

# Bayesian Brains and the Rényi divergence

Noor Sajid<sup>1, †</sup>, Francesco Faccio<sup>2, †</sup>, Lancelot Da Costa<sup>1, 3</sup>, Thomas Parr<sup>1</sup>, Jürgen Schmidhuber<sup>2</sup> & Karl Friston<sup>1</sup>

<sup>1</sup>Wellcome Centre for Human Neuroimaging, University College London, London, UK., <sup>2</sup>The Swiss AI Lab IDSIA, USI, SUPSI, Lugano, Switzerland.

<sup>3</sup>Department of Mathematics, Imperial College London, London, UK., <sup>†</sup>Equal Contribution,

Correspondence: noor.sajid.18@ucl.ac.uk & francesco@idsia.ch



## Motivation

The notion that the brain is Bayesian and performs some form of inference has attracted enormous attention in neuroscience. This casts the brain as realising a model about causes of sensation, that can predict observations and future behaviour.

Practically, this involves the optimisation of the variational free energy (derived from the KL-divergence). Then **behavioural variations are attributed to altered priors** over the generative model parameters. We offer an **alternative account for explaining behavioural variations using the Rényi divergences**.

## Rényi Divergence

Rényi divergences are a general class of divergences, indexed by an  $\alpha$  parameter, of which the KL-divergence is a special case:

$$D_\alpha [p(s|o)||q(s)] = \frac{1}{\alpha - 1} \log \int p(s|o)^\alpha q(s)^{1-\alpha} ds$$

This divergence family can provide different posterior estimates, at the minimum of the divergence with respect to  $q$ , that vary smoothly with  $\alpha$ . These differences are possible only when the true posterior, e.g., some bi-modal distribution, is not in the same family of distributions as the approximate posterior, e.g. a Gaussian distribution. The corresponding variational Rényi bound is:

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

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

## Properties of Rényi bound

Under the Rényi bound optimising  $\alpha \rightarrow 0^+$ , the approximate posterior covers the joint distribution. This happens because the approximate posterior is forced to be  $q(\cdot) > 0$ , whenever the joint is  $p(\cdot) > 0$  (i.e., zero-avoiding). This implies that low probability regions of the joint distribution may be overestimated.

Conversely,  $\alpha \rightarrow +\infty$  will favour posterior distributions that best fit the mode with the most mass (not necessarily the highest). This happens because the approximate posterior is forced to be small, whenever the joint is small (i.e., zero-forcing). This causes parts of the joint distribution to be excluded. Figure 2 illustrates this.

## Multi-armed bandit simulations

To illustrate differential preferences that arise under the Rényi bound we simulated the multi-armed bandit paradigm. This was formulated as a one-state Markov Decision Process where the agent could pull an arm and observe corresponding outcome (i.e., score) each time-step. The agent's objective was to identify, and select, the arm with the highest Sharpe ratio. The Sharpe ratio, a variance adjusted return measure, is defined as  $SR := \frac{R_t}{\sqrt{R_t}}$ .

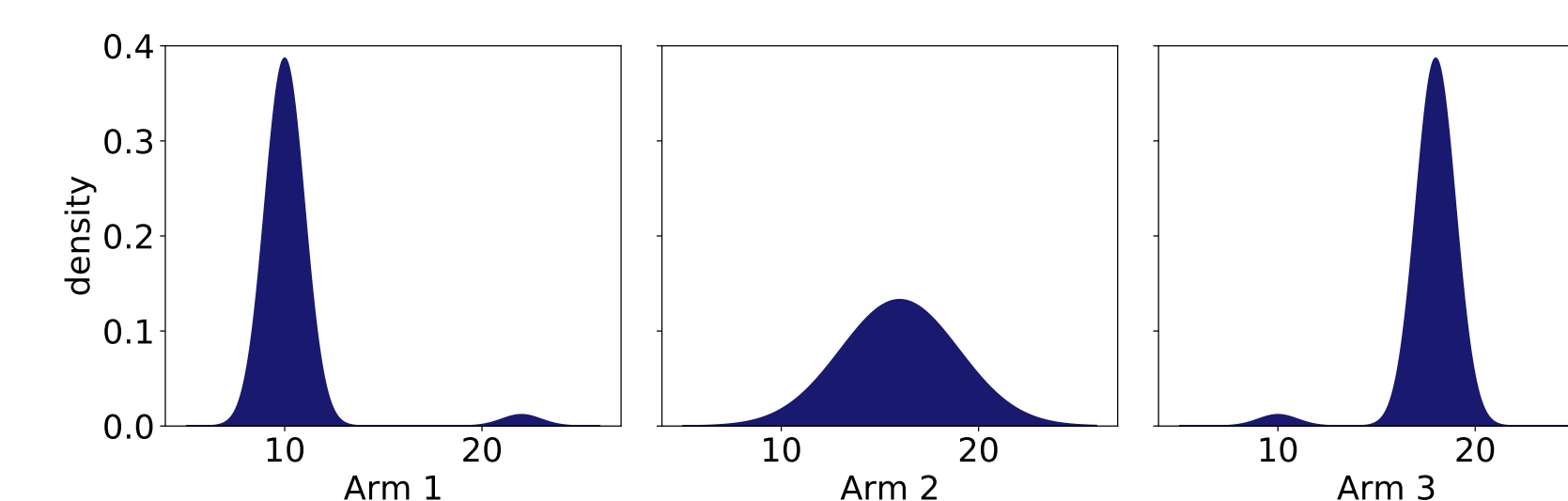


Figure 1: Score distribution for each arm

Each arm has a fixed distribution unknown to the agent: arm 1 is a bi-modal distribution with a Sharpe ratio of  $SR = 2.03$ , arm 2 is a Gaussian distribution with a Sharpe ratio of  $SR = 1.76$ , and arm 3 is another bi-modal distribution with an expected return of  $SR = 6.20$ . Thus, arm 3 was the best choice i.e., the arm with the maximal Sharpe ratio. 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 pulled arm at iteration  $t$ .

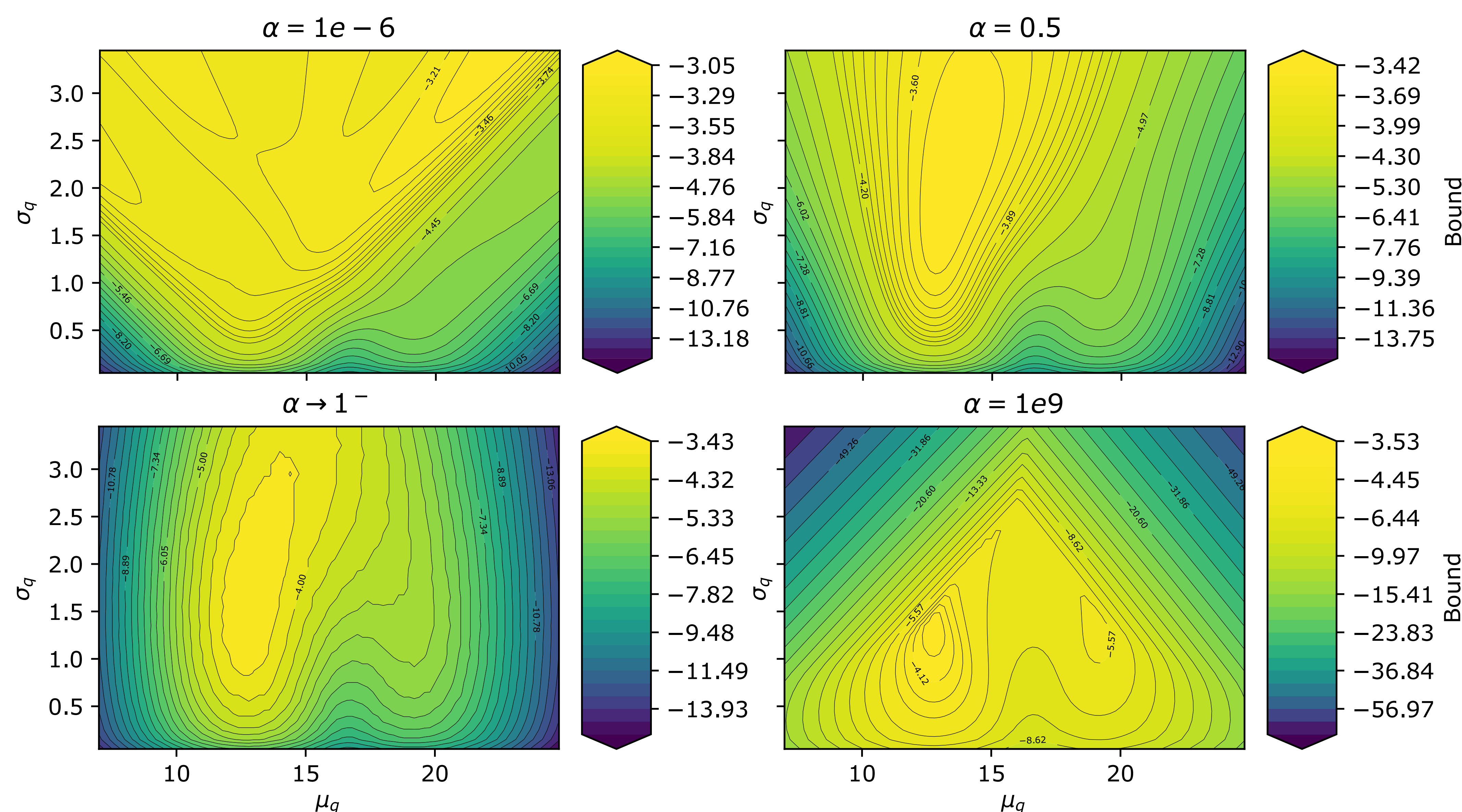


Figure 2: The Rényi bound as a function of the variational posterior for arm 1.

We simulated 4000 iterations of the agents optimising the Rényi bound with  $\alpha$  values:  $\rightarrow 0^+$ ,  $0.5$ ,  $\rightarrow 1^-$ ,  $2$ ,  $10$ ,  $\rightarrow +\infty$ , during this task. Figure 3 shows the behavioural variations, measured using cumulative regret.

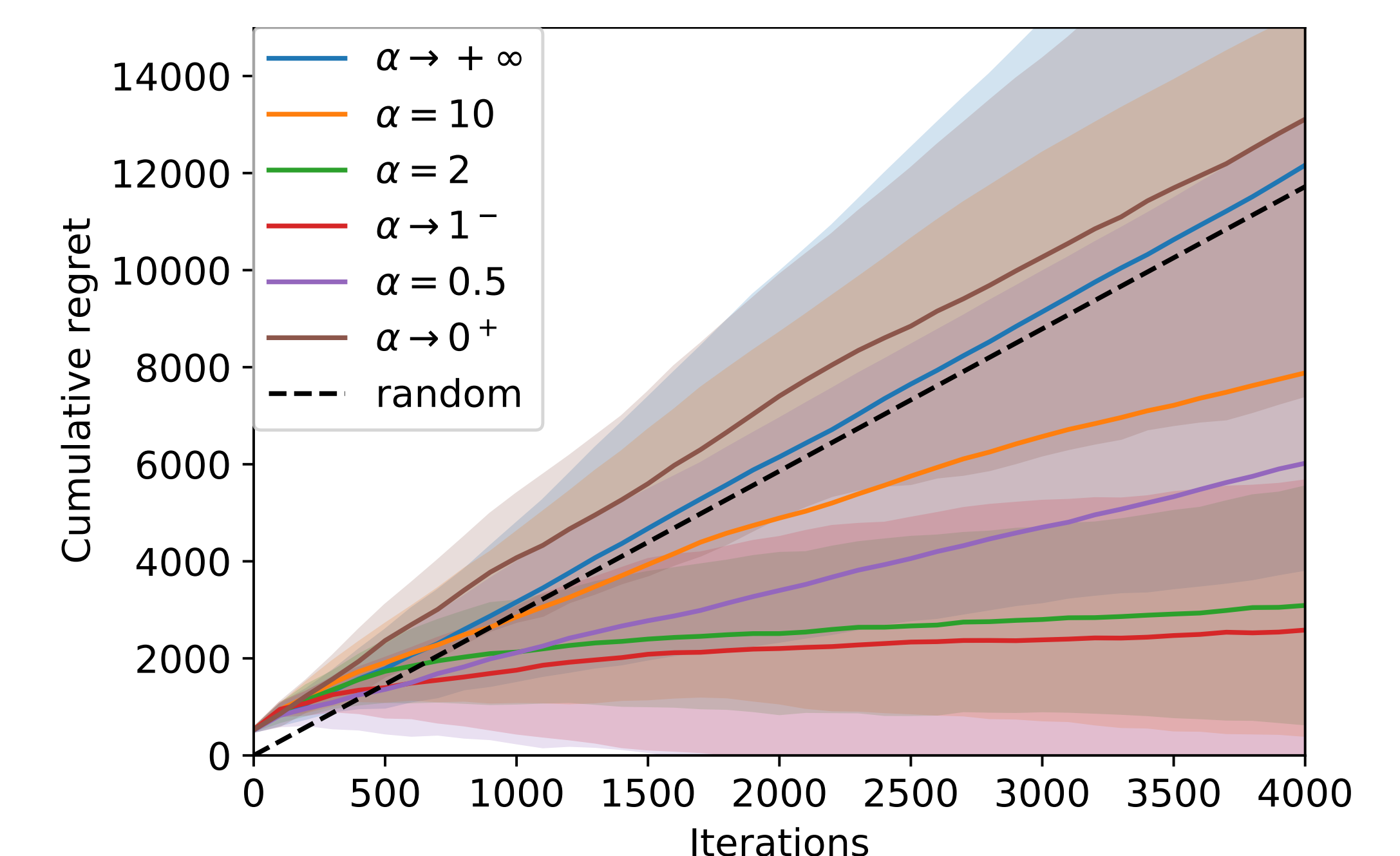


Figure 3: Regret under the Rényi bound.

Optimising  $\alpha \rightarrow 1^-$  leads to the lowest cumulative regret, &  $\alpha \rightarrow \infty$ ;  $\rightarrow 0^+$  with the highest. Next, we plot the variational posterior for arm 1 where approximating the second mode leads to sub-optimal action selection. We see different action selection strategies for agents optimising  $\alpha \rightarrow \infty$ ;  $\rightarrow 0^+$  that are consequence of bound properties. For agents optimising  $\alpha \rightarrow \infty$ , the posterior gets stuck in the second mode due to the low variance leading to the sub-optimal arm selection. For  $\alpha \rightarrow 0^+$  optimising agents mass-covering behaviour leads to high probability that a sub-optimal arm is selected.

## Conclusions

We show that 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  $\alpha$  parameter that alter the bound strength, induce different inferences and consequent behavioural variations.

[1] Doya, et al (2007). Bayesian brain: Probabilistic approaches to neural coding. MIT press. [2] Rényi (1961). On measures of entropy and information. & [3] Li & Turner (2017). Rényi divergence variational inference.