Brain Bandit: A Biologically Grounded Neural Network for Efficient Control of Exploration

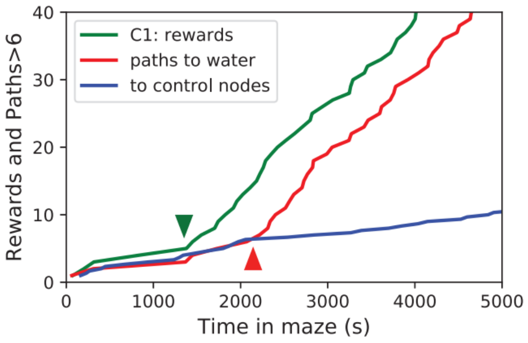
Chen Jiang[1], Jiahui An[1,2], Yating Liu[3,1], Ni Ji[1,2]

Chinese Institute for Brain Research, Beijing
Chinese Academy of Medical Sciences
China Agricultural University

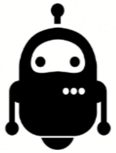
ICLR 2025, Singapore

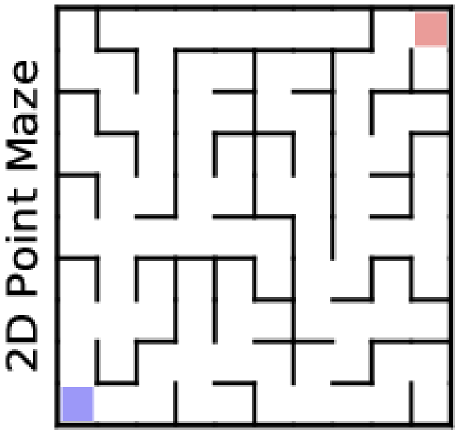
How do humans and animals learn fast & with few samples?

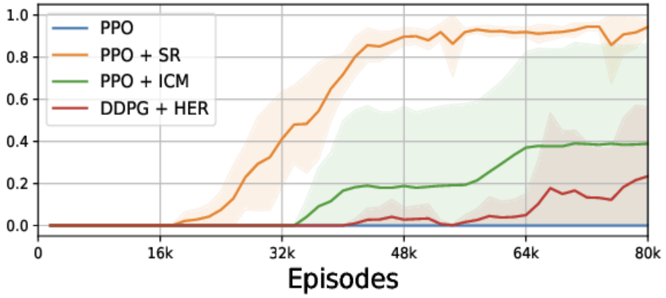
Animals



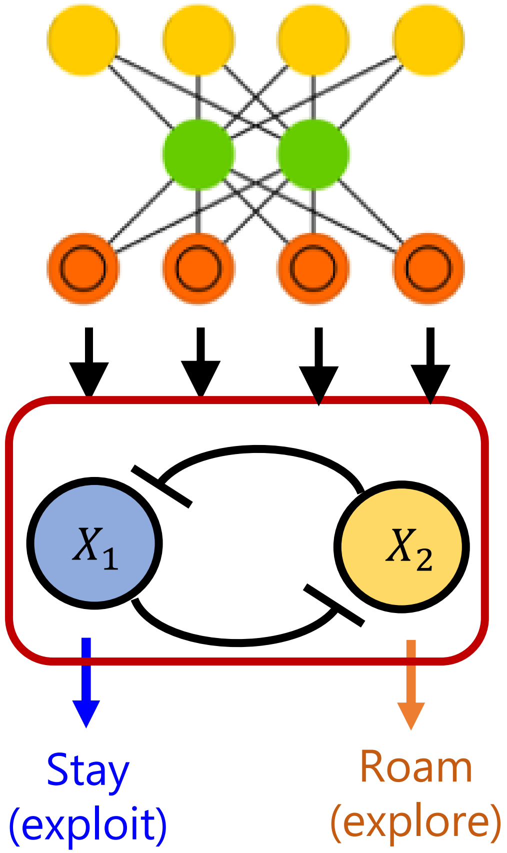
Agents







Biological neural networks



Context encoding

Decision-making

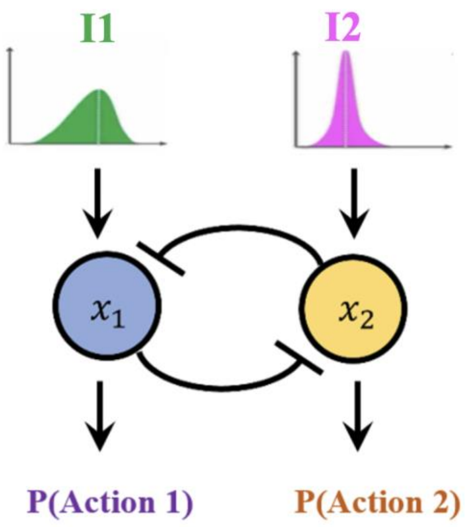
Explore-exploit

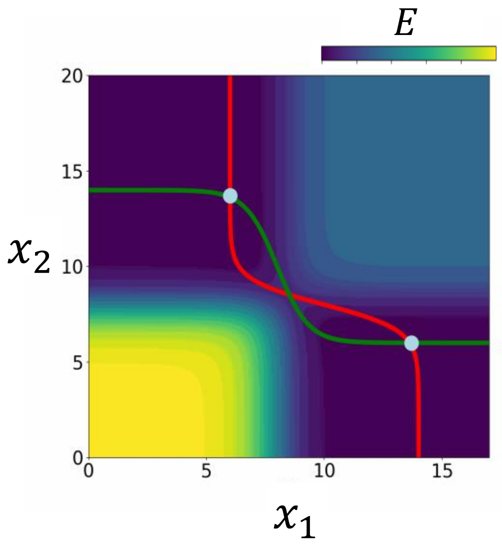
Ji et al. *eLife* 2021b

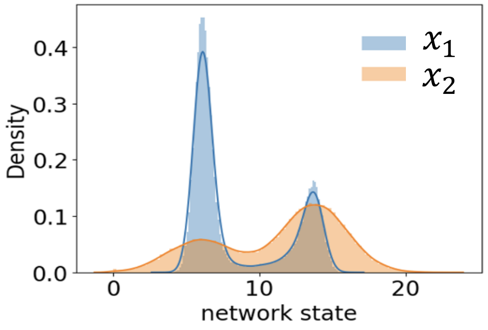
See also: X-J Wang 2008
Heric and Sakata, 2019
Mei et al., 2023

A bio-grounded neural network for explore-exploit balance

Brain Bandit Network (BBN)







Continuous Hopfield network (with stochastic input)

$$\tau_i \frac{dx_i}{dt} = \underbrace{-\gamma_i x_i}_{\text{self-decay}} + \underbrace{\sum_{j\neq i}^{N} w_{ij} f(x_j)}_{\text{mutual inhibition}} + \underbrace{b_i}_{\text{activity baseline}} + \underbrace{\bar{I}_i}_{\text{mean external input}} + \underbrace{\sigma_i dW(t)}_{\text{Wiener noise (uncertainty in input)}}$$

Hopfield energy

$$E = \left\{-\frac{1}{2}\sum_{i,j,i\neq j}^{N} w_{ij} f(x_i) f(x_j) + \sum_{i}^{N}\left[x_i f(x_i) - \int_0^{x_i} f(x)dx\right] - \sum_{i}^{N} b_i f(x_i)\right\} - \left\{\sum_{i}^{N} \bar{I}_i f(x_i)\right\}$$

State probability distribution (Boltzmann)

$$P(\boldsymbol{x}) \propto e^{-E(\boldsymbol{x})/D(\boldsymbol{\sigma})}$$

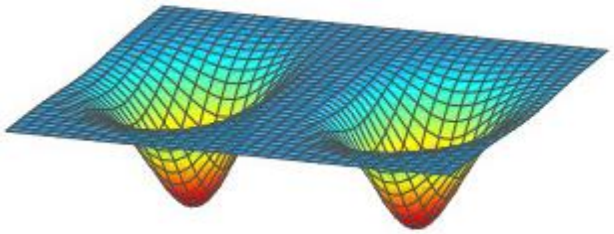
Hopfield, *PNAS* 1984
Ackley, Hinton, Sejnowski, *Cognitive Science* 1985

BBN samples the Bayesian posterior of the decision variable

Hopfield energy

$$E = \overbrace{\left\{-\frac{1}{2}\sum_{i,j,i\neq j}^{N} w_{ij} f(x_i) f(x_j) + \sum_{i}^{N}\left[x_i f(x_i) - \int_0^{x_i} f(x)dx\right] - \sum_{i}^{N} b_i f(x_i)\right\}}^{E_{internal}} - \overbrace{\left\{\sum_{i}^{N} \bar{I}_i f(x_i)\right\}}^{E_{external}}$$

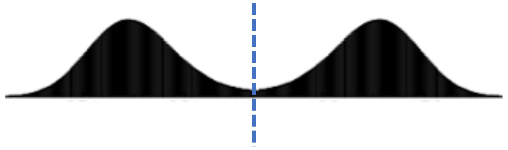
$E_{internal}$



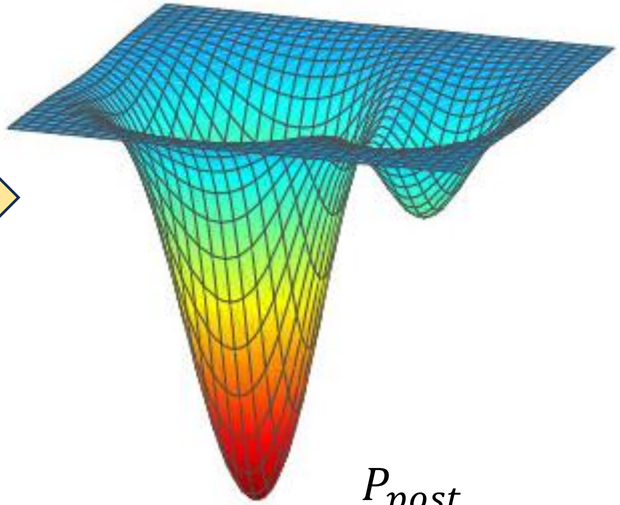
hypothesis 1

hypothesis 2

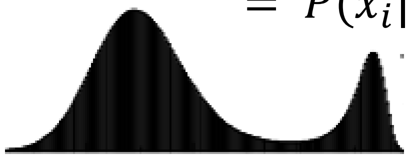
$P_{prior} = P(x_i)$



$E|\boldsymbol{I} = E_{internal} + E_{external}$



P_{post}
$= P(x_i|\boldsymbol{I})$



Hinton and Sejnowski, 1983
(for discrete Hopfield network)

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
Washington, D. C., June, 1983

OPTIMAL PERCEPTUAL INFERENCE

Geoffrey E. Hinton — Computer Science Department, Carnegie-Mellon University

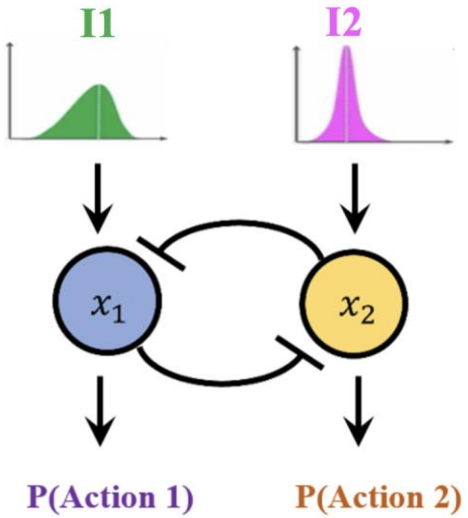
Terrence J. Sejnowski — Biophysics Department, The Johns Hopkins University

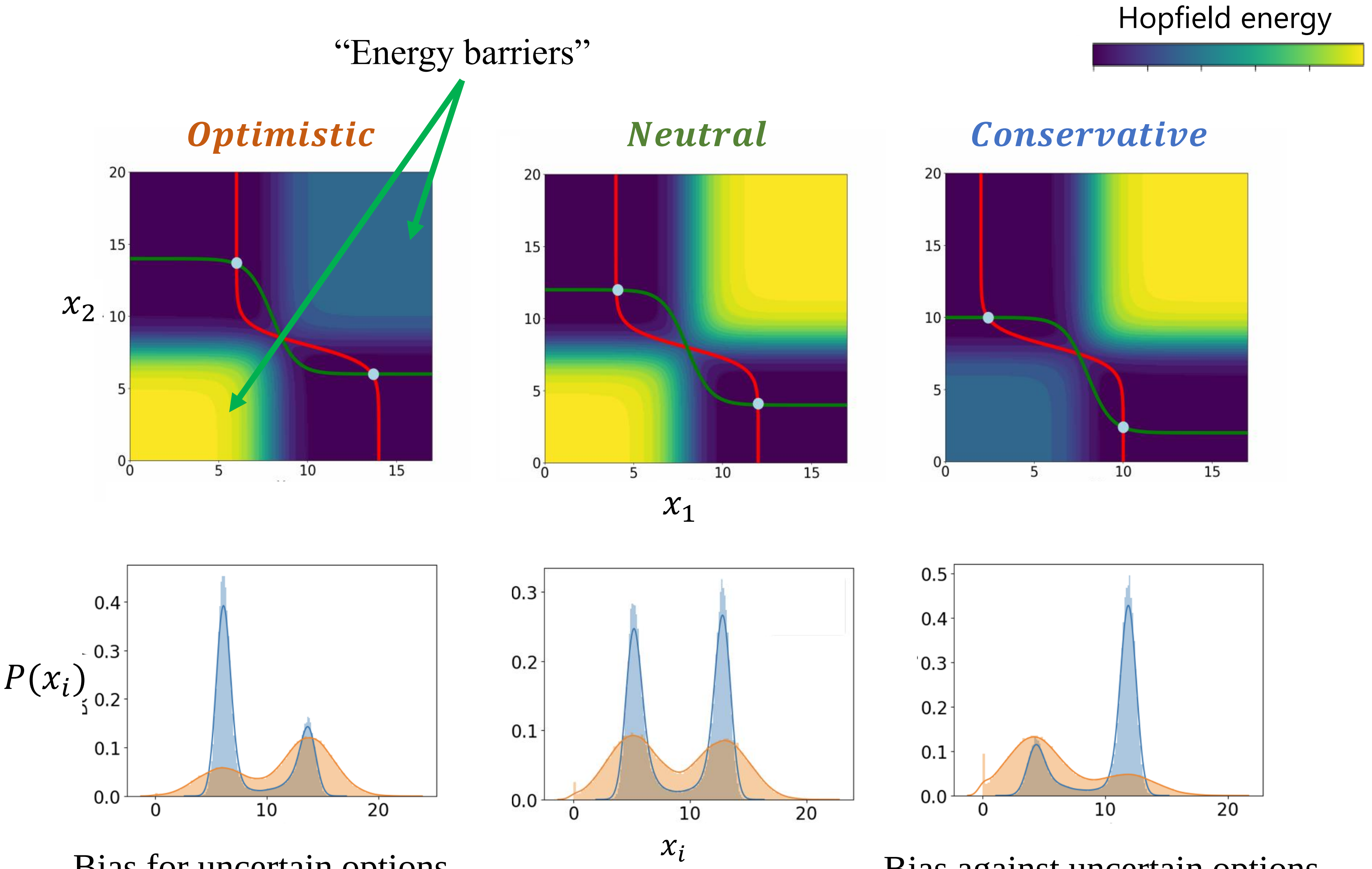
Hopfield energy — **Bayesian inference**

$$p_k = \frac{1}{(1+e^{-\Delta E_k/T})} = 1/(1+e^{-(\ln\frac{p(h)}{p(\bar{h})} + \ln\frac{p(e|h)}{p(e|\bar{h})})})$$

$$\doteq 1/(1+\frac{p(\bar{h})\,p(e|\bar{h})}{p(h)\,p(e|h)})$$

BBN generates tunable bias for or against input uncertainty





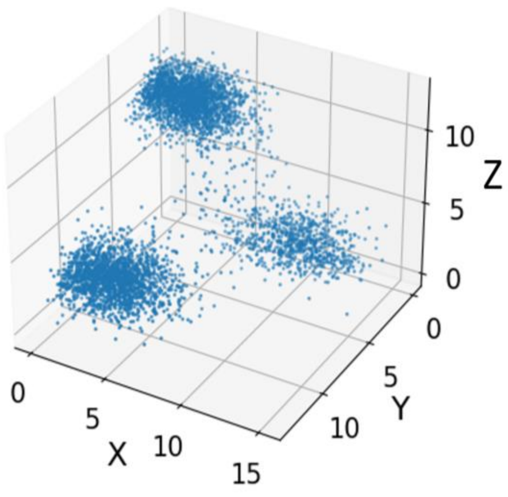
Bias for uncertain options

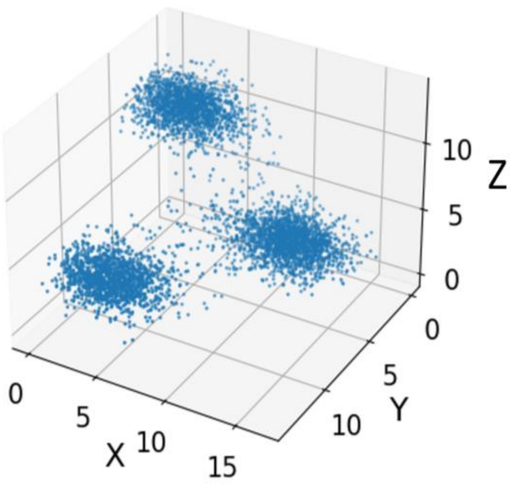
Bias against uncertain options

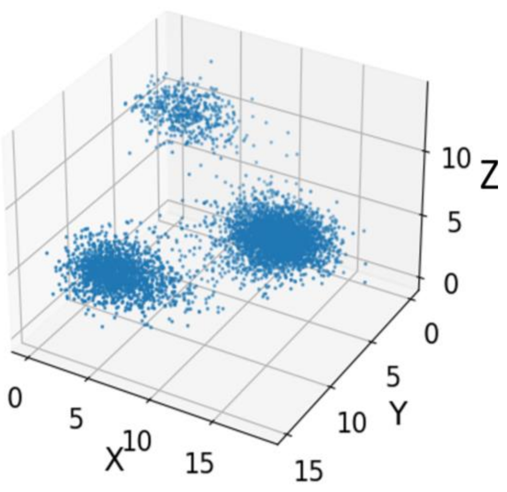
Uncertainty bias scales to high-D without re-tuning

Optimistic | Neutral | Conservative

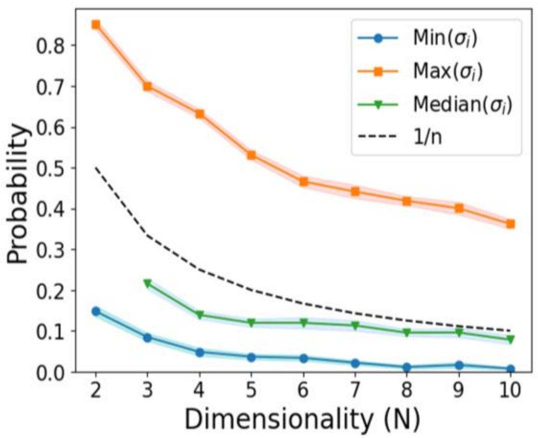
3D BBN

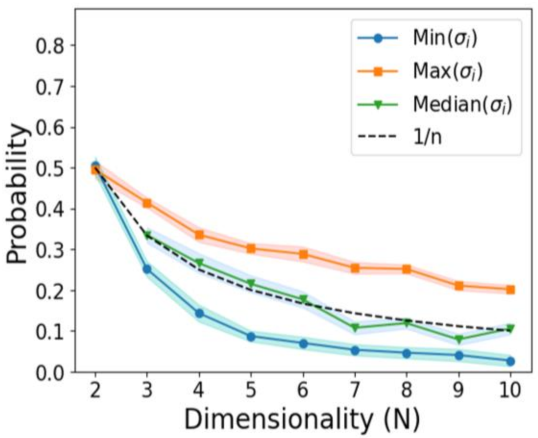


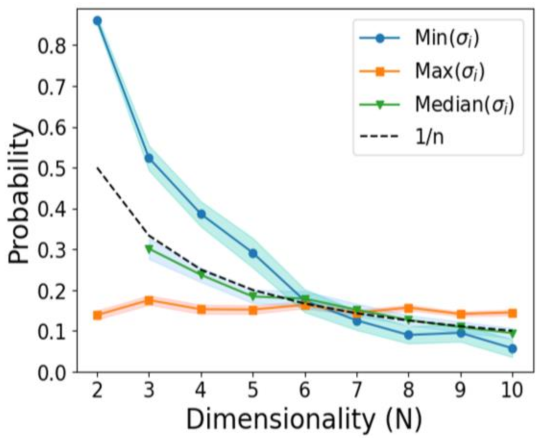




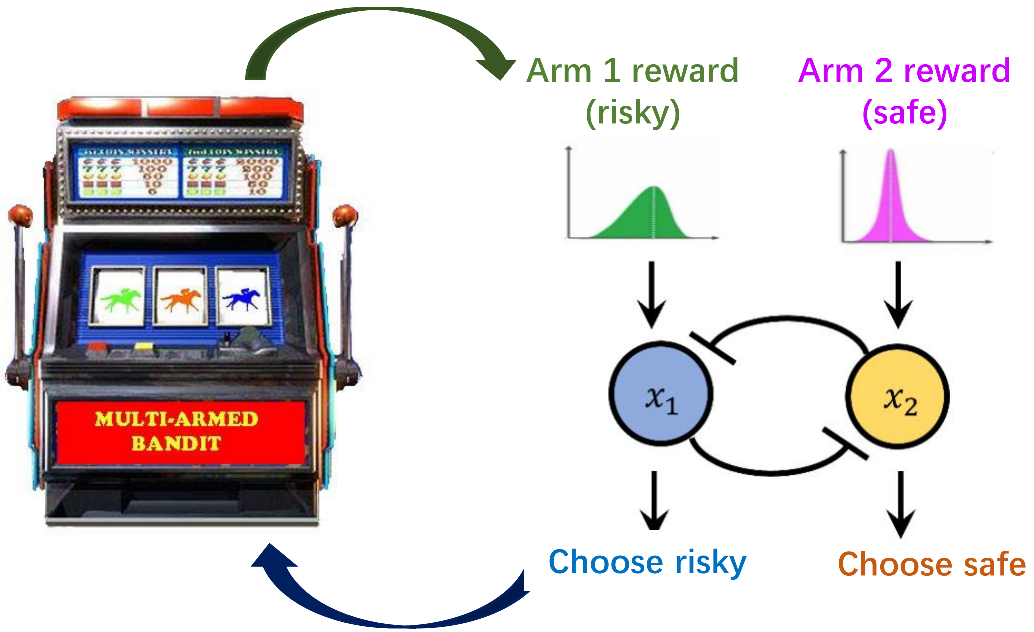
N-D BBN



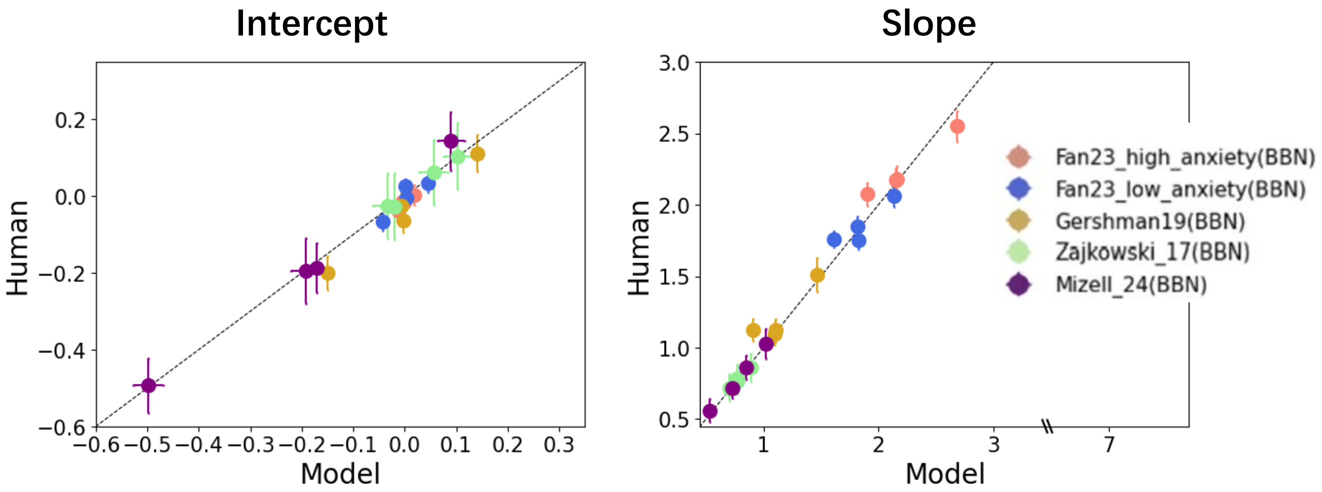




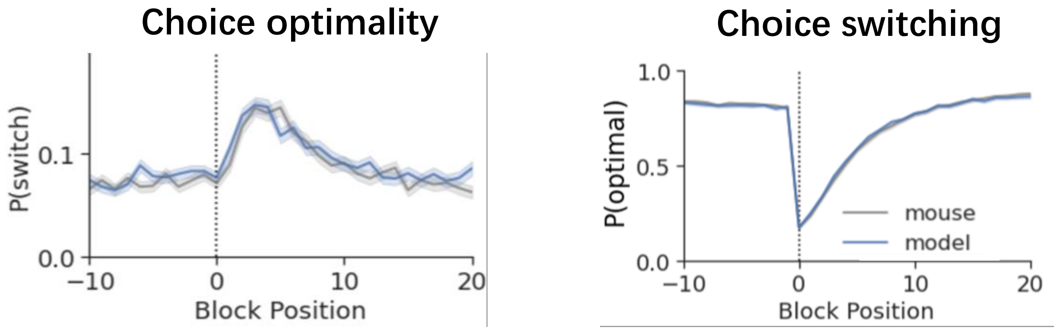
BBN closely captures human and animal bandit choice patterns



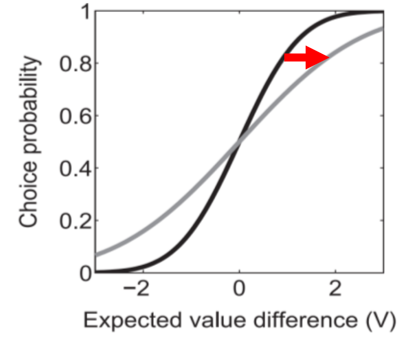
Humans
(grouped by age, mental state, task horizon)



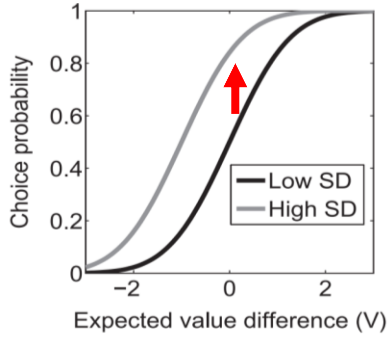
Mice
(switching bandit)



Thompson Sampling (Bayesian inf.): slope shift



Upper Confidence Bound (optimistic): Intercept shift



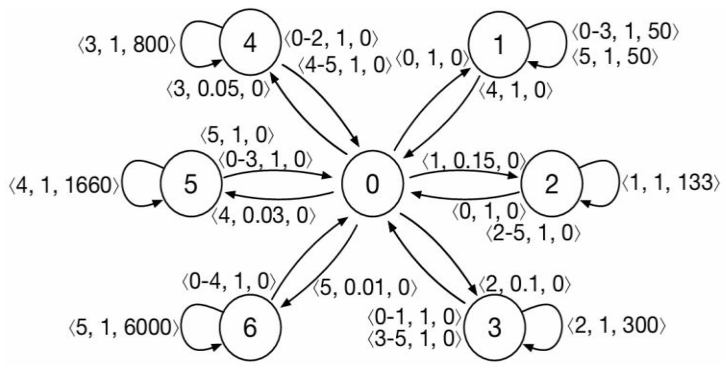
Wilson et al., 2014
Gershman, Cognition 2018

Data from Beron et al., 2022

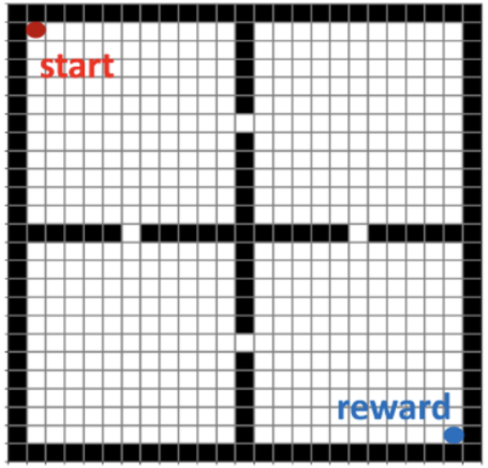
Model for exploration behavior across species

Optimistic BBN promotes efficient exploration in MDP tasks

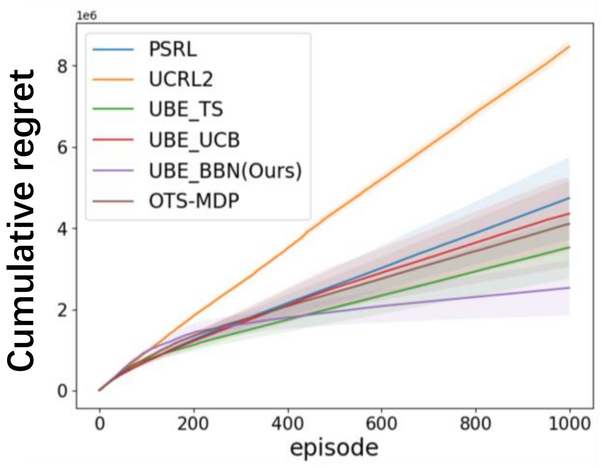
SixArms

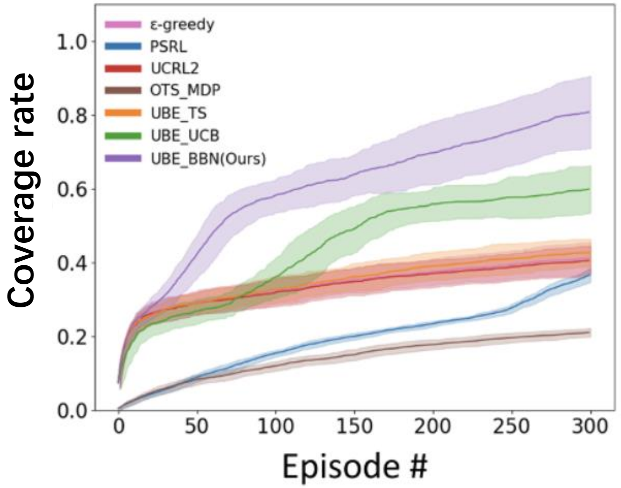


FourRooms

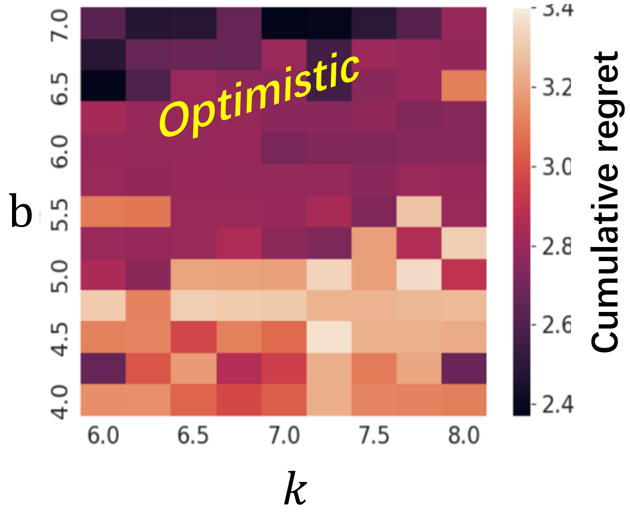


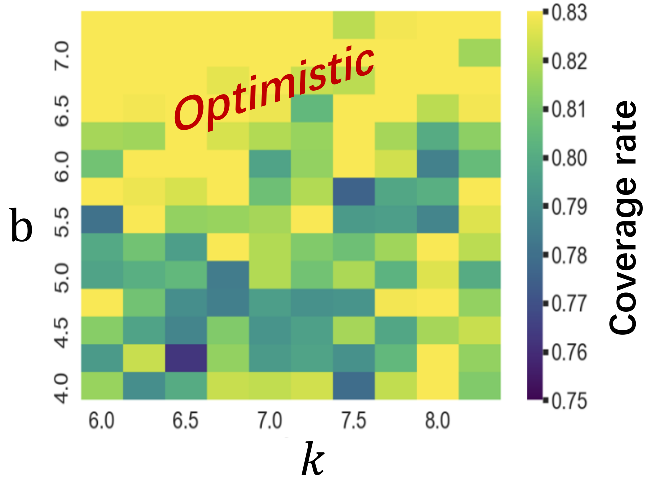
Improved performance



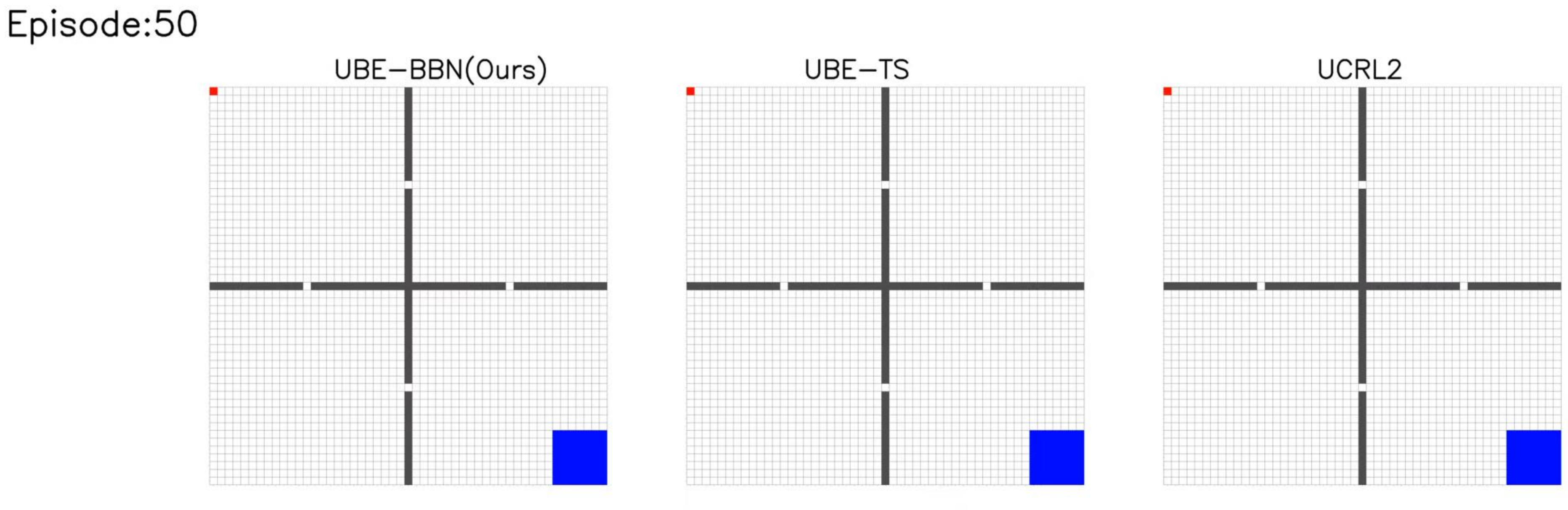


No need for fine-tuning



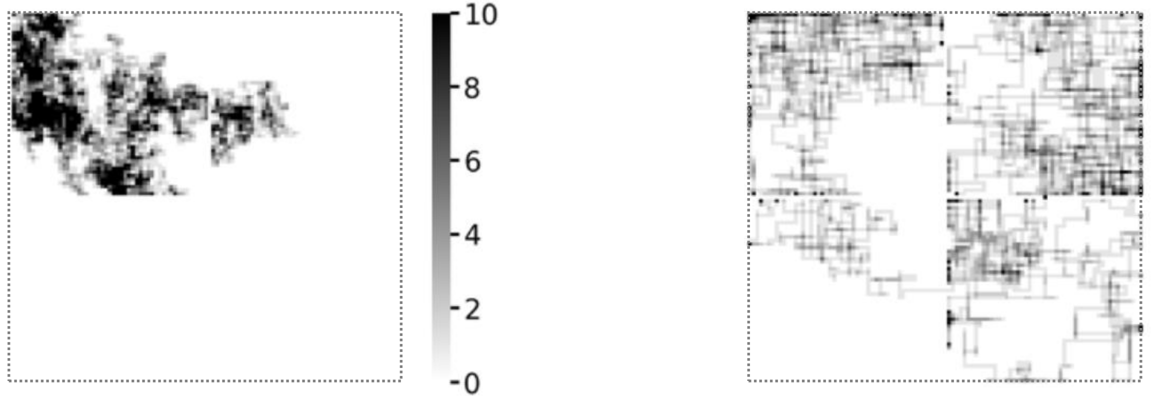


Exploiting attractor persistence further enhances exploration



w/ persistence (memory)

w/o persistence (memory)



See also: Amin et al., ICML 2021
Osband et al., ICLR 2020

Current limitations & future promises

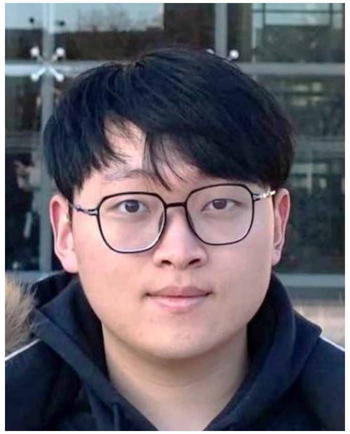
- How to estimate uncertainty in continuous space MDP tasks? How does the brain do it?

- How does the network learn? →Hebbian learning may work

- Computational cost with SDEs?
  1. Replace simulation with better analytical approximation
  2. Neuromorphic device

Looking for postdocs, students & collaborators!

Thank you for listening!



Chen Jiang
(now @ McGill)

Jiahui An

Yating Liu

NSFC
National Natural Science Foundation of China

Paper *Code* *Lab*