Now, the  $\alpha$  values scale the terms in the Rényi bound with equation 4.14 having an influence on the final bound estimate. For comparability, we introduced the gamma prior to a simple gaussian system. As shown in equation 4.7, this introduces additional terms that scale the free energy  $\mathcal{F}$ . We expect the scaling from the  $\alpha$  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 (see Figure 2). The numerical simulation illustrates that optimization of these objectives, for appropriate priors  $(\alpha_p, \beta_p)$  or the  $\alpha$  value, can lead to (extremely) different variational densities.

Interestingly, the two variational objectives exhibit a similar optimization landscape under specific parameterizations. For example, a striking (local) minimum of  $-33.14$  nats is observed when  $\alpha_p$  is approximately 1,  $\beta_p$  is greater than 0.8, and  $\alpha < 5$ . However, this is constrained to a small space of posterior  $\mu_q$  estimates. Outside these posterior parameters, the optimization landscape differs. Importantly, this difference becomes more acute when considering  $\sigma_q$ . Here,  $\sigma_q$  represents 1-dimensional  $\Sigma_q$ . This suggests hyperpriors may be particularly important in shaping the correspondence between the two variational objectives. However, the optimization profile can differ under inappropriate priors (i.e., a misalignment between prior beliefs and  $\alpha$  value) and lead to divergences in the estimated variational density (see Figure 2).

Briefly, we do not observe a direct correspondence in the optimization landscapes (and the variational posterior) for certain priors or  $\alpha$  value. These numerical analyses demonstrate that the Rényi divergences account for behavioral differences in a way that is formally distinct from a change in priors, through manipulation of the  $\alpha$  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 Multiarmed Bandit Simulation

---

In this section, we illustrate the differential preferences that arise naturally under the Rényi bound. For this, we simulated the multiarmed bandit (MAB) paradigm (Auer et al., 2002; Lattimore & Szepesvári, 2020) using three arms. The MAB environment was formulated as a one-state Markov decision process (MDP) that is, the environment remains in the same state independent 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

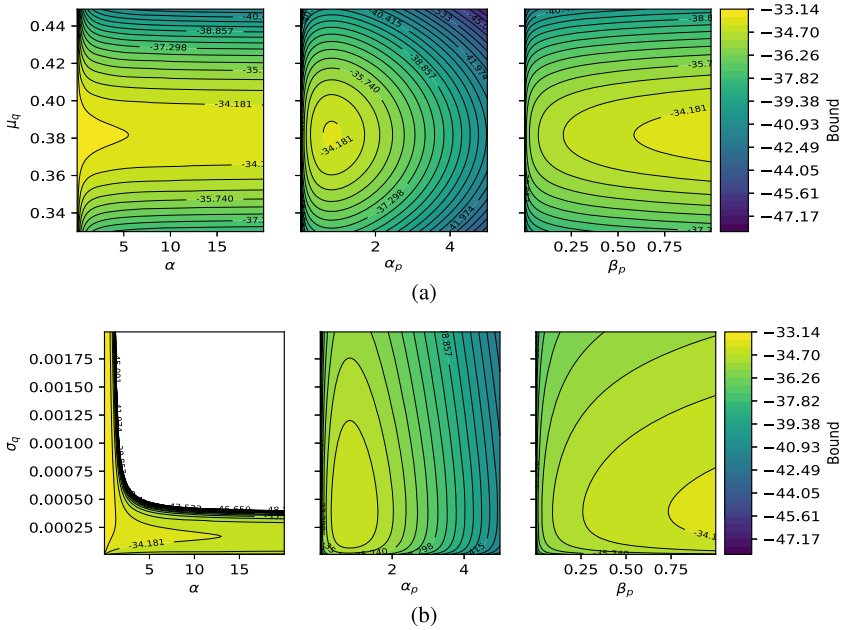


Figure 2: Heat map of variational bounds as a function of estimated sufficient statistics:  $\mu_q$  (a) and  $\sigma_q$  (b). Here,  $\sigma_q$  represents 1-dimensional  $\Sigma_q$ . These graphics plot the optimization landscape for changing priors or  $\alpha$  values. The first column plots the Rényi bound, as a function of  $\alpha$  on the x-axis and  $\mu_q$  (a) or  $\sigma_q$  (b) on the y-axis. Similarly, the next two columns plot the free energy, as a function of  $\alpha_p$  (center column) or  $\beta_p$  (right column) on the x-axis and  $\mu_q$  (a) or  $\sigma_q$  (b) on the y-axis. The variational bound ranges from  $-33$  (yellow) to  $-47$  nats (blue). The empty region in panel b for different  $\alpha$  values in the Renyi bound is a consequence of the (positive-definiteness) constraint imposed on  $\Sigma_q$  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, hyperparameters are fixed with  $\mu_q = 0.4$ ,  $\sigma_q = 1e - 4$ ,  $\alpha_p = 0.8$ ,  $\beta_p = 0.8$ ,  $\lambda_p = 0.8$ ,  $x = \{r : r = 1.1 \times n, n \in \{0, 1, \dots, 19\}\}$ ,  $y = 0.4 \times x$ ,  $\Sigma_l = \mathbb{I}_{20}$ .

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, given the expected return of an arm  $R = \mathbb{E}[R_t]$ , the Sharpe ratio is defined as  $SR := \frac{\mathbb{E}[R_t]}{\sqrt{\mathbb{V}[R_t]}}$ , 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  $\alpha$  influence the sufficient

statistics of the variational posterior and ensuing behavior. Practically, this means we sample from the posterior distribution for each state (i.e., arm) and select actions that maximize the Sharpe ratio. The Sharpe ratio affords an action selection criterion that accommodates posterior uncertainty about hidden states, which underwrites choice behavior. For example, posterior estimates for some (suboptimal) 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 suboptimal arms with uncertain payoff may be selected with unduly high probability. The Sharpe ratio precludes this, penalizing arms with high posterior uncertainty.

We modeled each arm with a fixed multimodal distribution (a mixture of gaussians) unknown to the agent, characterizing this as stationary stochastic bandit setting. Explicitly, this entailed the following parameterization for each arm:

$$p(s) = \sum_i^2 \omega_i \mathcal{N}(\mu_i, \Sigma_i), \quad (5.1)$$

$$p(o|s) = \mathcal{N}(s, 1.0), \quad (5.2)$$

$$q(s) = \mathcal{N}(\mu_q, \Sigma_q), \quad (5.3)$$

$$\sum_i^2 \omega_i = 1, \omega_i > 0, \quad (5.4)$$

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.<sup>7</sup> However, due to the multimodal prior, 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 Figure 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  $t$ .

Optimizing the Rényi bound under different  $\alpha$  values led to varying posterior estimates and accompanying behavioral differences manifested by distinct arm choices. To show this, we simulated six agents optimizing the Rényi bound for distinct  $\alpha$  values:  $\rightarrow +\infty, 10, 2, \rightarrow 1^-, 0.5, \rightarrow 0^+$  – across

<sup>7</sup>That is, a fully factorized variational distribution. For further details see Minka (2005), Parr, Sajid, and Friston (2020), and Sajid, Convertino, and Friston (2021).



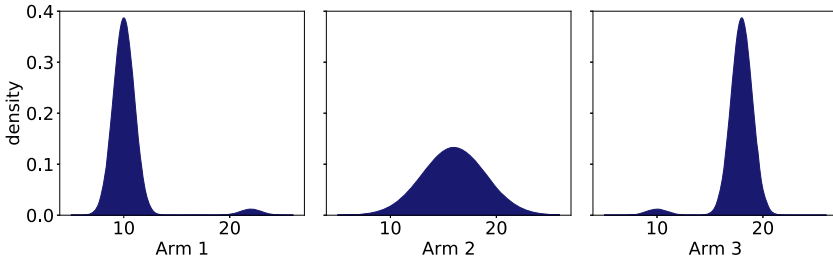


Figure 3: Score distribution for each arm. The panels plot the score distributions for each arm. The  $x$ -axis is the  $s \sim q(s)$  and  $y$ -axis the score density. Arm 1 has a multimodal distribution of  $\mu_1^1 = 10$  ( $\Sigma_1^1 = 1$ ) and  $\mu_2^1 = 22$  ( $\Sigma_2^1 = 1$ ) with  $\omega_1^1 = 0.97$  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 multimodal 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.

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,

$$i^* = \operatorname{argmax}_i \frac{s_i}{\Sigma_q^i}, \tag{5.5}$$

where  $\Sigma_q^i$  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  $\mu_q$  optimized under each bound. This avoided premature convergence to suboptimal 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 multimodal 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 & 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 optimization strategies could be used here (e.g., Momentum or RMSProp; Soydaner, 2020). Additionally, for each arm, there was a separate memory buffer and optimization process. The agent learned the score distribution

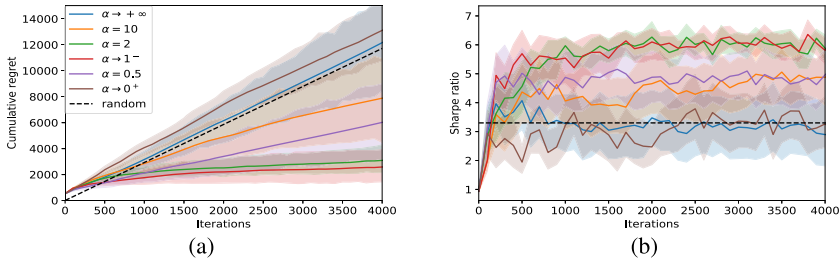


Figure 4: Regret (a) and Sharpe ratio (b) under the Rényi bound. (a) The line plot illustrates the cumulative regret across the 4000 iterations for each agent optimizing a particular Rényi bound. The  $x$ -axis denotes the iteration and  $y$ -axis the accompanying cumulative regret. (b) The line plot 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 optimizing Rényi bound for  $\alpha \rightarrow +\infty$ , orange for  $\alpha = 10$ , green for  $\alpha = 2$ , red for  $\alpha \rightarrow +1^-$ , purple for  $\alpha = 0.5$ , and brown for  $\alpha \rightarrow 0^+$ . 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  $\alpha \rightarrow +1^-$  and  $\alpha = 2$  obtained the best performance.

through the memory buffer that stored the previous 1000 observations. At each iteration, the observations in memory were used to optimize 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  $\alpha$  parameter. To assess the performance of each  $\alpha$ , we plot the accumulated regret and the accompanying Sharpe ratio in Figure 4. We observe that optimizing  $\alpha \rightarrow +1^-$ ; 2 leads to the lowest cumulative regret and a high Sharpe ratio. Conversely, optimizing  $\alpha \rightarrow 0^+$ ;  $\rightarrow +\infty$  leads to the highest cumulative regret and lowest Sharpe ratio.

To investigate this further, we plot the variational bounds for arm 1 under different  $\alpha$  parameters (see Figure 5). Recall from Figure 3 that if the variational posterior fits the right-hand-side mode, this results in suboptimal 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  $\alpha \rightarrow +\infty$ ;  $\rightarrow 0^+$  from the property of their variational bound. For agents optimizing  $\alpha \rightarrow +\infty$ , the approximate posterior fit the right-hand-side mode of the distribution due to its lower variance (i.e., mode-seeking behavior). Conversely, agents with  $\alpha \rightarrow 0^+$  would exhibit

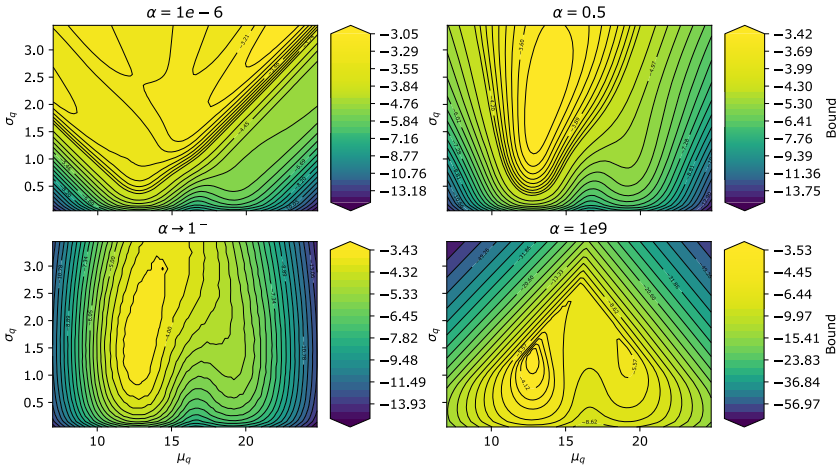


Figure 5: The Rényi bound as a function of the variational posterior. Here,  $\sigma_q$  represents 1-dimensional  $\Sigma_q$ . The contour plots show the optimization landscape for each  $\alpha$ . For  $\alpha = 1e9$ , we observe two optima; for small  $\alpha$  ( $1e-6$ ), the optimal solution exhibits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

mass-covering, high-variance posterior estimates. In contrast, agents optimizing  $\alpha \rightarrow 1^-$ ; 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 (see Figure 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  $\alpha$  parameterization, then there is a natural space and explanation for behavioral variations. In short, the shape of the posterior that underwrites ensuing behavior depends sensitively on the functional form of the variational bound.

## 6 Discussion

This article accounts for behavioral variations among agents using Rényi divergences and their associated variational bounds. These divergences are Rényi relative entropies<sup>8</sup> and satisfy similar properties as the KL divergence (Rényi, 1961; Van Erven & Harremos, 2014). Rényi divergences depend on an  $\alpha$  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 behavior. This provides a

<sup>8</sup>The Rényi entropy provides a parametric family of measures of information (Rényi, 1961).

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  $\alpha$  parameter. Yet note that a similar account is possible if actions depended on an expected free energy functional (Friston et al., 2017; Han, Doya, & Tani, 2021; Parr & Friston, 2019; van de Laar, Senoz, Özçelikkale, & Wymeersch, 2021), intrinsic reward (Schmidhuber, 1991, 2006; Storck et al., 1995; Sun, Gomez, & 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 behavioral variations) that vary smoothly with  $\alpha$ . As illustrated, in the bimodal scenario under our Rényi divergence definition, large, positive  $\alpha$  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 (the parts with the small total mass) to be excluded. Thus, the estimated variational posterior might be underestimated. Conversely, with small  $\alpha$  values, the approximation tries to cover the entire distribution, eventually forming an upper bound when  $\alpha \rightarrow 1$  (see Table 1). This happens because  $\alpha \rightarrow 1$  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  $\alpha$  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 parameterization shapes the preferences of the model. We observe this in our simple multiarmed bandit setting where large  $\alpha$  values seek to fit the posterior modes that lead to greater consistency in preferences over which arm to select. Conversely, small  $\alpha$  values try to cover the posterior distribution that led to greater flexibility over the choice of arm.

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

**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 optimization 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 behavior. On the basis of the above, we appeal to Rényi bounds, in addition to altered priors, to model behavioral variations. By committing to the Rényi bound, we provide an alternative perspective on how variant (or suboptimal) behavior can be modeled. This leads to a conceptual reversal of the standard variational free energy schemes, including predictive processing (Bogacz, 2017b; Buckley, Kim, McGregor, & Seth, 2017). That is, we can illustrate behavioral 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 suboptimal inferences using altered generative models or alternative variational bounds? This turns out to be significant in light of our numerical analysis (see 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 for variational objectives for lower levels to be modulated by higher levels through priors over  $\alpha$  values, a procedure of meta-inference. This is analogous to including precision priors over model parameters that have been associated with different neuromodulatory systems, such as state transition precision with noradrenergic and sensory precision with cholinergic systems (Fountas, Sajid, Mediano, & Friston, 2020; Parr & Friston, 2017). Consequently, this temporal separation of  $\alpha$  parameterizations may provide an interesting research avenue for understanding the role of neuromodulatory systems and how they facilitate particular behaviors (Angela & Dayan, 2002, 2005).

**6.2 Generalized Variational Inference.** The Rényi bound provides a generalized 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 behavioral preferences, directly via the optimization objective, without changing prior beliefs. Other variational objectives can be derived from other general families of divergences such as f-divergences and Wasserstein distances (Ambrogioni et al., 2018; Dieng, Tran, Ranganath, Paisley, & Blei, 2016; Regli & Silva, 2018), which can improve the statistical properties of the variational bounds for particular applications (Wan, Li, & Hovakimyan, 2020; Zhang, Bird, Habib, Xu, & Barber, 2019). Future work could generalize the arguments presented here and examine how these different divergences shape behavior 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 optimization 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  $\beta$  or  $\alpha$  value.

Implementations of the Rényi bound are constrained by sampling biases and interesting differences in optimization landscape. Indeed, when  $\alpha$  is extremely large, even if the approximate posterior distribution belongs to the same family as the true posterior, the optimization 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 optimizing 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 optimization 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 parameterizations of the variational posterior might negate the impact of the  $\alpha$  values. Specifically, we noted that if the true posterior is in the same family of distributions as the variational posterior, then changing the  $\alpha$  value does not have an impact on the shape of the variational posterior and, consequently, the system's behavior. However, complex parameterizations are computationally expensive and can still be inappropriate. Therefore, this departure from vanilla variational inference provides a useful explanation for different behaviors 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 affect 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, & Friston, 2019).

We illustrate that the Rényi divergences, and their associated bounds, provide a complementary (but alternate) formulation to manipulation of priors for evaluating behavioral variations. Empirically, this poses an interesting question: Are observed differences in choice behavior a consequence of  $\alpha$  values (i.e., optimization 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  $\alpha \rightarrow 0$  values provide a more graceful way of accounting for uncertainty or keeping options open while

making inferences about hidden states. We leave further links to human choice behavior for future work.

## 7 Conclusion

---

We offer an account of behavioral 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 behavioral differences for a fixed generative model that are formally distinct from a change of priors. This is accomplished by changes in an  $\alpha$  parameter that alters the bound's strength, inducing different inferences and consequent behavioral 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 emphasize 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 behaviors and  $\alpha < 1$  values to mode-covering behaviors when priors are held constant.

## Software Note

The code required to reproduce the simulations and figures is available at <https://github.com/ucbtms/renyibounds>.

## Acknowledgments

---

N.S. is funded by Medical Research Council (MR/S502522/1). F.F. is funded by the ERC Advanced Grant (742870) and the Swiss National Supercomputing Centre (CSCS, project s1090). L.D. 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). K.J.F. is funded by the Wellcome Trust (203147/Z/16/Z and 205103/Z/16/Z).

## Conflicts of Interest

The authors declare no conflict of interest.

## References

---

Amari, S.-i. (2012). *Differential-geometrical methods in statistic*. Berlin: Springer Science & Business Media.



- Amari, S.-i., & Cichocki, A. (2010). Information geometry of divergence functions. *Bulletin of the Polish Academy of Sciences. Technical Sciences*, 58(1), 183–195. 10.2478/v10175-010-0019-1
- Ambrogioni, L., Güçlü, U., Güçlütürk, Y., Hinne, M., Maris, E., & van Gerven, M. A. (2018). *Wasserstein variational inference*. arXiv:1805.11284.
- Angela, J. Y., & Dayan, P. (2002). Acetylcholine in cortical inference. *Neural Networks*, 15(4–6), 719–730. 10.1016/S0893-6080(02)00058-8, PubMed: 12371522
- Angela, J. Y., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4), 681–692. 10.1016/j.neuron.2005.04.026, PubMed: 15944135
- Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2), 235–256. 10.1023/A:1013689704352
- Barber, D., & van de Laar, P. (1999). Variational cumulant expansions for intractable distributions. *Journal of Artificial Intelligence Research*, 10, 435–455. 10.1613/jair.567
- Beal, M. J. (2003). *Variational algorithms for approximate Bayesian inference*. PhD diss., University College London.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Berlin: Springer.
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877. 10.1080/01621459.2017.1285773
- Bogacz, R. (2017a). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, 76, 198–211. 10.1016/j.jmp.2015.11.003
- Bogacz, R. (2017b). A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology*, 76, 198–211. 10.1016/j.jmp.2015.11.003.
- Buckley, C. L., Kim, C. S., McGregor, S., & Seth, A. K. (2017). The free energy principle for action and perception: A mathematical review. *Journal of Mathematical Psychology*, 81, 55–79. 10.1016/j.jmp.2017.09.004
- Burbea, J. (1984). *Informative geometry of probability spaces* (Tech. Rep.). Pittsburgh: University of Pittsburgh, Pennsylvania 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*. arXiv:2001.07203.
- Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation*, 7(5), 889–904. 10.1162/neco.1995.7.5.889, PubMed: 7584891
- Dieng, A. B., Tran, D., Ranganath, R., Paisley, J., & Blei, D. M. (2016). *Variational inference via  $\chi$ -upper bound minimization*. arXiv:1611.00328.
- Doya, K., Ishii, S., Pouget, A., & Rao, R. P. (2007). *Bayesian brain: Probabilistic approaches to neural coding*. Cambridge, MA: MIT Press.
- FitzGerald, T. H., Schwartenbeck, P., Moutoussis, M., Dolan, R. J., & Friston, K. (2015). Active inference, evidence accumulation, and the urn task. *Neural Computation*, 27(2), 306–328. 10.1162/NECO\_a\_00699
- Fountas, Z., Sajid, N., Mediano, P. A., & Friston, K. (2020). *Deep active inference agents using Monte-Carlo methods*. arXiv:2006.04176.
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, 29(1), 1–49. 10.1162/NECO\_a\_00912
- Friston, K. J., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 6(4), 187–224. 10.1080/17588928.2015.1020053, PubMed: 25689102



- Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). The anatomy of choice: Dopamine and decision-making. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 369(1655). 10.1098/rstb.2013.0481, PubMed: 25267823
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman and Hall/CRC.
- Han, D., Doya, K., & Tani, J. (2021). *Goal-directed planning by reinforcement learning and active inference*. arXiv:2106.09938.
- Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, 3(2012), 1–14. 10.3389/fpsyg.2012.00096, PubMed: 22279440
- Huzurbazar, V. S. (1955). Exact forms of some invariants for distributions admitting sufficient statistics. *Biometrika*, 42(3/4), 533–537. 10.2307/2333405
- Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37(2), 183–233. 10.1023/A:1007665907178
- 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. 10.1016/j.tins.2004.10.007, PubMed: 15541511
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79–86. 10.1214/aoms/1177729694
- Lattimore, T., & Szepesvári, C. (2020). *Bandit algorithms*. Cambridge: Cambridge University Press.
- Li, Y., & Turner, R. E. (2017). Rényi divergence variational inference. In I. Guyon, Y. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems*, 30 (pp. 1073–1081). Red Hook, NY: Curran.
- Metelli, A. M., Papini, M., Faccio, F., & Restelli, M. (2018). *Policy optimization via importance sampling*. arXiv:1809.06098.
- Millidge, B., Tschantz, A., & Buckley, C. L. (2020). *Predictive coding approximates backprop along arbitrary computation graphs*. arXiv:2006.04182.
- Minka, T. (2005). *Divergence measures and message passing*. Microsoft Research report UK MSR-TR-2005-173.
- Murphy, K. P. (2007). *Conjugate Bayesian analysis of the gaussian distribution*. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.126.4603>
- Parisi, G. (1988). *Statistical field theory*. New York: Basic Books.
- Parr, T., & Friston, K. J. (2017). Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface*, 14(136), 20170376. 10.1098/rsif.2017.0376
- Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics*, 113(5–6), 495–513. 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. *Scientific Reports*, 9(1), 1889. 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., . . . Chintala, 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). Red Hook, NY: Curran.

- Penny, W. (2012). Bayesian models of brain and behavior. *ISRN Biomathematics*, 2012, 785791. 10.5402/2012/78579110.5402/2012/785791
- Perrykkad, K., & Hohwy, J. (2020). Fidgeting as self-evidencing: A predictive processing account of non-goal-directed action. *New Ideas in Psychology*, 56, 100750. 10.1016/j.newideapsych.2019.100750
- Phan, M., Abbasi-Yadkori, Y., & Domke, J. (2019). *Thompson sampling with approximate inference*. arXiv:1908.04970.
- Regli, J.-B., & Silva, R. (2018). *Alpha-beta divergence for variational inference*. arXiv:1805.01045.
- Rényi, A. (1961). On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1: Contributions to the theory of statistics. Regents of the University of California.
- Sajid, N., Convertino, L., & Friston, K. (2021). Cancer niches and their kikuchi free energy. *Entropy*, 23(5), 609. 10.3390/e23050609, PubMed: 34069097
- Sajid, N., Ball, P. J., Parr, T., & Friston, K. J. (2021). Active inference: demystified and compared. *Neural Computation*, 33, 674–712. 10.1162/neco\_a\_01357, PubMed: 33400903
- Sajid, N., Parr, T., Gajardo-Vidal, A., Price, C. J., & Friston, K. J. (2020). Paradoxical lesions, plasticity and active inference. *Brain Communications*, 2, fcaa164. 10.1093/braincomms/fcaa164, PubMed: 33376985
- 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. FKI-126-90). Institut für Informatik, Technische Universität München. <https://people.idsia.ch/~juergen/FKI-126-90ocr.pdf>
- Schmidhuber, J. (1991). Curious model-building control systems. In *Proc. International Joint Conference on Neural Networks* (vol. 2, pp. 1458–1463). Piscataway, NJ: IEEE.
- Schmidhuber, J. (1992). Learning complex, extended sequences using the principle of history compression. *Neural Computation*, 4(2), 234–242. 10.1162/neco.1992.4.2.234
- Schmidhuber, J. (2006). Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science*, 18(2), 173–187. 10.1080/09540090600768658
- Schmidhuber, J., & Heil, S. (1995). Predictive coding with neural nets: Application to text compression. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems* (pp. 1047–1054). Cambridge, MA: MIT Press.
- 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. *Med. Hypotheses*, 84(2), 109–117. 10.1016/j.mehy.2014.12.007, PubMed: 25561321
- Schwöbel, S., Kiebel, S., & Marković, D. (2018). Active inference, belief propagation, and the Bethe approximation. *Neural Computation*, 30(2) 1–38.
- Sharpe, W. F. (1994). The Sharpe ratio. *Journal of Portfolio Management*, 21(1), 49–58. 10.3905/jpm.1994.409501
- 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. *Neuroscience and Biobehavioral Reviews*, 107, 473–491. 10.1016/j.neubiorev.2019.09.002, PubMed: 31518636
- Soydancer, D. (2020). A comparison of optimization algorithms for deep learning. *International Journal of Pattern Recognition and Artificial Intelligence*, 34(13), 2052013. 10.1142/S0218001420520138, PubMed: 34267412
- Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, 112, 92–97. 10.1016/j.bandc.2015.11.003, PubMed: 26809759
- Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: 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* (vol. 2, pp. 159–164).
- Sun, Y., Gomez, F., & Schmidhuber, J. (2011). Planning to be surprised: Optimal Bayesian exploration in dynamic environments. In J. Schmidhuber, K. R. Thórisson, & M. Looks (Eds.), *Proceedings of the 4th International Conference on Artificial General Intelligence* (pp. 41–51). Berlin: Springer. 10.1007/978-3-642-22887-2\_5
- Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). Learning action-oriented models through active inference. *PLOS Computational Biology*, 16(4), e1007805. 10.1371/journal.pcbi.1007805, PubMed: 32324758
- van de Laar, T., Senoz, I., Özçelikkale, A., & Wymeersch, H. (2021). *Chance-constrained active inference*. arXiv:2102.08792.
- Van Erven, T., & Harremoës, P. (2014). Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7), 3797–3820. 10.1109/TIT.2014.2320500
- Wainwright, M. J., & Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2), 1–305. 10.1561/2200000001
- Wan, N., Li, D., & Hovakimyan, N. (2020). f-divergence variational inference. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems*, 33. Red Hook, NY: Curran.
- 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. *Cognitive Neuropsychology*, 28, 65–108. 10.1080/02643294.2011.621937, PubMed: 22122115
- Whittington, J. C. R., & Bogacz, R. (2017). An approximation of the error backpropagation algorithm in a predictive coding network with local Hebbian synaptic plasticity. *Neural Comput.*, 29(5), 1229–1262. 10.1162/NECO\_a\_00949
- Zhang, M., Bird, T., Habib, R., Xu, T., & Barber, D. (2019). *Variational f-divergence minimization*. arXiv:1907.11891.

---

Received July 12, 2021; accepted November 29, 2021.