Bayesian brains and the Rényi divergence

Noor Sajid[1,] *, Lancelot Da Costa[1,2], Francesco Faccio[3], Thomas Parr[1] and Karl Friston[1]

noor.sajid.18@ucl.ac.uk,l.da-costa@imperial.ac.uk,francesco@idsia.ch,{thomas.parr.12; k.friston}@ucl.ac.uk

[1] Wellcome Centre for Human Neuroimaging, University College London, London, UK.
[2] Department of Mathematics, Imperial College London, London, UK.
[3]Swiss AI Lab IDSIA, Lugano, Switzerland.
*****Corresponding author

Abstract Under the Bayesian brain hypothesis, behavioural variations can be attributed to altered priors over the generative model (hyper-)parameters. This provides a particular explanation as to why individuals may exhibit inconsistent behavioural preferences when faced with similar observations. Here, we offer an alternative account for explaining behavioural variations using Rényi divergences, and their associated Rényi variational bounds. The Rényi bound provides a formal way to establish behavioural differences through its α parameter, ceteris paribus. This is accomplished by α changes that alter the bound, induce different posterior estimates, and consequent variations in behaviour. Thus, it looks as if individuals have different priors, and have reached different conclusions. We highlight that these Rényi bounds are analogous to the variational free energy or evidence lower bound and can be derived using the same assumptions. However, this α parameterisations is only relevant, i.e., shape preferences, when the true posterior is not in the same family of distributions as the assumed (simpler) approximate density – common for complex real-world scenarios. Consequently, this departure from vanilla variational inference provides a useful explanation for differences in behavioural preferences of biological (or artificial) agents – under the assumption that the brain performs variational Bayesian inference. We exemplify this formulation through numerical analysis of gaussian systems, and simulations of a multi-armed bandit task.

Keywords: Bayesian brain hypothesis, variational inference, Rényi divergences, Kullback-Leibler divergence, Rényi bound, variational free energy, evidence lower bound

---

Contents

1.0 Introduction

The notion that the brain is Bayesian and performs some form of inference has attracted enormous attention in neuroscience (Doya et al 2007, Knill & Pouget 2004). It takes the view that the brain embodies a model about causes of sensation, that allow for predictions about observations (Dayan et al 1995, Hohwy 2012) and future behaviour (Friston 2019, Friston et al 2017). Practically, this involves the optimisation of a free energy functional (or evidence lower bound) (Bogacz 2017, Friston et al 2012, Penny 2012), using variational inference (Blei et al 2017, Wainwright & Jordan 2008), to make appropriate predictions. The free energy functional can be derived from the Kullback-Leibler (KL)-divergence, which measures the dissimilarity between true and approximate posterior densities. Under this formulation, behavioural variations can be attributed to altered priors over the generative model (hyper-) parameters, given the same (variational) free energy functional (Friston et al 2014b, Schwartenbeck et al 2015). This has been used to simulate variations in choice behaviour (FitzGerald et al 2015, Friston et al 2014a, Friston et al 2015) and behavioural deficits (Parr & Friston 2017, Sajid et al 2020a, Smith et al 2019).

Conversely, distinct behavioural profiles could be attributed to differences in the variational objective, given the same set of priors. In this paper, we present this alternative account for explaining phenotypic variations using Rényi divergences (Amari 2012, Amari & Cichocki 2010, Amari 2009, Rényi 1961, Van Erven & 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 & Jordan 2008)[1]. Broadly, variational inference is the process of approximating a posterior probability through application of variational calculus. 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. Now, these Rényi divergences, instead of the KL-divergence, can be used to derive a (generalised) variational inference objective called the Rényi-bound (Li & Turner 2016). 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 α changes, on a continuous scale, that gives rise to different posterior estimates, and consequent behavioural variations (Minka 2005). Thus, it will look as if individuals have different priors, as they 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 divergences depend on an α parameter that controls the strength of the bound and induces different posterior estimates. Consequently, the resultant system behaviour may vary, and point towards different priors that could have altered the variational posterior form. For this, we have to 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 value. This entails a natural space of explanation for observed behavioural variation. To make the link to behaviour, we will assume a prior preference for particular values of the observable outcome. The distribution from which actions are sampled attributes greater probability to those actions for which the anticipated distribution over outcomes maximises the expected (log) prior preferences i.e., softmax action strategy (Sutton & Barto 1998). Explicitly, under our Rényi bound definition, small α values lead to mass-seeking approximate[2] posteriors i.e., greedy preferences for a particular outcome. Conversely, $\alpha \to \infty$ can result in mode-covering approximate posteriors —resulting in a greater range of actions for which there are plausible outcomes consistent with prior preferences. We reserve further description of this for later sections. Hence, variable

[1] Indeed, previous work has developed divergence-based lower bounds that give tighter bounds e.g., Barber D, de van Laar P. 1999. Variational cumulant expansions for intractable distributions. *Journal of Artificial Intelligence Research* 10: 435-55 However, these can be more difficult to optimise despite better approximations.

[2] Here, we use approximate and variational posterior interchangeably.

individual preferences could be attributed to differences in the variational optimisation objective. This contrasts with standard variational objective where precision over priors is altered to illustrate these divergent behaviour profiles. In what follows, we present, and validate, this departure from standard variational inference that can explain behaviour 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 2.0). Section 3.0 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 under different priors (section 4.0). In section 5.0, we validate the approach through a series of simulations in gaussian systems and a multi-armed bandit setting with bimodal observation distribution. Our simulations demonstrate that variational Bayesian agents, optimising a generalised variational bound (i.e., Rényi bound) undergird by different α values, can naturally account for variations in choice behaviour. We conclude with a brief discussion about the implications of our work in understanding behavioural variations and future directions.

2.0 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 et al 1999, Wainwright & Jordan 2008). It works by defining 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 most practical applications that characterize the brain as performing Bayesian inference including predictive coding (Millidge et al 2020, Spratling 2017, Whittington & Bogacz 2017), active inference (Da Costa et al 2020, Friston et al 2017, Sajid et al 2021, Tschantz et al 2020), etc. and take the perspective of variational inference as part of it.

2.1 KL-divergence and the variational free energy

To derive the variational free energy, we consider a simple system with two random variables. These are $\boldsymbol{s \epsilon S}$ denoting hidden states of the system (e.g., it rained last night) and $\boldsymbol{o \epsilon O}$ the observations (e.g., the grass is wet). The joint density over these variables:

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

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

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

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[3]. For this, we introduce a variational density $q(.)$ that can be easily integrated. The following equations illustrate how through a few simple moves we can derive the quantities of interest. We assume that $p(s \mid o)$ and $q(s)$ are non-zero:

$$\begin{aligned}
\log p(o) &= \log p(o) + \int \log \frac{p(s \mid o)}{p(s \mid o)} ds \\
&= \int q(s) \log p(o) ds + \int q(s) \log \frac{p(s \mid o)}{p(s \mid o)} ds \\
&= \int q(s) \log \frac{p(s \mid o)\, p(o)}{p(s \mid o)} ds = \int q(s) \log \frac{p(s,o)}{p(s \mid o)} ds \\
&= \int q(s) \log \frac{q(s)}{q(s)} ds + \int q(s) \log p(s,o) ds + \int q(s) \log \frac{1}{p(s \mid o)} ds \\
&= \underbrace{\int q(s) \log \frac{1}{q(s)} ds + \int q(s) \log p(s,o) ds}_{ELBO} + \underbrace{\int q(s) \log \frac{q(s)}{p(s \mid o)} ds}_{KL-Divergence}
\end{aligned} \tag{3}$$

The first two summands of the last equality are the evidence lower bound (Welbourne et al), and the last summand presents the (reverse) KL-divergence between the approximate and true posterior. If $q(.)$ and $p(.)$ 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 (F) which upper bounds the negative log evidence. Therefore, we rewrite the last equality:

[3] 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 (*Wainwright MJ, Jordan MI. 2008. Graphical models, exponential families, and variational inference. Foundations and Trends® in Machine Learning 1: 1-305*).

$$
\begin{aligned}
-\log p(o) &= -\left[\int q(s) \log \frac{1}{q(s)} ds + \int q(s) \log p(s,o) ds + \int q(s) \log \frac{q(s)}{p(s \mid o)} ds \right] \\
&\leq -\int q(s) \log p(s,o) ds + \int q(s) \log q(s) ds \\
&= -\left\langle \log p(s,o) \right\rangle_{q(s)} + H(q(s)) \\
&= \underbrace{D_{KL}\left[q(s) \,\|\, p(s) \right]}_{complexity} - \underbrace{\left\langle \log p(o \mid s) \right\rangle_{q(s)}}_{accuracy} = F
\end{aligned}
\tag{4}
$$

The last line is the commonly presented decomposition of the variational free energy summands: complexity and accuracy (Friston 2010, Sajid et al 2020b). The accuracy term represents how well observed data can be predicted and complexity is a regularisation term. The variational free energy objective favours accurate explanations for sensory observations that are maximally consistent with prior beliefs.

In this setting, illustrations of behavioural variations i.e., differences in variational posterior estimations can result from altered priors over the (hyper-)parameters of the generative model e.g., change in precision over the likelihood function (Friston et al 2014b). We reserve a more detailed description of hyper-prior parameterisations and their impact on belief updating for section 4.0.

3.0 Rényi divergences, and their variational bound

We are interested in defining a (general) variational objective that can offer an alternate account of behavioural variations. For this, we may relax the KL divergence to general divergence objectives, i.e., non-negative functions $D[\bullet \| \bullet]$ that satisfy $D\left[q(s) \,\|\, p(s \mid o)\right] = 0$ if and only if $q(s) = p(s \mid o)$ for all s. For our purposes, we focus on Rényi divergences, a general class of divergences that comprises the KL-divergence (Table 1). This has the advantage of being computationally tractable, and satisfies many additional properties (Amari 1985, Amari 2012, Rényi 1961, Van Erven & Harremos 2014). Renyi-divergence is defined as (Li & Turner 2016, Rényi 1961):

$$
D_{\alpha}\left[p(s \mid o) \,\|\, q(s) \right] \doteq \frac{1}{\alpha - 1} \log \int p(s \mid o)^{\alpha} q(s)^{1-\alpha} ds \tag{5}
$$

where $\alpha \in \mathbb{R} + \backslash\{1\}$

An analogous definition holds for the discrete case, by replacing the densities with probabilities and the integral by a sum.

This family of divergences can provide different posterior estimates, at the minimum of the divergence with respect to *q*, that vary smoothly with α. 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, e.g. a Gaussian distribution. In this multimodal scenario, small positive α -parameter ($\alpha < 1$) in the Rényi divergence will favour

distributions that best fit the mode with the most mass (not necessarily the highest). This happens because forces the approximate posterior to be small (e.g., $q(.) = 0$), whenever the true posterior is small (i.e., zero-forcing). This causes parts of the true posterior to be excluded. Conversely, with $\alpha \to \infty$ values, the approximate posterior aims to cover the entire true posterior. This happens because the approximate posterior is forced to be $q(.) > 0$, whenever the true posterior is $p(.) > 0$ (i.e., zero-avoiding). This implies that low probability regions of the true posterior may be overestimated.

Other (non-Rényi) divergences in the literature are also parameterized by α: the I divergence, Amari's α-divergence and the Tsallis divergence. All of these divergences are equivalent in that the value of one can be deduced from another, see appendix A1. This allows the results presented in this paper to be mapped to these different divergences along the correspondences drawn in appendix A1.

3.1 Rényi bound

The accompanying variational bound for these 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 & Turner 2016):

$$p(o) = \frac{p(o,s)}{p(s \mid o)} \Rightarrow$$

$$p(o)^{1-a}\, p(s \mid o)^{1-a} = p(o,s)^{1-a}$$

$$\int q(s)^{\alpha}\, p(o)^{1-a}\, p(s \mid o)^{1-a} ds = \int q(s)^{\alpha}\, p(o,s)^{1-a} ds$$

$$\log \int q(s)^{\alpha}\, p(o)^{1-a}\, p(s \mid o)^{1-a} ds = \log \int q(s)^{\alpha}\, p(o,s)^{1-a} ds$$

$$\log p(o)^{1-a} + \log \int q(s)^{\alpha}\, p(s \mid o)^{1-a} ds = \log \int q(s)^{\alpha}\, p(o,s)^{1-a} ds$$

$$\log p(o)^{1-a} = \log \left[\frac{\int q(s)^{\alpha}\, p(o,s)^{1-a} ds}{\int q(s)^{\alpha}\, p(s \mid o)^{1-a} ds} \right]$$

$$\log p(o) = \frac{1}{1-a} \log \left[\frac{\int q(s)^{\alpha}\, p(o,s)^{1-a} ds}{\int q(s)^{\alpha}\, p(s \mid o)^{1-a} ds} \right]$$

$$\log p(o) = \underbrace{\frac{1}{1-a} \log \int q(s)^{\alpha}\, p(o,s)^{1-a} ds}_{\text{Renyi Bound}} + \underbrace{\frac{1}{a-1} \log \int q(s)^{\alpha}\, p(s \mid o)^{1-a} ds}_{\text{Renyi Divergence } D_{\alpha}[q(s) \| p(s \mid o)]} \tag{7}$$

We assume that $p(s \mid o)$ and $q(s)$ are non-zero and $\alpha \in \mathbb{R} + \backslash\{1\}$. The negative Rényi bound can be regarded as being analogous to the variational free energy objective (F) (Eq~4). The following equations highlight this correspondence:

$$-\log p(o) = -\left[\frac{1}{1-\alpha}\log\int q(s)^{\alpha}\, p(o,s)^{1-a}ds + \frac{1}{\alpha-1}\log\int q(s)^{\alpha}\, p(s\mid o)^{1-a}ds\right] \leq -\frac{1}{1-\alpha}\log\int q(s)^{\alpha}\, p(o,s)^{1-a}ds \quad (8)$$

where $\alpha \in \mathbb{R}+\backslash\{1\}$.

Table 1 provides an overview of this, and other examples of the negative Rényi bound.

α	**Definition:** $D_{\alpha}\left[p(s\mid o)\,\|\,q(s)\right]$	**Rényi Bound:** $-D_{\alpha}\left[q(s)\,\|\,p(s,o)\right]$	**Comment**
$\alpha \to 1$	$\int p(s\mid o)\log\frac{p(s\mid o)}{q(s)}ds$	$-D_{KL}\left[q(s)\,\|\,p(s)\right]+\left\langle\log p(o\mid s)\right\rangle_{q(s)}$ $\left\langle\log q(s)\right\rangle_{p(s,o)} - H\left(p(s,o)\right)$	Kullback-Leibler (KL) divergence: $D_{KL}[q\,\|\,p]$ & $D_{KL}[p\,\|\,q]$. The first bound is for $-D_{KL}\left[q\,\|\,p\right]$, and the second bound is for $-D_{KL}\left[p\,\|\,q\right]$
$\alpha = 0.5$	$-2\log\left(1-Hel^{2}\left[p(s\mid o),q(s)\right]\right)$ $-2\log\int\sqrt{p(s\mid o)q(s)}ds$	$2\log\left(1-Hel^{2}\left[p(s,o),q(s)\right]\right)$ $2\log\int\sqrt{p(s,o)q(s)}ds$	(First equation) Function of the square Hellinger distance: $Hel^{2}(p,q)\doteq 1-\int\sqrt{p(s\mid o)q(s)}ds$ (Birgé 1986, Gibbs & Su 2002). (Second equation) It can be also derived as the Bhattacharyya divergence. Both are symmetric in their arguments.(Van Erven & Harremos 2014)
$\alpha = 2$	$\log\left[1+\chi^{2}\left(p(s\mid o)\,\|\,q(s)\right)\right]$	$-\log\left[1+\chi^{2}\left(q(s)\,\|\,p(s,o)\right)\right]$	Proportional to χ^{2}-divergence: $\chi^{2}(p,q)\doteq\int\frac{(p-q)^{2}}{q}ds=\int\frac{p^{2}}{q}ds-1$ (Gibbs & Su 2002).
$\alpha \to +\infty$	$\log\max_{s\in S}\frac{p(s\mid o)}{q(s)}$	$-\log\max_{s\in S}\frac{q(s)}{p(s,o)}$	Minimum description length or the worst case regret of using q instead of p (Grünwald 2007, Hinton & Zemel 1994)

Table 1: Examples of (normalized) Rényi divergences (Li & Turner 2016, Minka 2005, Van Erven & Harremos 2014) for different values of α, and the accompanying (negative) Rényi bounds. We omit $a \to 0^{+}$ because it is not a divergence. These particular divergences have a

non-decreasing order: $Hel^2(p,q) \leq D_{1/2}[p \| q] \leq D_1[p \| q] \leq D_2[p \| q] \leq \chi^2(p,q)$ (Van Erven & Harremos 2014). This implies that sometimes it might be easier to derive risk bounds for $D_{1/2}[p \| q]$, compared to $D_1[p \| q]$, as empirically demonstrated in (Li & Turner 2016) (See appendix Figure A2).

We would expect similar variations in the estimation of the approximate posterior under the Rényi bound as with the Rényi divergences. This follows from the inclusivity (or zero-forcing) and exclusivity (or zero-avoiding) properties of the Rényi bound (Li & Turner 2016). Hence, the Rényi bound should provide a formal account of behavioural differences through different α parameters. That is, we would expect a natural shift in behavioural preferences as we move from small α values to large positive α values, given a consistent set of priors. Section 5.3 demonstrates this shift in preferences for a multi-armed bandit setting.

4.0 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 priors under a standard variational objective. We examine the relationship between the Rényi bound and the variational free energy on a simple Gaussian system to see whether the same kinds of behaviour can be produced through the standard variational objective (Eq~4) given particular prior beliefs or the Rényi bound (Eq~8) with constant prior beliefs but altered α value. If this were to be the case, we would be able to re-write the variational free energy under different precision hyperpriors as the Rényi bound, where hyperparameters 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 stationary points) – with respect to the posterior under some different priors or α value.

Interestingly, it is the same kind of prior beliefs that must be assumed when modelling variations in behavioural choices. This speaks to different priors where agents can exhibit conservative or greedy choice behaviour. Practically, this may be a result of either lending one contribution more precision through weighting the log probability under the standard variational objective (Eq~4), or by indirectly altering the priors by taking the log of the probability to the power of alpha (Eq~8). To illustrate these equivalences, we consider a simple system. We start by deriving the exact variational free energy for a gaussian system with gamma priors over the variance. Conversely, to derive the exact Rényi bound for a similar system, we assume a simple gaussian parameterisation. The differences in parameterisations allow us to cast the precision prior as being equivalent to the α parameter i.e., one must either alter the precision prior or the α values to demonstrate behavioural differences.

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 states of the system and $o \in O$ the observations (Figure 1). The variational family is parameterised as a Gaussian. This is formalised as:

$$\begin{aligned} p(s \mid \lambda_p) &\sim N\left(s; 0, (\lambda_p \Lambda_p)^{-1}\right) Gam(\lambda_p; \alpha_p, \beta_p) \\ p(o \mid s, \lambda_l) &\sim N\left(o; sx, (\lambda_l \Lambda_l)^{-1}\right) Gam(\lambda_l; \alpha_l, \beta_l) \\ q(s) &\sim N\left(s; \mu_q, \Sigma_q^{-1}\right) \end{aligned} \tag{9}$$

where $\Sigma_k = (\lambda_k \Lambda_k)^{-1}$. We use these quantities to derive the (negative) variational free energy i.e., ELBO (Appendix 2.0 for the derivation):

$$\begin{aligned} -F = &-\frac{1}{2}\left[\mu_q{}^T (\Sigma_{KL} - \Sigma_q^{-1}) \mu_q - 2\mu_q{}^T \Sigma_{KL}\, \mu_{KL} + tr(\Sigma_{KL}\, \Sigma_q^{-1}) + o^T \Sigma_l^{-1} o + 2\lambda_p \beta_p + 2\lambda_l \beta_l \right] \\ &+ \frac{1}{2} \log\left[\frac{|\Sigma_l^{-1}||\Sigma_p^{-1}|}{|\Sigma_q^{-1}| 2\pi^{(m-n)}} \right] + \log \frac{\lambda_l{}^{(a_l - 1)} \beta_l^{a_l} \lambda_p{}^{(a_p - 1)} \beta_p^{a_p}}{\Gamma(a_l)\Gamma(a_p)} \end{aligned} \tag{10}$$

where:
$$\begin{aligned} \Sigma_{KL} &= \Lambda_p \lambda_p + \lambda_l x^T \Lambda_l x - \Sigma_q^{-1} \\ \mu_{KL} &= \Sigma_{KL}{}^{-1} (\lambda_l x^T \Lambda_l o - \Sigma_q^{-1} \mu_q) \end{aligned}$$

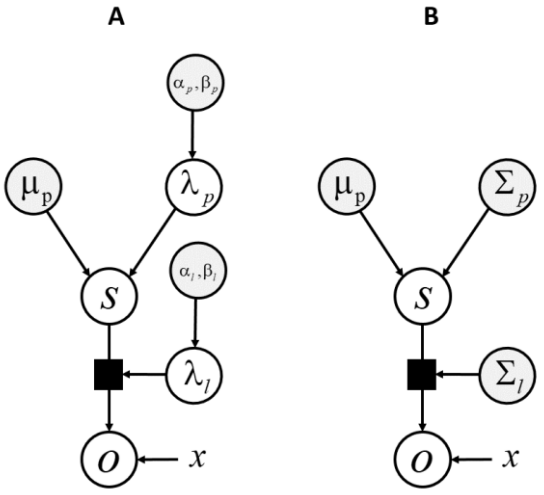

Figure 1. Graphical model for the Gaussian-Gamma (A), and Gaussian (B) system. Here white circles represent random variables, grey circles represent priors and x is the parameter governing the mean.

4.2 Rényi-bound for a Gaussian system

Next, we consider a similar system for deriving the Rényi-bound. Here, we do not need to include priors that shape preferences. Therefore, the densities are parameterised using Gaussians:

$$\begin{aligned} p(s) &\sim N\left(s;0,\Sigma_p\right) \\ p(o\mid s) &\sim N\left(o;sx,\Sigma_l\right) \\ q(s) &\sim N\left(s;\mu_q,\Sigma_q\right) \end{aligned} \tag{11}$$

We use these quantities to derive the Rényi bound, R_B (Appendix 3.0 for the derivation):

$$\begin{aligned} R_B = &\frac{1}{2(1-\alpha)}\left({\mu_\alpha}^T{\Sigma_\alpha}^{-1}\mu_\alpha - \alpha{\mu_q}^T\Sigma_q^{-1}\mu_q + (\alpha-1)o^T\Sigma_l^{-1}o\right) \\ &+\frac{1}{2(1-\alpha)}\log\left[\frac{|\Sigma_\alpha|}{2\pi^{(1-\alpha)(m-n)}|\Sigma_l|^{1-\alpha}|\Sigma_p|^{1-\alpha}|\Sigma_q|^{\alpha}}\right] \end{aligned} \tag{12}$$

where:
$$\begin{aligned} \Sigma_\alpha &= \left((1-\alpha)(\Sigma_p^{-1}+x^T\Sigma_l^{-1}x)+\alpha\Sigma_q^{-1}\right)^{-1} \\ \mu_\alpha &= \Sigma_\alpha\left(\alpha\Sigma_q^{-1}\mu_q+(1-\alpha)x^T\Sigma_l^{-1}o\right) \end{aligned}$$

under the assumption that Σ is positive-definite. Accordingly, for a univariate Gaussian system the following:

$$\Sigma_q < \frac{\alpha}{(\alpha-1)}\underbrace{\begin{bmatrix}\Sigma_p^{-1} & \Sigma_p^{-1}x^T \\ \Sigma_p\Sigma_p^{-1} & \Sigma_l^{-1}+x\Sigma_p^{-1}x^T\end{bmatrix}}_{\mathrm{cov}(o,s)} \tag{13}$$

must be satisfied for invertibility (Burbea 1984, Metelli et al 2018). Therefore, for the numerical simulations we constrain $\Sigma_q < \Sigma_l$. This is perfectly reasonable since the univariate gaussian variational posterior can perfectly approximate the true posterior for all alpha values.

4.3 Numerical analysis

Using the derived bounds (Eq~10,12), we illustrate how the variational objective may change as the appropriate priors or the α value is altered, ceteris paribus (Figure 1). Our numerical simulation illustrates that optimisation of these objectives can lead to (extremely) different estimated variational density.

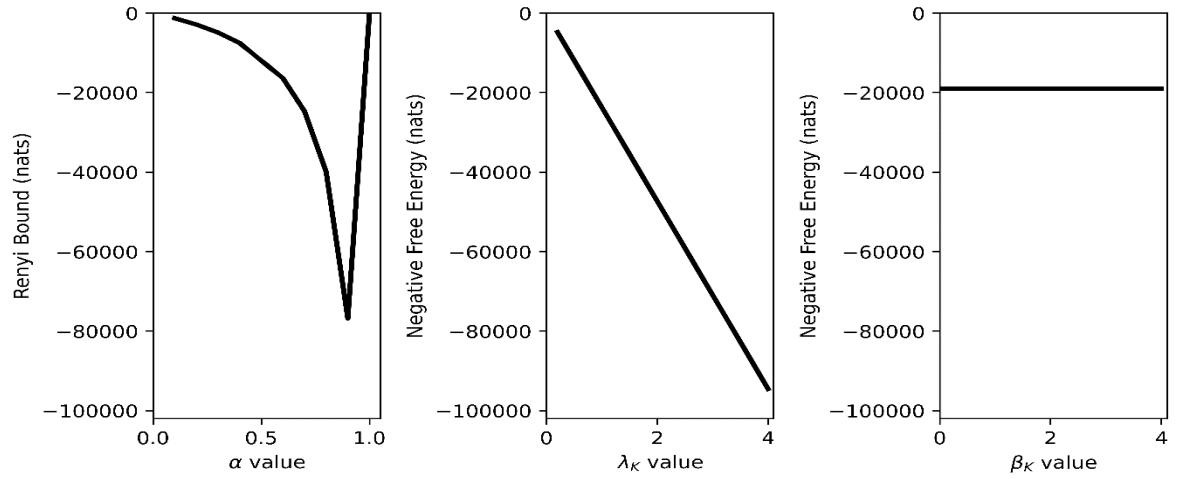

Figure 1 (*Correspondence between Rényi bound and (negative) variational free energy*) The graphs plot the exact bound for the varying priors or α value for the systems described in Section 4.1 and 4.2, at a particular μ_q. In each plot everything, except the stated priors or α value, was kept consistent. First panel shows the Rényi bound (y-axis) for different α -values (x-axis); this is limited1.0 since $\alpha > 1$ cannot be calculated in closed form. Panel 2 (and 3) plot the (negative) variational free energy (y-axis) for different β_k (and λ_k) values. Here, k denotes all three λ or β quantities: $\lambda_p, \lambda_l, \lambda_q$ and $\beta_p, \beta_l, \beta_q$. For 2-D illustration purposes, here all values of either λ or β were set to the same number.

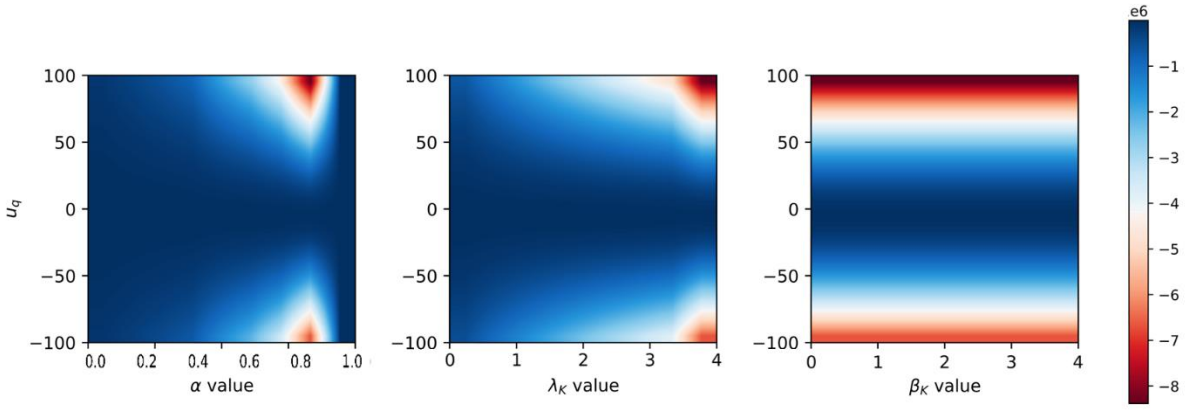

Figure 2 (*Variational optimisation landscape*) These graphics plot the optimisation landscape for changing priors or α values for different variational posteriors (i.e., μ_q). First panel plots it for the Rényi bound, where the x-axis denotes the different α values and y-axis the μ_q. Similarly, the next two panels plots the (negative) free energy, where the x-axis denotes the different λ_k / β_k values and y-axis the μ_q. The variational bound quantities range from 0 nats (dark blue) to -100000 nats (red). For 2-D illustration purposes, here all values of either λ or β were set to the same number.

Further investigations of the variational optimisation landscape for different values of λ_k / β_k or α (presented in Figure 1for $\mu_q = 100$) under different variational means, μ_q, are shown in Figure 2. The analysis illustrates that although there can exist similarities in optimisation landscapes (and the variational posterior) for certain priors or α value, this is constrained to a small subspace of posterior estimates. Outside these posterior values, the variational optimisation landscape diverges. These numerical analyses demonstrate that the Rényi divergences account for behavioural differences in a different way, than a change in priors, through manipulation of the α parameter. Conversely the standard variational objective could require multiple alterations to the (hyper-)parameters altered to exhibit 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.

5.0 Simulations

In this section, we illustrate how α value changes may impact variational posterior estimates, and consequent behaviour. For this, we simulated inferences drawn from data sampled from a bi-modal system, comprising a mixture of Gaussians, with a Gaussian variational density, under a mean-field assumption[4]. Therefore, the true posterior will not be in the same family of distributions as the assumed variational posterior. This specification allows us to illustrate how different α values lead to differences in variational posterior estimations (and choice behaviour). For simulations where α parameter changes are irrelevant for modelling behavioural variations see appendix 2.0.

Next, to illustrate behavioural variations, we simulated the multi-armed bandit setting using multivariate gaussian parameterisations for three different arms. In what follows, we optimise the variational objectives, i.e., *Rényi* bound, with respect to the approximate density.

5.1. Rényi bounds for a multimodal system

We simulate inference about a multimodal system with different α values to elucidate differences in variational posterior estimation. The system is parameterised as:

$$
\begin{aligned}
&p(s) \sim \sum_i^2 \varpi_i N\left(\mu_i, \Sigma_i\right) \\
&p(o,s) \sim N\left(\mu_p, 1.0\right) p(s) \\
&q(s) \sim N(\mu_q, \Sigma_q)
\end{aligned} \tag{17}
$$

$$\sum_i^2 \varpi_i = 1$$

Here, the variational density, $q(s)$, form is constrained as Gaussian, with an arbitrary mean and variance. However due to the multivariate gaussian prior, the true posterior can take a more

[4] Mean-field assumption: explain

complex form. Thus, the true posterior may not be in the same family of distributions as the assumed variational posterior. This causes differences in posteriors to be made evident under different Rényi bounds. Under this setting, we can simulate scenarios where different α values lead to behavioural variations.

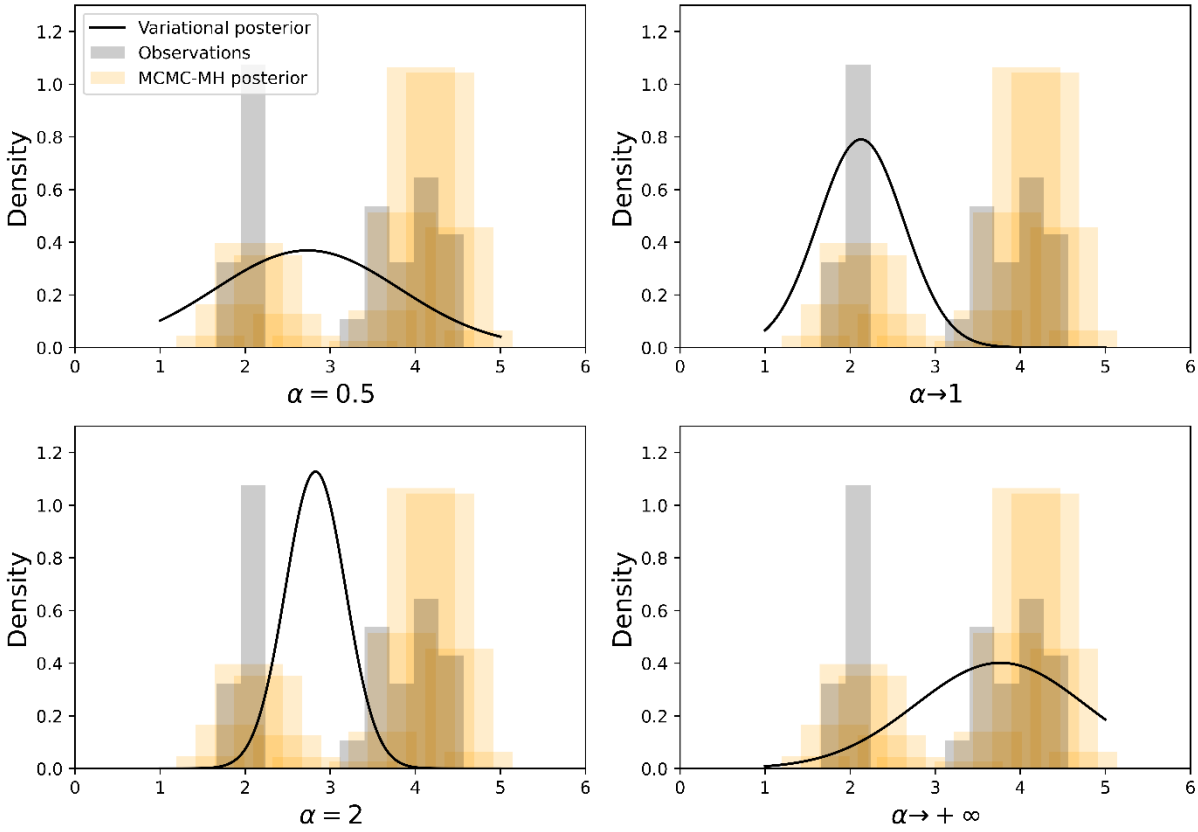

Figure 3. (*Estimated variational posteriors for a mixture of gaussian system*) The figures plot the variational posteriors, under different variational objectives, parameterised by the α. We plot results for: $\alpha \in \{0.5, 1^+, 2, +\infty\}$. Each plot shows the *s* value on the x-axis, and its density on the y-axis. The black line is the variational posterior at the minimum of the Renyi bound, the orange bars are the Metropolis-Hastings sampling posterior and grey bars represent the samples from a mixture of Gaussians with weights [0.6, 0.4]. For each bound, we see different variational posteriors after an appropriate number of gradient descent iterations.

	Mean (μ)	**Variance (σ)**
$\alpha = 0.5$	2.7335	1.0824
$\alpha \to 1^+$	2.1276	0.5048
$\alpha = 2$	2.8272	0.3539

$\alpha \to +\infty$	3.7665	0.9940

Table 2: Estimated mean and variance of the Gaussian variational posterior for a mixture of Gaussian system using different α.

Using this system, we estimate the gaussian variational posterior, with four different Rényi bounds parameterised by: $\alpha \in \{0.5, 1^+, 2, +\infty\}$ (Table 1). Appendix 4.0 presents the pseudocode for the optimisation procedure. Everything is kept consistent across each simulation i.e., had it not been for the change in the α value they would be no difference between the simulations. The results from this are shown in Figure 3 and Table 2. This analysis illustrates the differences in posteriors produced by the different α values. For example, the $\alpha \to 1^+$ approximates one mode – consequently, underestimating the variational posterior.

Whilst simple, the bimodal system highlights interesting behavioural shifts that might arise when the variational posterior is not in the same family of distributions as the true posterior.

5.2. Rényi bounds for multi-armed bandits

To make the behavioural differences concrete, we simulate the multi-armed bandit setting using multivariate gaussian parameterisations for three different arms i.e., stationary stochastic bandit. We treat this as *planning as inference* problem (Botvinick & Toussaint 2012); allowing us to accommodate and illustrate the differential preferences that arise naturally under the Rényi bound. However, these differences are context dependent.

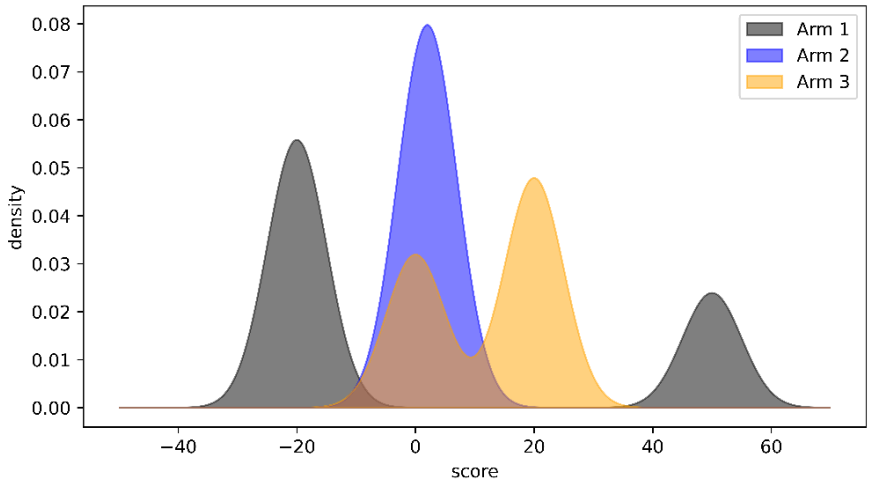

Figure 4. (*Score distributions*) The figure plots the score distributions for the three arms of the multi-armed bandit. The x-axis denotes the score and y-axis the score density. Here, black is for Arm 1, blue for Arm 2 and orange for Arm 3.

The environment set-up is straightforward, at each round, a particular arm is pulled and an observation (i.e., score) corresponding to this is received. This is a one-state Markov Decision Process (MDP), such that regardless of which arm is pulled, environment will remain in the same state. Therefore, the task of the agent (system) is to judge which arm will give the appropriate observation and continue accumulating scores over X trials. Here each arm has a fixed observation distribution unknown to the agent. These distributions allow us to model behavioural variations. Each arm is modelled separately using the parameterisations presented in Eq~17. Under these distributions, Arm 2 has an expected return of 0, Arm 1 has an expected return of -13 and Arm 3 an expected return of 12. Accordingly, Arm 3 is the optimal/best policy under this MDP i.e., the arm with the maximal expected reward.

Using this setting, we illustrate how differences in optimisation objective results in behavioural variations across 1000 rounds of the agent pulling these particular arms. For this, we model four agents optimising the different Rényi bounds for α values: $0.5, \to 1^{+}, 2, \to \infty$. During each round, the Renyi bound is used to estimate the posterior distribution over each arm, and a softmax action strategy is used to select the arm. Appendix 5.0 details the pseudo-code for this. Each agent was simulated 50 times and the accumulated score is shown in Figure 5. The agents, where the Rényi bound α parameterisation is 2 or $\to \infty$ show the highest accumulated score. The behavioural variation can be explained by action selection differences; Arm 3 is preferred by these agents (Figure 6). Of these, agents with the Rényi bound objective of $\alpha \to \infty$ have a slightly higher accumulated score, but larger variance across trials (Figure 5 – red line). This particular variance is due to the differences in estimated posterior score for each arm. The agents optimising Rényi bound objective of $\alpha = 0.5$ and $\alpha \to 1^{+}$ have similar cumulative scores (Figure 5) but demonstrate different arm selection strategies (Figure 6). These agents have preferences for selecting Arm 3 as well, alongside different preferences over Arm 1. Agents optimising $\alpha = 0.5$ choose Arm 1 the highest number of times. Conversely, agents optimising $\alpha \to 1^{+}$ choose Arm 2 least amount of times.

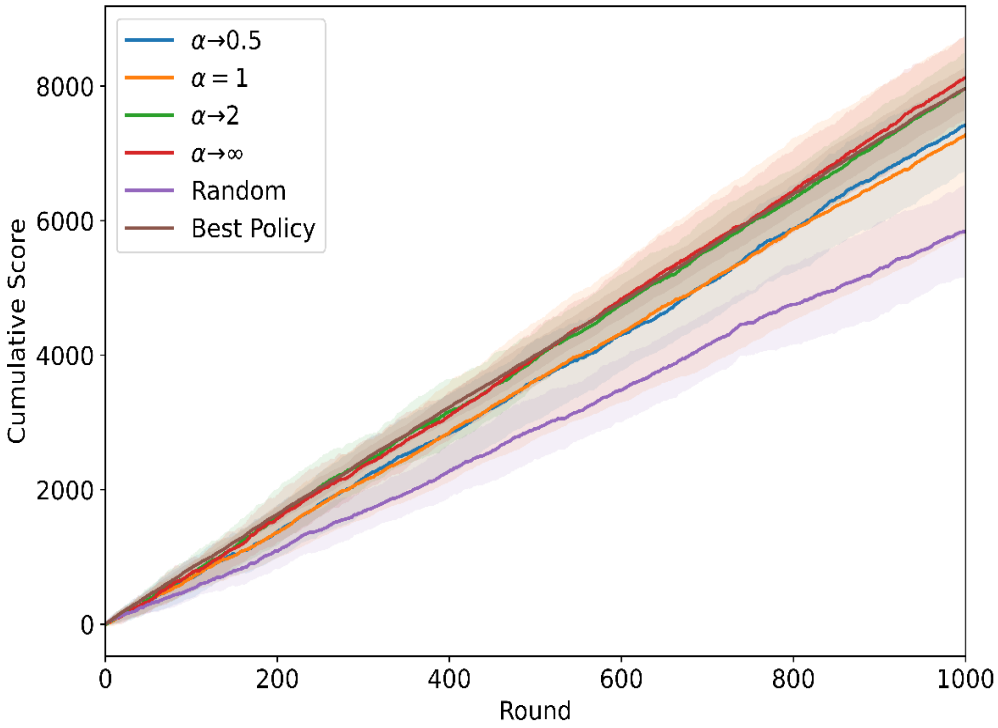

Figure 5. (*Cumulative score*) The figure shows the cumulative score across the 1000 rounds for each agent. Each agent was simulated 50 times, with the dark lines representing mean and shaded area the 95% confidence interval. The x-axis denotes the round and y-axis the accompanying cumulative score. Here, blue is for agents optimising Rényi Bound for $\alpha = 0.5$, orange for agents optimising Rényi Bound for $\alpha = 1$, green for agents optimising Rényi Bound for $\alpha = 2$ and red for agents optimising Rényi Bound for $\alpha \to \infty$. Purple and brown lines represent following a random (i.e., any arm) or best policy (i.e., selecting arm with the highest expected return), respectively.

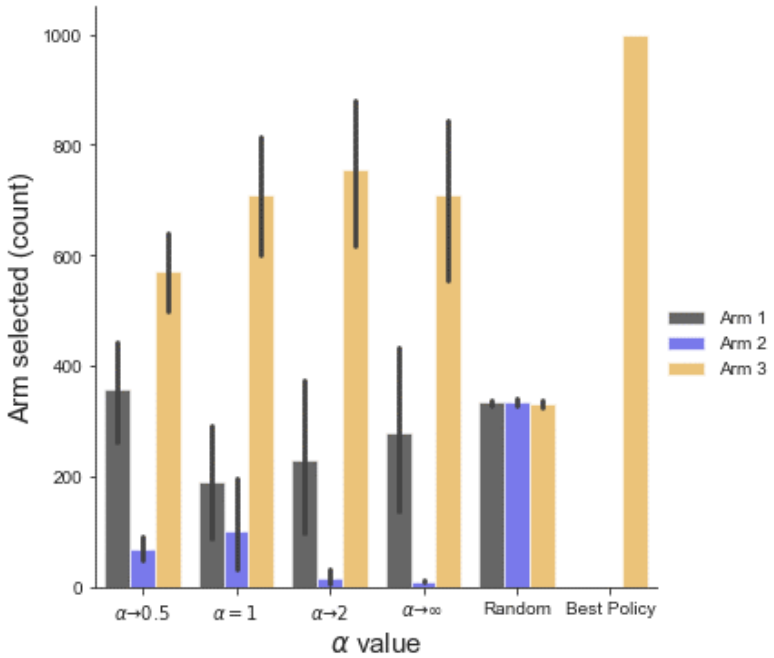

Figure 6. (*Arms selected*) The bars plot the arms selected by each agent across the 1000 trials. The height of the bar represents the total count of a particular arm being selected and the black lines show the 95% confidence internals. The x-axis denotes the α values, random (i.e., any arm) and best policy (i.e., selecting arm 2). The black bars denote Arm 1, blue bars are for Arm 2 and orange bars for Arm 3.

Our simulations highlight 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 variations in behaviour. In other words, the shape of the posterior will underwrite the ensuring behaviour, under planning as inference.

6.0 Discussion

This paper considers an alternative explanation for behavioural variations using Rényi divergences and their associated variational bounds. These divergences are Rényi relative entropies[5], and satisfy similar properties as the KL divergence (Rényi 1961, Van Erven & Harremos 2014). The divergences depend on an α parameter that controls the strength of the bound and induces different posterior estimates. Consequently, the resultant system behaviour may vary, and point towards different priors that could have altered the variational posterior form. For this, we have to assume that systems (or 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 value. This entails a natural space and explanation as to why some people are more risk averse and others not. In other words, the shape of the posterior predicated on the variational objective, will determine the kinds of preferences the agents have. Here, posteriors can be written down as a function of α by committing to variational Rényi bounds. A similar account would also be true if we appealed to expected free energy functional (Friston et al 2017, Parr & Friston 2019, van de Laar et al 2021), intrinsic reward (Schmidhuber 1991, Schmidhuber 2006, Sun et al 2011) or any class of objective functions that puts information gain and epistemic value into the game.

Rényi divergences account for behavioural differences in a different way than a change in priors i.e., through the α manipulations. This stems from the ability to disentangle different modes of preferences. Our simple multi-armed bandit setting illustrates this e.g., large α values exhibit greater consistency in preferences. Conversely, to standard variational objectives where priors over multiple different (hyper-) parameters needed to be altered to exhibit similar functionals, and behaviours. For example, under the active inference process theory multiple behavioural deficits have been illustrated by manipulation of the precision over the priors (Parr & Friston 2017, Sajid et al 2020a, Schwartenbeck & Friston 2016). Although there has been some focus upon priors, and some on the form of the variational posterior (Schwöbel et al

[5] The Rényi entropy provides a parametric family of information measures: Rényi A. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics1961*. The Regents of the University of California.

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 when interacting with its environment. Practically, this entails the optimisation of a variational functional to make appropriate predictions. However, there are no clear formulae for implementing such systems, and what sorts of things can be manipulated to show altered behaviour. Consequently, it is perfectly rational to appeal to Rényi bounds, instead of altered priors, to model any sort of behavioural variations. By committing to this 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. 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 investigate implementations of variational inference in the brain. That is, do we consider sub-optimal (i.e., altered priors) generative models and/or variable variational bound when modelling behavioural differences? This is particularly relevant in light of our numerical analysis (section 4.3) that shows no strict correspondence between the two formulations.

In a deep temporal system like the brain, it is important to speculate if different cortical hierarchies might be performing inference under different variational objectives. One might expect higher, and slower, levels to have fixed variational objectives. However, these could potentially be modulating the inferential procedure at the lower level, through priors over α values – via appropriate meta-inference procedures. This is analogous to including precision priors over model parameters that have been associated with different neuromodulatory systems e.g., state transition precisions with noradrenergic and sensory precision with cholinergic systems (Fountas et al 2020, Parr & Friston 2017). Consequently, this temporal separation of α parameterisations, provides an interesting research avenue, for investigating neuromodulatory systems and how they facilitate particular behaviours (Angela & Dayan 2002, Angela & Dayan 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 altering 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 et al 2016, Regli & Silva 2018), etc., which can improve the statistical properties of the variational bounds for particular applications (Wan et al 2020, Zhang et al 2019). Future work will generalize the arguments presented here and examine how these different divergences shape behaviour when planning as inference.

6.3 Limitations and future directions

We illustrate a weak correspondence between the Rényi bound and the variational free energy under particular priors. This is dependent on the variational posterior and specific α values. Additionally, we highlighted that multiple prior manipulations were required to even get this

correspondence. This link needs to be explored further. Our future work will look to investigate this in more complex systems to show what sorts of prior modifications are critical in establishing similar optimisation landscapes for different variational bounds. This will entail further exploring the association between the variational posterior and λ_k, β_k 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 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 doesn't 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 values does not necessarily impact the shape of the variational posterior – and by default 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 the different behaviours that biological (or artificial) agents might make – 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 et al 2019).

7.0 Conclusion

We offer an alternative account for explaining behavioural variations using Rényi divergences and their associated Rényi variational bounds. We show how the Rényi bound provides a way to establish behavioural differences given a generative model that is different than 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. We highlight that these bounds are analogous to the variational free energy (or evidence lower bound) and can be derived using the same assumptions. The formulation is illustrated through numerical analysis and demonstrates that positive α values give rise to risk averse behaviours and negative α values to risk sensitive behaviours when priors are held constant.

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

Funding: NS is funded by Medical Research Council (MR/S502522/1). LD is supported by the Fonds National de la Recherche, Luxembourg (Project code: 13568875). 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

Appendix

A1.0 Note on divergences indexed by α :

Many divergences in the literature are indexed with the parameter α (see Table A1). These divergences turn out to be equivalent to the Rényi divergence as we can identify one-to-one mappings between them. This appendix disambiguates these divergences and shows the correspondences between them.

Divergence	Formulation
I-divergence (Nielsen & Nock 2011)	$D_\alpha^I[p \,\|\, q] = \int p^\alpha q^{1-\alpha} ds$
Amari's α-divergence (Amari 1985, Amari 2009)	$D^{AM}{}_\alpha[p \,\|\, q] = \frac{4}{1-\alpha^2}\left(1 - \int p^{\frac{1+\alpha}{2}} q^{\frac{1-\alpha}{2}} ds\right)$
Tsallis' divergence (Nielsen & Nock 2011)	$D^T{}_\alpha[p \,\|\, q] = \frac{1}{\alpha-1}\left(\int p^\alpha q^{1-\alpha} ds - 1\right)$
Rényi divergence	$D_\alpha[p \,\|\, q] = \frac{1}{\alpha-1}\log \int p^\alpha q^{1-\alpha} ds$

Table A1. Other divergence families indexed with α.

Amari's α-divergence plays an important role in information geometry as it induces a dually-flat geometry on the space of probability measures, and furthermore, when extended to positive measures, it is the only intersection between f-divergences and Bregman divergences, two important families of divergences (Amari 2009, Ay & Gibilisco 2016).

All of the divergences shown in Table A.1 are equivalent, in the sense that there are one-to-one mappings between them.

The Tsallis and Amari's divergences are linear functions of the I-divergence:

$$D^{T}{}_{\alpha}[p \| q] = \frac{1}{\alpha - 1}\left(D^{I}_{\alpha}[p \| q] - 1\right)$$
$$D^{AM}{}_{\alpha}[p \| q] = \frac{4}{1-\alpha^2}\left(1 - D^{I}_{\frac{1+\alpha}{2}}[p \| q]\right) \quad \text{(A1)}$$

As a consequence, the Amari α-divergence is a scalar multiple of the Tsallis divergence, under the correspondence $\beta = \frac{1+\alpha}{2}$

$$D^{AM}{}_{\alpha}[p \| q] = \frac{1}{\beta} D^{T}{}_{\beta}[p \| q] \quad \text{(A2)}$$

Finally, the Rényi divergence is a monotonic function of the I-divergence:

$$D_{\alpha}[p \| q] = \frac{1}{\alpha - 1} \log D^{I}_{\alpha}[p \| q] \quad \text{(A3)}$$

A1.1 Duality of Rényi divergence:

$$D_{\alpha}[p \| q] = \frac{1}{\alpha - 1} \log \int p^{\alpha} q^{1-\alpha} ds$$
$$D_{1-\alpha'}[p \| q] = -\frac{1}{a'} \log \int p^{1-\alpha'} q^{\alpha'} ds$$
$$\frac{\alpha'}{1-\alpha'} D_{1-\alpha'}[q \| p] = \frac{1}{\alpha' - 1} \log \int p^{1-\alpha'} q^{\alpha'} ds \quad \text{(A4)}$$

$$\Rightarrow D_{\alpha}[p \| q] = \frac{\alpha}{1-\alpha} D_{1-\alpha}[q \| p]$$

where $\alpha \notin \{0^{+}, 1^{+}\}$ (Li & Turner 2016).

A2.0 Variational free energy for Gaussian-Gamma distribution

Here, we work through the variational free energy for the system described in Section 4.1. s, o are the random variables of interest, x the parameter governing the mean and $\sum_k$ is the precision parameter:

$$p(s \mid \lambda_p) \sim N\left(s; 0, (\lambda_p \Lambda_p)^{-1}\right) Gam(\lambda_p; \alpha_p, \beta_p)$$
$$p(o \mid s, \lambda_l) \sim N\left(o; sx, (\lambda_l \Lambda_l)^{-1}\right) Gam(\lambda_l; \alpha_l, \beta_l)$$
$$q(s) \sim N\left(s; \mu_q, (\lambda_q \Lambda_q)^{-1}\right)$$
$$\Sigma_k = (\lambda_k \Lambda_k)^{-1}$$

Where:

$$p(s \mid \lambda_p) = \frac{|\lambda_p \Lambda_p|^{1/2}}{2\pi^{n/2}} \exp\left[-\frac{\lambda_p}{2} s^T \Lambda_p s\right] \frac{\beta_p^{a_p}}{\Gamma(a_p)} \lambda_p^{a_p - 1} \exp\left[-\lambda_p \beta_p\right]$$

$$p(o \mid s, \lambda_l) = \frac{|\lambda_l \Lambda_l|^{1/2}}{2\pi^{n/2}} \exp\left[-\frac{\lambda_l}{2} (o - sx)^T \Lambda_l (o - sx)\right] \frac{\beta_l^{a_l}}{\Gamma(a_l)} \lambda_l^{a_l - 1} \exp\left[-\lambda_l \beta_l\right]$$

$$\Rightarrow p(o, s \mid \lambda_l, \lambda_p) \propto \exp\left[-(\lambda_p \beta_p + \lambda_l \beta_l)\right] \exp\left[-\frac{1}{2}\left(\lambda_l o^T \Lambda_l o + s^T (\Lambda_p \lambda_p + \lambda_l x^T \Lambda_l x) s - 2\lambda_l s^T x^T \Lambda_l o\right)\right]$$

$$q(s) = \frac{|\lambda_q \Lambda_q|^{1/2}}{2\pi^{n/2}} \exp\left[-\frac{1}{2}(s - \mu_q)^T \lambda_q \Lambda_q (s - \mu_q)\right]$$

The joint distribution for this system:

$$\begin{aligned}
&p(o, s \mid \lambda_l) \\
&= \frac{|\lambda_l \Lambda_l|^{1/2}}{2\pi^{n/2}} \exp\left[-\frac{\lambda_l}{2} (o - sx)^T \Lambda_l (o - sx)\right] \frac{\beta_l^{a_l}}{\Gamma(a_l)} \lambda_l^{a_l - 1} \exp\left[-\lambda_l \beta_l\right] \\
&\times \frac{|\lambda_p \Lambda_p|^{1/2}}{2\pi^{n/2}} \exp\left[-\frac{\lambda_p}{2} s^T \Lambda_p s\right] \frac{\beta_p^{a_p}}{\Gamma(a_p)} \lambda_p^{a_p - 1} \exp\left[-\lambda_p \beta_p\right] \\
&= \frac{|\lambda_l \Lambda_l|^{1/2} |\lambda_p \Lambda_p|^{1/2}}{2\pi^{m/2}} \cdot \frac{\beta_l^{a_l}}{\Gamma(a_l)} \frac{\beta_p^{a_p}}{\Gamma(a_p)} \lambda_p^{a_p - 1} \lambda_l^{a_l - 1} \\
&\exp\left[-\frac{1}{2}[\lambda_l (o - sx)^T \Lambda_l (o - sx) + \lambda_p s^T \Lambda_p s]\right] \exp\left[-(\lambda_p \beta_p + \lambda_l \beta_l)\right] \\
&\propto \exp\left[-(\lambda_p \beta_p + \lambda_l \beta_l)\right] \exp\left[-\frac{1}{2}\left(\lambda_l o^T \Lambda_l o + s^T (\Lambda_p \lambda_p + \lambda_l x^T \Lambda_l x) s - 2\lambda_l s^T x^T \Lambda_l o\right)\right]
\end{aligned} \tag{A5}$$

Note, $p(o, s)$ is a quadratic function of its components, and a Gaussian distribution.We use these probability distributions to derive the quantity of interest:

$$\begin{aligned}
&E_{q(s)}\left[\log p(s, o) - \log q(s)\right] \\
&= E_{q(s)}\left[\log\left(\frac{|\lambda_l \Lambda_l|^{1/2} |\lambda_p \Lambda_p|^{1/2}}{2\pi^{m/2}} \cdot \frac{\beta_l^{a_l}}{\Gamma(a_l)} \frac{\beta_p^{a_p}}{\Gamma(a_p)} \lambda_p^{a_p - 1} \lambda_l^{a_l - 1} \cdot \right.\right. \\
&\left.\left. \exp\left[-(\lambda_p \beta_p + \lambda_l \beta_l)\right] \exp\left[-\frac{1}{2}\left(\lambda_l o^T \Lambda_l o + s^T (\Lambda_p \lambda_p + \lambda_l x^T \Lambda_l x) s - 2\lambda_l s^T x^T \Lambda_l o\right)\right]\right)\right] \\
&- E_{q(s)}\left[\log\left(\frac{|\lambda_q \Lambda_q|^{1/2}}{2\pi^{n/2}} \exp\left[-\frac{1}{2}(s - \mu_q)^T \lambda_q \Lambda_q (s - \mu_q)\right]\right)\right]
\end{aligned}$$

$$= E_{q(s)}\left[\log\frac{|\lambda_l\Lambda_l|^{1/2}|\lambda_p\Lambda_p|^{1/2}}{|\lambda_q\Lambda_q|^{1/2}\,2\pi^{(m-n)/2}}\cdot\frac{\beta_l^{a_l}\beta_p^{a_p}}{\Gamma(a_l)\Gamma(a_p)}.\lambda_p^{a_p-1}\lambda_l^{a_l-1}\right]+E_{q(s)}\left[-(\lambda_p\beta_p+\lambda_l\beta_l)\right]$$
$$+E_{q(s)}\left[-\frac{1}{2}\left(\lambda_l o^T\Lambda_l o+s^T(\Lambda_p\lambda_p+\lambda_l x^T\Lambda_l x)s-2\lambda_l s^T x^T\Lambda_l o-(s-\mu_q)^T\lambda_q\Lambda_q(s-\mu_q)\right)\right]$$
$$= E_{q(s)}\left[\log\frac{|\lambda_l\Lambda_l|^{1/2}|\lambda_p\Lambda_p|^{1/2}}{|\lambda_q\Lambda_q|^{1/2}\,2\pi^{(m-n)/2}}\cdot\frac{\beta_l^{a_l}\beta_p^{a_p}}{\Gamma(a_l)\Gamma(a_p)}.\lambda_p^{a_p-1}\lambda_l^{a_l-1}\right]+E_{q(s)}\left[-(\lambda_p\beta_p+\lambda_l\beta_l)-\frac{1}{2}\lambda_l o^T\Lambda_l o\right]$$
$$+E_{q(s)}\left[-\frac{1}{2}\left(s^T(\Lambda_p\lambda_p+\lambda_l x^T\Lambda_l x)s-2\lambda_l s^T x^T\Lambda_l o-(s-\mu_q)^T\lambda_q\Lambda_q(s-\mu_q)\right)\right]$$

(A6)

First, let us focus on the term inside the third term of the second equality. To avoid clutter we replace:

$$\Lambda_{KL}=\Lambda_p\lambda_p+\lambda_l x^T\Lambda_l x$$
$$\Sigma_q=(\lambda_q\Lambda_q)^{-1} \tag{A7}$$

We now substitute this below:

$$E_{q(s)}\left[-\frac{1}{2}\left(s^T\Lambda_{KL}s-2\lambda_l s^T x^T\Lambda_l o-(s-\mu_q)^T\Sigma_q^{-1}(s-\mu_q)\right)\right]$$
$$= E_{q(s)}\left[-\frac{1}{2}\left(s^T\Lambda_{KL}s-2\lambda_l s^T x^T\Lambda_l o-s^T\Sigma_q^{-1}s+2s^T\Sigma_q^{-1}\mu_q-\mu_q^T\Sigma_q^{-1}\mu_q\right)\right] \tag{A8}$$
$$= E_{q(s)}\left[-\frac{1}{2}\left(s^T\left(\Lambda_{KL}-\Sigma_q^{-1}\right)s-2s^T\left(\lambda_l x^T\Lambda_l o-\Sigma_q^{-1}\mu_q\right)-\mu_q^T\Sigma_q^{-1}\mu_q\right)\right]$$

We can now complete the square in s to obtain an unnormalized Gaussian. For this, we replace with the following:

$$\sum_{KL}=\Lambda_{KL}-\Sigma_q^{-1}$$
$$\mu_{KL}=\sum_{KL}^{-1}(\lambda_l x^T\Lambda_l o-\Sigma_q^{-1}\mu_q) \tag{A9}$$

This gives:

$$= E_{q(s)}\left[-\frac{1}{2}\left((s-\mu_{KL})^T\sum_{KL}(s-\mu_{KL})-\mu_{KL}{}^T\sum_{KL}\mu_{KL}-\mu_q^T\Sigma_q^{-1}\mu_q\right)\right] \tag{A10}$$

Now, since the $(s-\mu_{KL})^T\sum_{KL}(s-\mu_{KL})\in\mathbb{R}$, we can write it as $tr\{(s-\mu_{KL})(s-\mu_{KL})^T\sum_{KL}\}$ using standard trace trick (of switching the expectation and the trace):

$$=-\frac{1}{2}\left[tr\{E_{q(s)}[(s-\mu_{KL})(s-\mu_{KL})^T]\sum_{KL}\}\right]-\frac{1}{2}\left[-\mu_{KL}{}^T\sum_{KL}\mu_{KL}-\mu_q^T\Sigma_q^{-1}\mu_q\right]$$
$$=-\frac{1}{2}\left[(\mu_q-\mu_{KL})^T\sum_{KL}(\mu_q-\mu_{KL})+tr(\kappa\Sigma_q)\right]-\frac{1}{2}\left[-\mu_{KL}{}^T\sum_{KL}\mu_{KL}-\mu_q^T\Sigma_q^{-1}\mu_q\right] \tag{A11}$$

We now add this back to the full equation (A6):

$$= \log\left[\frac{|\sum_l^{-1}|^{1/2}|\sum_p^{-1}|^{1/2}}{|\sum_q^{-1}|^{1/2}\,2\pi^{(m-n)/2}} \cdot \frac{\beta_l^{a_l}\beta_p^{a_p}}{\Gamma(a_l)\Gamma(a_p)} . {\lambda_p}^{a_p-1}\lambda_l^{a_l-1} \right] - \lambda_p\beta_p - \lambda_l\beta_l - \frac{1}{2}\lambda_l o^T \Lambda_l o$$
$$-\frac{1}{2}\left[(\mu_q - \mu_{KL})^T \sum_{KL} (\mu_q - \mu_{KL}) + tr(\kappa\Sigma_q) \right] - \frac{1}{2}\left[-{\mu_{KL}}^T \sum_{KL} \mu_{KL} - \mu_q^T \Sigma_q^{-1} \mu_q \right] \quad \text{(A12)}$$

From this we get the form of the variational free energy when the distributions are parameterised as gaussian-gamma:

$$-F = -\frac{1}{2}\left[{\mu_q}^T (\Sigma_{KL} - \Sigma_q^{-1})\mu_q - 2{\mu_q}^T \sum_{KL} \mu_{KL} + tr(\kappa\Sigma_q^{-1}) + o^T \sum_l^{-1} o + 2\lambda_p\beta_p + 2\lambda_l\beta_l \right]$$
$$+\frac{1}{2}\log\left[\frac{|\sum_l^{-1}||\sum_p^{-1}|}{|\sum_q^{-1}|\,2\pi^{(m-n)}} \right] + \log\frac{{\lambda_l}^{(a_l-1)}\beta_l^{a_l}{\lambda_p}^{(a_p-1)}\beta_p^{a_p}}{\Gamma(a_l)\Gamma(a_p)} \quad \text{(A13)}$$

A3.0 Renyi bound for Gaussian distribution

The probability density function for the random variables, s, o and x is the parameter governing the means:

$$p(s) = \frac{1}{2\pi^{n/2}|\Sigma_p|^{1/2}} \exp\left(-\frac{1}{2} s^T \Sigma_p^{-1} s \right)$$
$$p(o \mid s) = \frac{1}{2\pi^{n/2}|\Sigma_l|^{1/2}} \exp\left(-\frac{1}{2}(o - sx)^T \Sigma_l^{-1} (o - sx) \right)$$
$$\Rightarrow p(o, s) = p(o \mid s)\, p(s)$$

$$p(o, s) = \frac{1}{2\pi^{m/2}|\Sigma_l|^{1/2}|\Sigma_p|^{1/2}} \exp\left(-\frac{1}{2}\left((o - sx)^T \Sigma_l^{-1}(o - sx) + s^T \Sigma_p^{-1} s \right) \right)$$
$$\propto \exp\left(-\frac{1}{2}\left(o^T \Sigma_l^{-1} o - 2s^T x^T \Sigma_l^{-1} o + s^T \left[\Sigma_p^{-1} + x^T \Sigma_l^{-1} x \right] s \right) \right) \quad \text{(A13)}$$

Note, that $p(o, s)$ is a quadratic function of its components s, o, and a Gaussian distribution.

$$q(s) = \frac{1}{2\pi^{n/2}|\Sigma_q|^{1/2}} \exp\left(-\frac{1}{2}(s - \mu_q)^T \Sigma_q^{-1}(s - \mu_q) \right)$$

We now supplement these quantities into the negative Rényi bound, and rewrite using the defined quantities:

$$D_\alpha\left[q(s) \,||\, p(s, o) \right] = \frac{1}{1-\alpha} \log \int q(s)^\alpha \, p(s, o)^{1-\alpha} \, ds \quad \text{(A13)}$$

$$= \frac{1}{1-\alpha}\log\frac{1}{2\pi^{\alpha n/2}\,|\Sigma_q|^{\alpha/2}\;2\pi^{(1-\alpha)m/2}\,|\Sigma_l|^{1-\alpha/2}|\Sigma_p|^{1-\alpha/2}}\times$$

$$\int\exp-\frac{1}{2}\Big[a\Big[(s-\mu_q)^T\Sigma_q^{-1}(s-\mu_q)\Big]+(1-\alpha)\Big[\Big(o^T\Sigma_l^{-1}o-2s^Tx^T\Sigma_l^{-1}o+s^T\Big[\Sigma_p^{-1}+x^T\Sigma_l^{-1}x\Big]s\Big)\Big]\Big]ds$$

$$= \frac{1}{(1-\alpha)}\log\frac{\exp-0.5[(1-\alpha)o^T\Sigma_l^{-1}o]}{2\pi^{\alpha n/2}\,|\Sigma_q|^{\alpha/2}\;2\pi^{(1-\alpha)m/2}\,|\Sigma_l|^{1-\alpha/2}|\Sigma_p|^{1-\alpha/2}}\times$$

$$\int\exp-\frac{1}{2}\Big[(1-\alpha)s^T[\Sigma_p^{-1}+x^T\Sigma_l^{-1}x]s-2(1-\alpha)s^Tx^T\Sigma_l^{-1}o+\alpha s^T\Sigma_q^{-1}s+\alpha\mu_q{}^T\Sigma_q^{-1}\mu_q-2\alpha s^T\Sigma_q^{-1}\mu_q\Big]ds$$

First, let us focus on the term inside the integral. To avoid clutter we replace: $\Lambda_\alpha = \Sigma_p^{-1}+x^T\Sigma_l^{-1}x$

$$= \int\exp-\frac{1}{2}\Big[(1-\alpha)s^T\Lambda_\alpha s-2(1-\alpha)s^Tx^T\Sigma_l^{-1}o+\alpha s^T\Sigma_q^{-1}s+\alpha\mu_q{}^T\Sigma_q^{-1}\mu_q-2\alpha s^T\Sigma_q^{-1}\mu_q\Big]ds$$

$$= \int\exp-\frac{1}{2}\Big[s^T\Big((1-\alpha)\Lambda_\alpha+\alpha\Sigma_q^{-1}\Big)s-2s^T\Big(\alpha\Sigma_q^{-1}\mu_q+(1-\alpha)x^T\Sigma_l^{-1}o\Big)+\alpha\mu_q{}^T\Sigma_q^{-1}\mu_q\Big]ds$$

$$= \int\exp-\frac{1}{2}\Big[(s-\mu_\alpha)^T\Sigma_\alpha{}^{-1}(s-\mu_\alpha)-\mu_\alpha{}^T\Sigma_\alpha{}^{-1}\mu_\alpha+\alpha\mu_q{}^T\Sigma_q^{-1}\mu_q\Big]ds$$

$$= \exp-\frac{1}{2}\Big[-\mu_\alpha{}^T\Sigma_\alpha{}^{-1}\mu_\alpha+\alpha\mu_q{}^T\Sigma_q^{-1}\mu_q\Big]\int\exp-\frac{1}{2}\Big[(s-\mu_\alpha)^T\Sigma_\alpha{}^{-1}(s-\mu_\alpha)\Big]ds$$

$$= 2\pi^{\frac{n}{2}}\,|\Sigma|^{\frac{1}{2}}\exp-\frac{1}{2}\Big[-\mu_\alpha{}^T\Sigma_\alpha{}^{-1}\mu_\alpha+\alpha\mu_q{}^T\Sigma_q^{-1}\mu_q\Big]$$

where:
$$\Sigma_\alpha = \Big((1-\alpha)\Lambda_\alpha+\alpha\Sigma_q^{-1}\Big)^{-1}$$
$$\mu_\alpha = \Sigma_\alpha\Big(\alpha\Sigma_q^{-1}\mu_q+(1-\alpha)x^T\Sigma_l^{-1}o\Big)$$

We put it all together now:

$$D_\alpha\big[q(s)\,\|\,p(s,o)\big] = \frac{1}{(1-\alpha)}\log\left[\begin{array}{l}\dfrac{|\Sigma_\alpha|^{\frac{1}{2}}}{2\pi^{(\alpha-1)n/2}\,2\pi^{(1-\alpha)m/2}\,|\Sigma_l|^{1-\alpha/2}|\Sigma_p|^{1-\alpha/2}|\Sigma_q|^{\alpha/2}}\\ \times\exp[-\dfrac{1}{2}\Big[(1-\alpha)o^T\Sigma_l^{-1}o-\mu_\alpha{}^T\Sigma_\alpha{}^{-1}\mu_\alpha+\alpha\mu_q{}^T\Sigma_q^{-1}\mu_q\Big]\end{array}\right]$$

$$= \frac{1}{2(1-\alpha)}\Big(\mu_\alpha{}^T\Sigma_\alpha{}^{-1}\mu_\alpha-\alpha\mu_q{}^T\Sigma_q^{-1}\mu_q+(\alpha-1)o^T\Sigma_l^{-1}o\Big) \qquad \text{(A14)}$$

$$+\frac{1}{2(1-\alpha)}\log\left[\frac{|\Sigma_\alpha|}{2\pi^{(1-\alpha)(m-n)}\,|\Sigma_l|^{1-\alpha}|\Sigma_p|^{1-\alpha}|\Sigma_q|^{\alpha}}\right]$$

A4.0 Rényi bounds for a univariate gaussian system

Here, we consider a univariate gaussian system where changes to the α values do not impact behaviour. This holds true since the true (gaussian) posterior is in the same family of distributions as the assumed variational (gaussian) posterior. The system is defined as:

$$
\begin{aligned}
p(s) &\sim N\left(s; 0, 1.0\right) \\
p(o,s) &\sim N\left(o; sx, 1.0\right) p\left(s\right) \\
q(s) &\sim N\left(s; \mu_q, \Sigma_q\right)
\end{aligned}
\tag{A15}
$$

Using this system, we estimate the variational posterior, with four different Rényi bounds parameterised by the α values: $-\infty, -1, 0, 1$ (Table 1), using gradient descent. Appendix 4.0 presents the pseudocode for the optimisation procedure. Everything is kept consistent across each simulation i.e., had it not been for the change in the α value they would be no difference between the simulations. The results from this are shown in Figure 3 – as expected each estimate converged to the analytical posterior of 0.40 mean, μ_q and 0.0045 variance, Σ_q, after an appropriate number of gradient descent iterations and learning rate of 0.1. This simulation highlights that for certain parameterisations, changes to the Rényi variational objective, have no impact in modelling behavioural variations. This is because exact inference is possible: the true posterior is in the same family of distributions as the assumed variational posterior. Therefore, altering the α value has little to no effect on the variational posterior estimation.

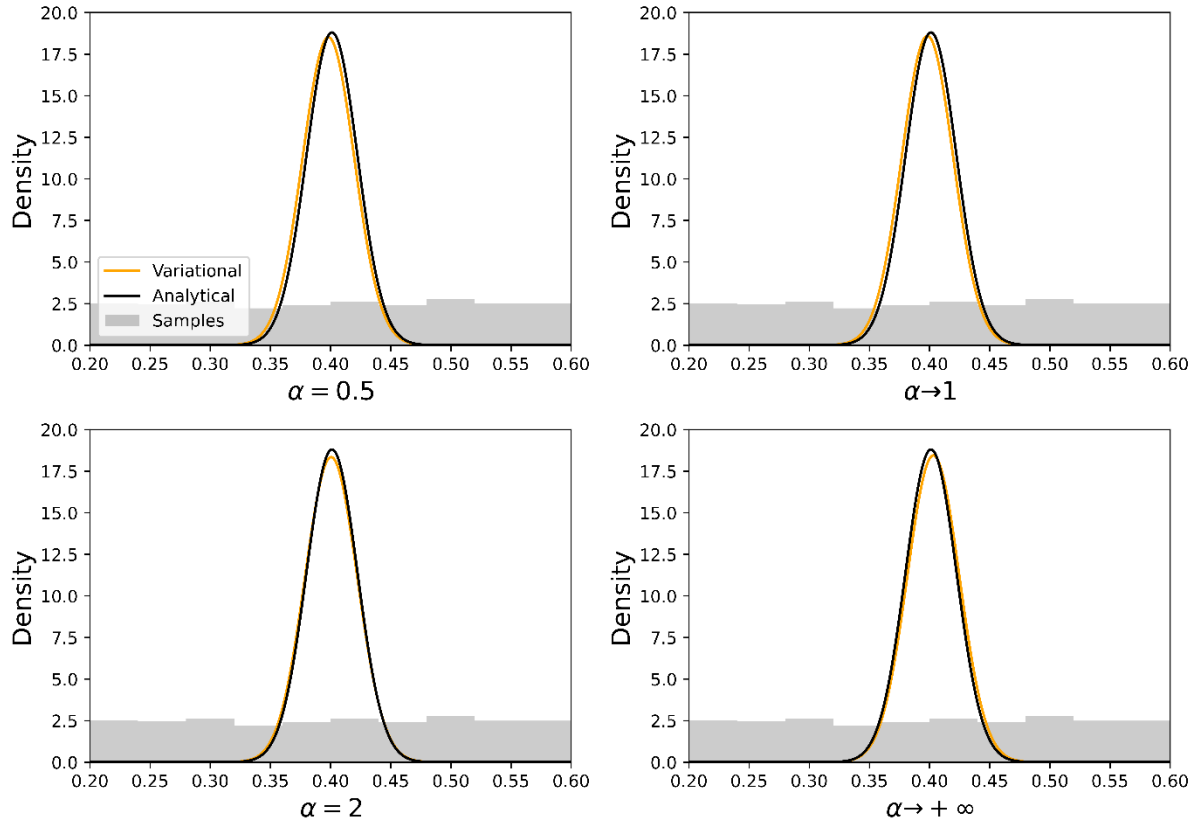

Figure A1. (*Estimated variational posteriors for a univariate gaussian system*) The figures plot the variational posteriors, under the different variational objectives, parameterised by the α. We plot results for α values: $0.5, 1, 2, \infty+$. Each plot shows the $q(s)$ value on the x-axis, and its density on the y-axis. The black line is the analytical posterior, the orange line the variational posterior and grey bars represent the samples. The samples are from a uniform distribution between 0.2 and 0.6. For each bound, we see a convergence to the analytical

posterior of 0.40 mean, μ and 0.0045 variance, σ after an appropriate learning rate, and 2048 gradient descent.

A5.0 Pseudo-code for Rényi bound optimisation

Input: a model $p(s,o)$ **,** observations

Output: a variational density $q(s)$

Initialise: variational density sufficient statistics e.g., $N(0,I)$ for gaussian

while the Rényi bound has not converged **do**

compute expectation using samples

compute gradient

optimise with gradient descent

end

return $q(s)$

A6.0 Pseudo-code for MAB planning as inference algorithm

Initialise: empty array for arm selected, observations and score

while t<T **do**

for arm in *range*(bandit-arms)

samples = observation[bandit-arms == arm]

perform inference using A2.0 and samples, return $q(s)$

end

softmax over posterior estimates for each arm

argmax arm

update arm selected array

calculate score and regret

update observation and score array

end

References

Amari S-i. 1985. Differential-geometrical methods in statistics. *Lecture Notes on Statistics* 28: 1

Amari S-i. 2012. *Differential-geometrical methods in statistics*. 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: 183-95

Amari S. 2009. α -Divergence Is Unique, Belonging to Both f-Divergence and Bregman Divergence Classes. *IEEE Transactions on Information Theory* 55: 4925-31

Ambrogioni L, Güçlü U, Güçlütürk Y, Hinne M, Maris E, van Gerven MA. 2018. Wasserstein variational inference. *arXiv preprint arXiv:1805.11284*

Angela JY, Dayan P. 2002. Acetylcholine in cortical inference. *Neural Networks* 15: 719-30

Angela JY, Dayan P. 2005. Uncertainty, neuromodulation, and attention. *Neuron* 46: 681-92

Ay N, Gibilisco P. 2016. *Information Geometry and Its Applications*. Springer.

Barber D, de van Laar P. 1999. Variational cumulant expansions for intractable distributions. *Journal of Artificial Intelligence Research* 10: 435-55

Beal MJ. 2003. Variational Algorithms for Approximate Bayesian Inference. *PhD. Thesis, University College London*

Birgé L. 1986. On estimating a density using Hellinger distance and some other strange facts. *Probability theory and related fields* 71: 271-91

Blei DM, Kucukelbir A, McAuliffe JD. 2017. Variational inference: A review for statisticians. *Journal of the American Statistical Association* 112: 859-77

Bogacz R. 2017. A tutorial on the free-energy framework for modelling perception and learning. *Journal of Mathematical Psychology* 76: 198--211

Botvinick M, Toussaint M. 2012. Planning as inference. *Trends Cogn Sci.* 16: 485-8

Burbea J. 1984. The convexity with respect to Gaussian distributions of divergences of order α. *Utilitas Mathematica* 26: 171-92

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

Dayan P, Hinton GE, Neal R. 1995. The Helmholtz machine. *Neural computation* 7: 889-904

Dieng AB, Tran D, Ranganath R, Paisley J, Blei DM. 2016. Variational Inference via $\chi $-Upper Bound Minimization. *arXiv preprint arXiv:1611.00328*

Doya K, Ishii S, Pouget A, Rao RP. 2007. *Bayesian brain: Probabilistic approaches to neural coding*. MIT press.

FitzGerald TH, Schwartenbeck P, Moutoussis M, Dolan RJ, Friston K. 2015. Active inference, evidence accumulation, and the urn task. *Neural computation* 27: 306-28

Fountas Z, Sajid N, Mediano PA, Friston K. 2020. Deep active inference agents using Monte-Carlo methods. *arXiv preprint arXiv:2006.04176*

Friston K. 2019. A free energy principle for a particular physics. *arXiv preprint arXiv:1906.10184*

Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. 2017. Active Inference: A Process Theory. *Neural computation* 29: 1-49

Friston K, Samothrakis S, Montague R. 2012. Active inference and agency: optimal control without cost functions. *Biol. Cybernetics*: [Epub ahead of print]

Friston K, Schwartenbeck P, FitzGerald T, Moutoussis M, Behrens T, Dolan RJ. 2014a. The anatomy of choice: dopamine and decision-making. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 369

Friston KJ. 2010. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience* 11: 127--38

Friston KJ, Rigoli F, Ognibene D, Mathys C, Fitzgerald T, Pezzulo G. 2015. Active inference and epistemic value. *Cognitive neuroscience* 6: 187--224

Friston KJ, Stephan KE, Montague R, Dolan RJ. 2014b. Computational psychiatry: the brain as a phantastic organ. *The Lancet Psychiatry* 1: 148-58

Gibbs AL, Su FE. 2002. On choosing and bounding probability metrics. *International statistical review* 70: 419-35

Grünwald PD. 2007. *The minimum description length principle*. MIT press.

Hinton GE, Zemel RS. 1994. Autoencoders, minimum description length, and Helmholtz free energy. *Advances in neural information processing systems* 6: 3-10

Hohwy J. 2012. Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology* 3: 1--14

Huzurbazar VS. 1955. Exact Forms of Some Invariants for Distributions Admitting Sufficient Statistics. *Biometrika* 42: 533-37

Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK. 1999. An introduction to variational methods for graphical models. *Machine learning* 37: 183-233

Knill DC, Pouget A. 2004. The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences* 27: 712-19

Li Y, Turner RE. *Advances in Neural Information Processing Systems2016*: 1073-81.

Metelli AM, Papini M, Faccio F, Restelli M. 2018. Policy Optimization via Importance Sampling. *Advances in Neural Information Processing Systems* 31: 5442-54

Millidge B, Tschantz A, Buckley CL. 2020. Predictive coding approximates backprop along arbitrary computation graphs. *arXiv preprint arXiv:2006.04182*

Minka T. 2005. Divergence measures and message passing, Microsoft Research

Nielsen F, Nock R. 2011. On R\'enyi and Tsallis entropies and divergences for exponential families. *arXiv preprint arXiv:1105.3259*

Parisi G. 1988. *Statistical Field Theory*. Basic Books.

Parr T, Friston KJ. 2017. Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface* 14: 20170376

Parr T, Friston KJ. 2019. Generalised free energy and active inference. *Biological Cybernetics* 113: 495-513

Parr T, Markovic D, Kiebel SJ, Friston KJ. 2019. Neuronal message passing using Mean-field, Bethe, and Marginal approximations. *Scientific Reports* 9: 1889

Penny W. 2012. Bayesian Models of Brain and Behaviour. *ISRN Biomathematics* 2012: 785791

Regli J-B, Silva R. 2018. Alpha-beta divergence for variational inference. *arXiv preprint arXiv:1805.01045*

Rényi A. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics1961*. The Regents of the University of California.

Sajid N, Ball PJ, Parr T, Friston KJ. 2021. Active inference: demystified and compared. *Neural computation*: 1-39

Sajid N, Parr T, Gajardo-Vidal A, Price CJ, Friston KJ. 2020a. Paradoxical lesions, plasticity and active inference. *Brain Communications*

Sajid N, Parr T, Hope TM, Price CJ, Friston KJ. 2020b. Degeneracy and Redundancy in Active Inference. *Cerebral Cortex*

Schmidhuber J. 1991. Curious model-building control systems. *In Proc. International Joint Conference on Neural Networks, Singapore. IEEE* 2: 1458–63

Schmidhuber J. 2006. Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connect Sci* 18: 173-87

Schwartenbeck P, FitzGerald TH, Mathys C, Dolan R, Wurst F, et al. 2015. Optimal inference with suboptimal models: addiction and active Bayesian inference. *Medical hypotheses* 84: 109-17

Schwartenbeck P, Friston K. 2016. Computational Phenotyping in Psychiatry: A Worked Example. *eNeuro* 3: ENEURO.0049-16.2016

Schwöbel S, Kiebel S, Marković D. 2018. Active Inference, Belief Propagation, and the Bethe Approximation. *Neural computation*: 1-38

Smith R, Lane RD, Parr T, Friston KJ. 2019. Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their potential clinical relevance. *Neuroscience & Biobehavioral Reviews* 107: 473-91

Spratling MW. 2017. A review of predictive coding algorithms. *Brain and Cognition* 112: 92--97

Sun Y, Gomez F, Schmidhuber J. 2011. Planning to Be Surprised: Optimal Bayesian Exploration in Dynamic Environments In *Artificial General Intelligence: 4th International Conference, AGI 2011, Mountain View, CA, USA, August 3-6, 2011. Proceedings*, ed. J Schmidhuber, KR Thórisson, M Looks, pp. 41-51. Berlin, Heidelberg: Springer Berlin Heidelberg

Sutton RS, Barto AG. 1998. *Introduction to Reinforcement Learning*. MIT Press. 342 pp.

Tschantz A, Seth AK, Buckley CL. 2020. Learning action-oriented models through active inference. *PLoS computational biology* 16: e1007805

van de Laar T, Senoz I, Özçelikkale A, Wymeersch H. 2021. Chance-Constrained Active Inference. *arXiv preprint arXiv:2102.08792*

Van Erven T, Harremos P. 2014. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory* 60: 3797-820

Wainwright MJ, Jordan MI. 2008. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning* 1: 1-305

Wan N, Li D, Hovakimyan N. 2020. f-Divergence Variational Inference. *Advances in Neural Information Processing Systems* 33

Welbourne SR, Woollams AM, Crisp J, Lambon-Ralph MA. 2011. The role of plasticity-related functional reorganization in the explanation of central dyslexias. *Cognitive Neuropsychology*

Whittington JCR, Bogacz R. 2017. An Approximation of the Error Backpropagation Algorithm in a Predictive Coding Network with Local Hebbian Synaptic Plasticity. *Neural computation* 29: 1229-62

Zhang M, Bird T, Habib R, Xu T, Barber D. 2019. Variational f-divergence minimization. *arXiv preprint arXiv:1907.11891*