An information-geometric approach to artificial curiosity

Alexander Nedergaard, Institute of Neuroinformatics, ETH Zurich

Artificial curiosity

$$r(s) + \beta \bar{r}(s), \quad \beta \in \mathbb{R}_{\geq 0}$$

Artificial curiosity

$$r(s) + \beta \bar{r}(s), \quad \beta \in \mathbb{R}_{\geq 0}$$

Optimal policy

$$\pi_\beta = \arg\max_\pi \int_{\mathcal{S}} \cdots \int_{\mathcal{S}} \left[\sum_{i=0}^{n} r(s_i) + \beta \bar{r}(s) \right] d\mu(s_0) \prod_{i=1}^{n} dM(s_i \,|\, s_{i-1})$$

Artificial curiosity

$$r(s) + \beta \bar{r}(s), \quad \beta \in \mathbb{R}_{\geq 0}$$

Optimal policy

$$\pi_\beta = \arg\max_\pi \int_{\mathcal{S}} \cdots \int_{\mathcal{S}} \left[\sum_{i=0}^{n} r(s_i) + \beta \bar{r}(s) \right] d\mu(s_0) \prod_{i=1}^{n} dM(s_i \,|\, s_{i-1})$$

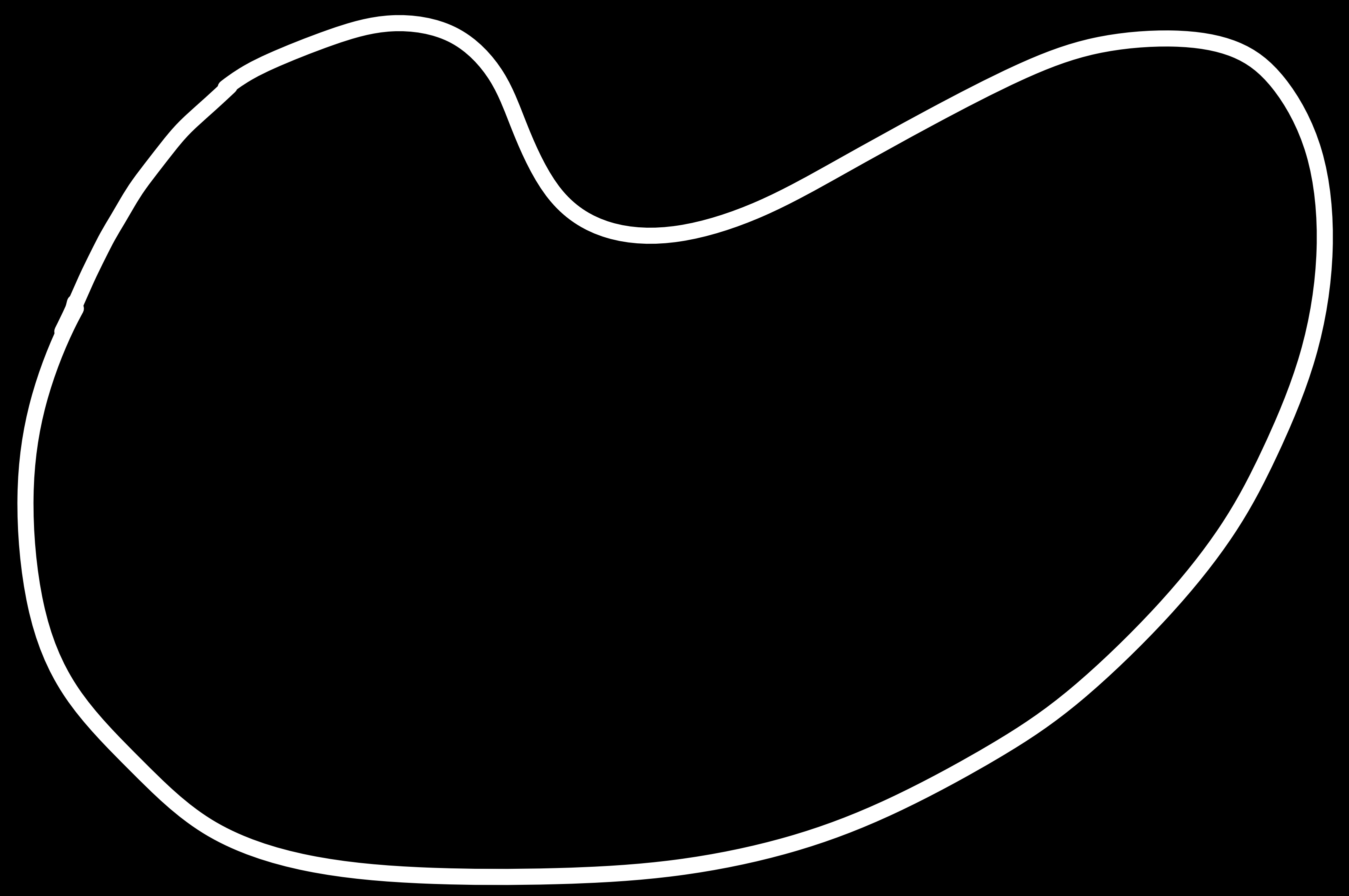
Occupancy (state visitation distribution)

$$p_\pi = M p_\pi$$

Exploration-exploitation trade-off via information geodesics

$$(\beta, p_{\pi_\beta})$$

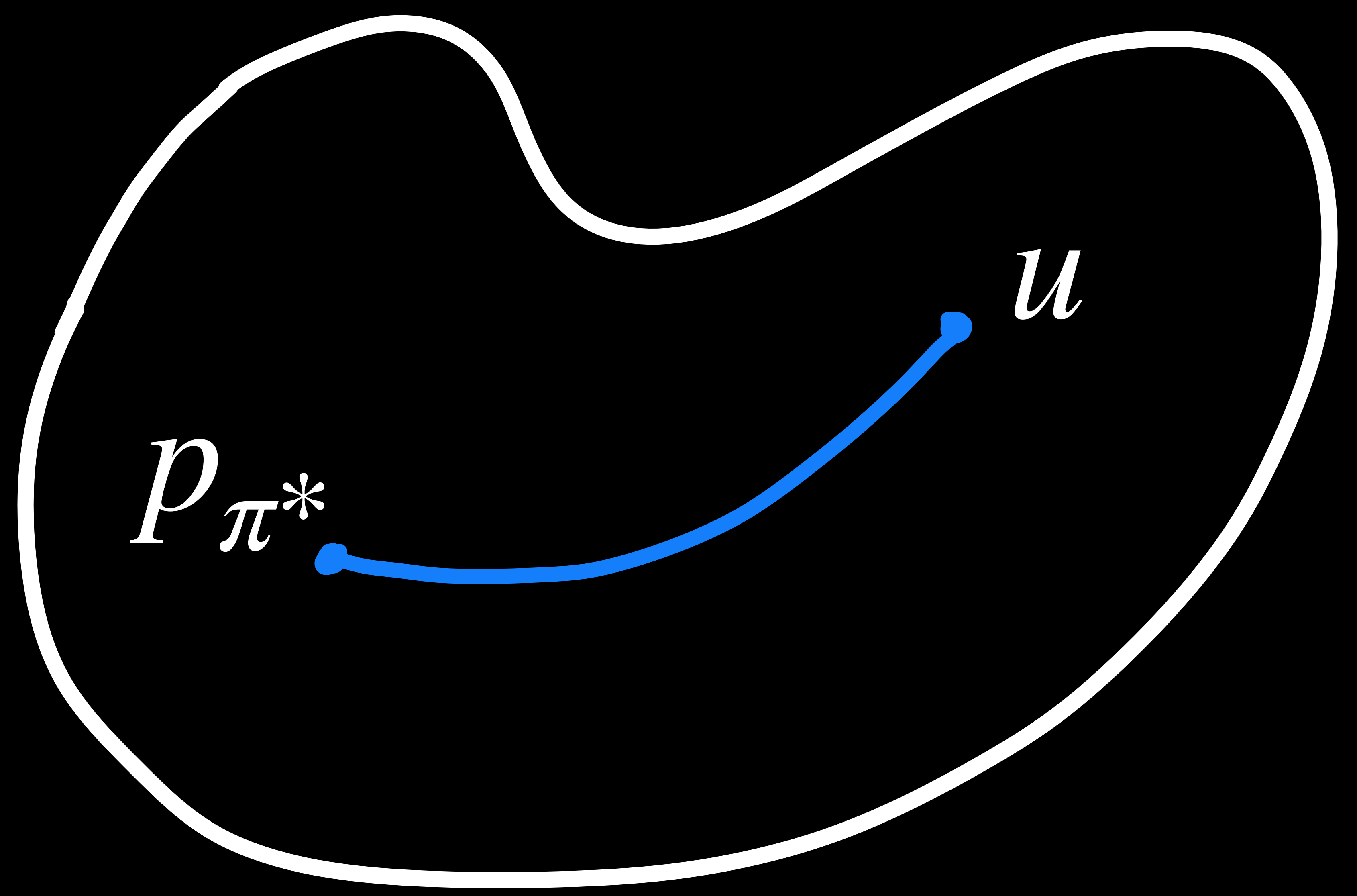
Occupancy manifold



Exploration-exploitation trade-off via information geodesics

$$(\beta, p_{\pi_\beta})$$

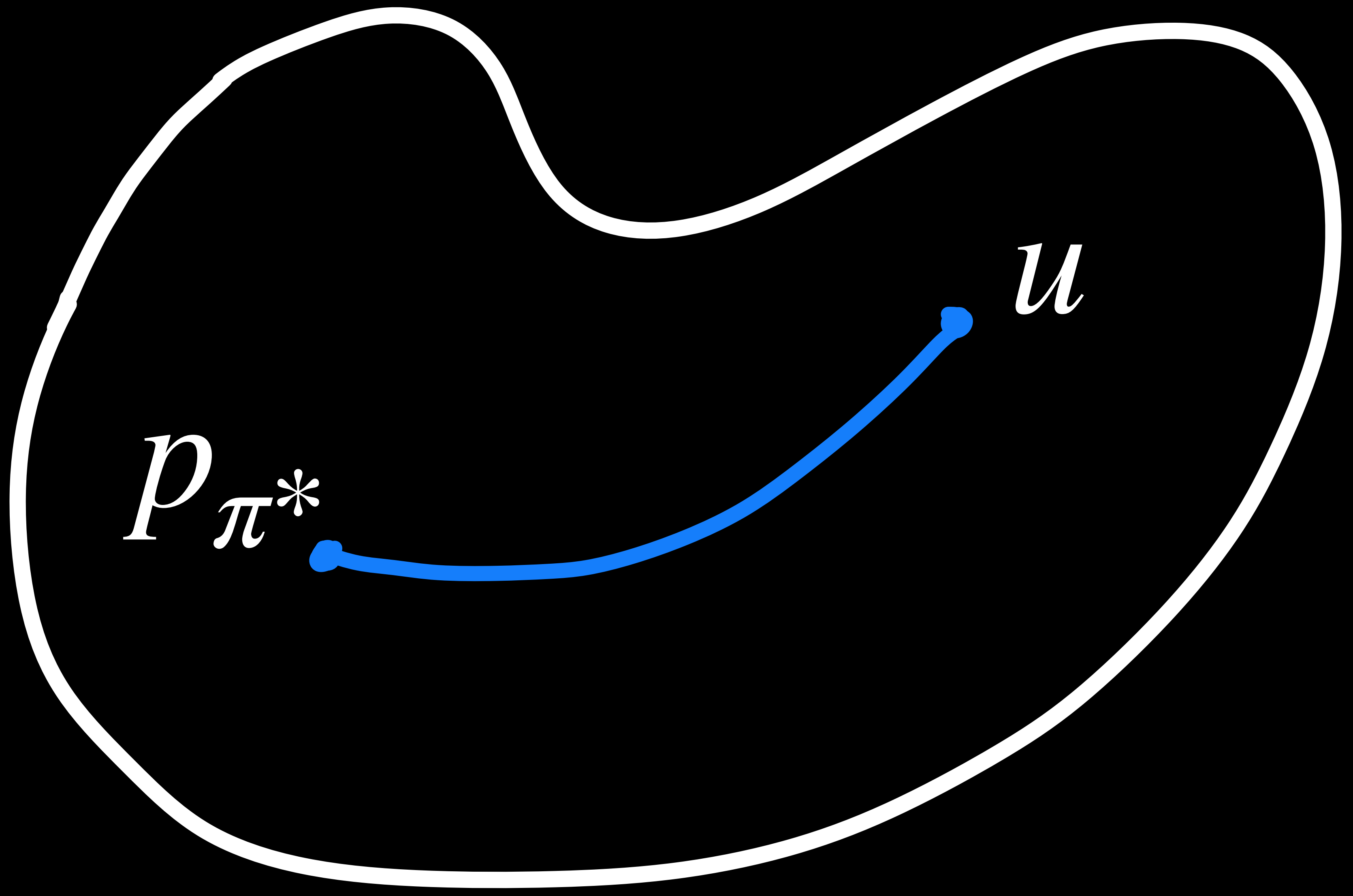
Occupancy manifold



Exploration-exploitation trade-off via information geodesics

$$(\beta, p_{\pi_\beta})$$

Occupancy manifold



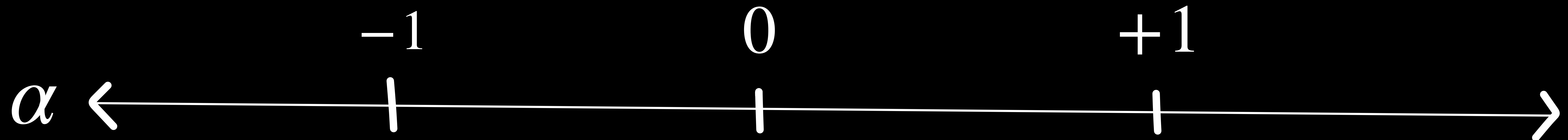
Information rewards

$$\bar{r}(s) = \frac{4}{1-\alpha^2}\left(\left[\frac{1}{p_\pi(s)}\right]^{\frac{\alpha+1}{2}} - 1\right), \quad \alpha \in \mathbb{R}$$

$\beta \sim$ "information exploited by the optimal agent to maximise rewards"

Information rewards unify exploration

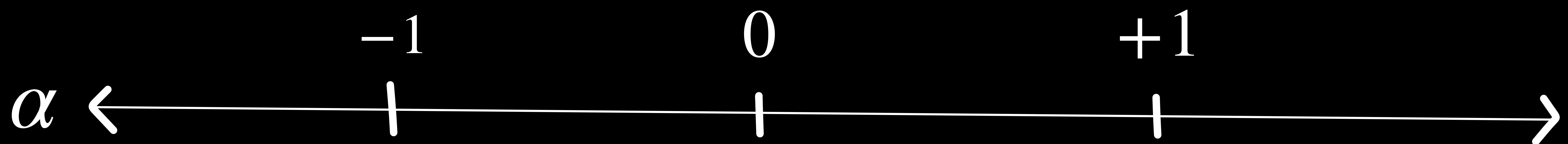
$$r(s) + \frac{4\beta}{1-\alpha^2}\left(\left[\frac{1}{p_\pi(s)}\right]^{\frac{\alpha+1}{2}} - 1\right), \quad \alpha \in \mathbb{R}, \beta \in \mathbb{R}_{\geq 0}$$



$\alpha \sim$ "curvature of the occupancy manifold"

Information rewards unify exploration

$$r(s) + \frac{4\beta}{1-\alpha^2}\left(\left[\frac{1}{p_\pi(s)}\right]^{\frac{\alpha+1}{2}} - 1\right), \quad \alpha \in \mathbb{R}, \beta \in \mathbb{R}_{\geq 0}$$



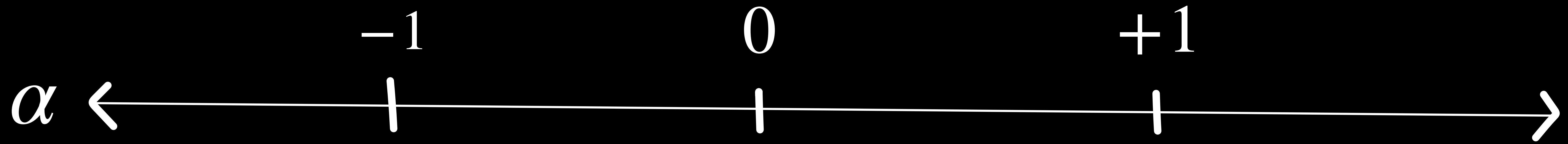
Count-based exploration

$$r(s) + \frac{\beta}{\sqrt{n(s)}}$$

$\alpha \sim$ “curvature of the occupancy manifold”

Information rewards unify exploration

$$r(s) + \frac{4\beta}{1-\alpha^2}\left(\left[\frac{1}{p_\pi(s)}\right]^{\frac{\alpha+1}{2}} - 1\right), \quad \alpha \in \mathbb{R}, \beta \in \mathbb{R}_{\geq 0}$$



Max entropy exploration

$$R(\pi) + \beta H(p_\pi)$$

Count-based exploration

$$r(s) + \frac{\beta}{\sqrt{n(s)}}$$

$\alpha \sim$ "curvature of the occupancy manifold"

Acknowledgements

Pablo Morales

Shun-ichi Amari