Novelty drives human exploration even when it is suboptimal

Alireza **Modirshanechi,** Wei-Hsiang **Lin,** He A. **Xu,** Michael H. **Herzog,** and Wulfram **Gerstner**

Ecole Polytechnique Fédérale de Lausanne (**EPFL**), Switzerland; Contact: **alireza.modirshanechi@epfl.ch**;

preprint →

1. Introduction

- How do humans **explore** environments with **sparse** rewards?

  Intrinsically motived RL algorithms have been proposed in computational and behavioral neuroscience as models of human exploration [1]. However, **different choices of intrinsic reward** result in fundamentally **different exploration strategies**. [2]

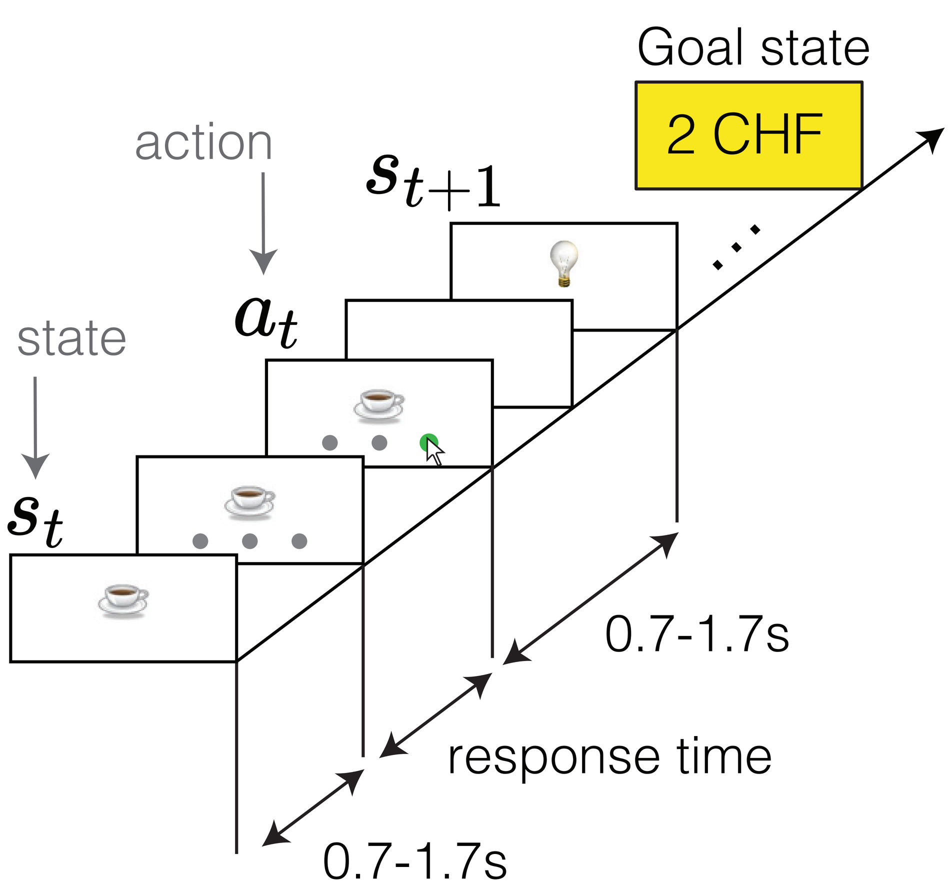
- **Which intrinsic reward** explains human exploration best?

  Our contribution: Inspired by the **"noisy TV" problem** in machine learning [3], we design an experimental paradigm where **three representative intrinsic rewards** (novelty [4,5], surprise [6,7], and information-gain [8-10]) make **different behavioral predictions**. We test these predictions against **human behavior.**

[1] Gottlieb and Oudeyer, 2018;
[2] Aubret et al., 2022;
[3] Burda et al., 2019;
[4] Bellemare et al., 2016;
[5] Xu and Modirshanechi et al., 2021;
[6] Kobayashi et al., 2019;
[7] Pathak et al., 2017;
[8] Itti and Baldi, 2009;
[9] Schmidhuber, 2010;
[10] Horvath et al., 2021

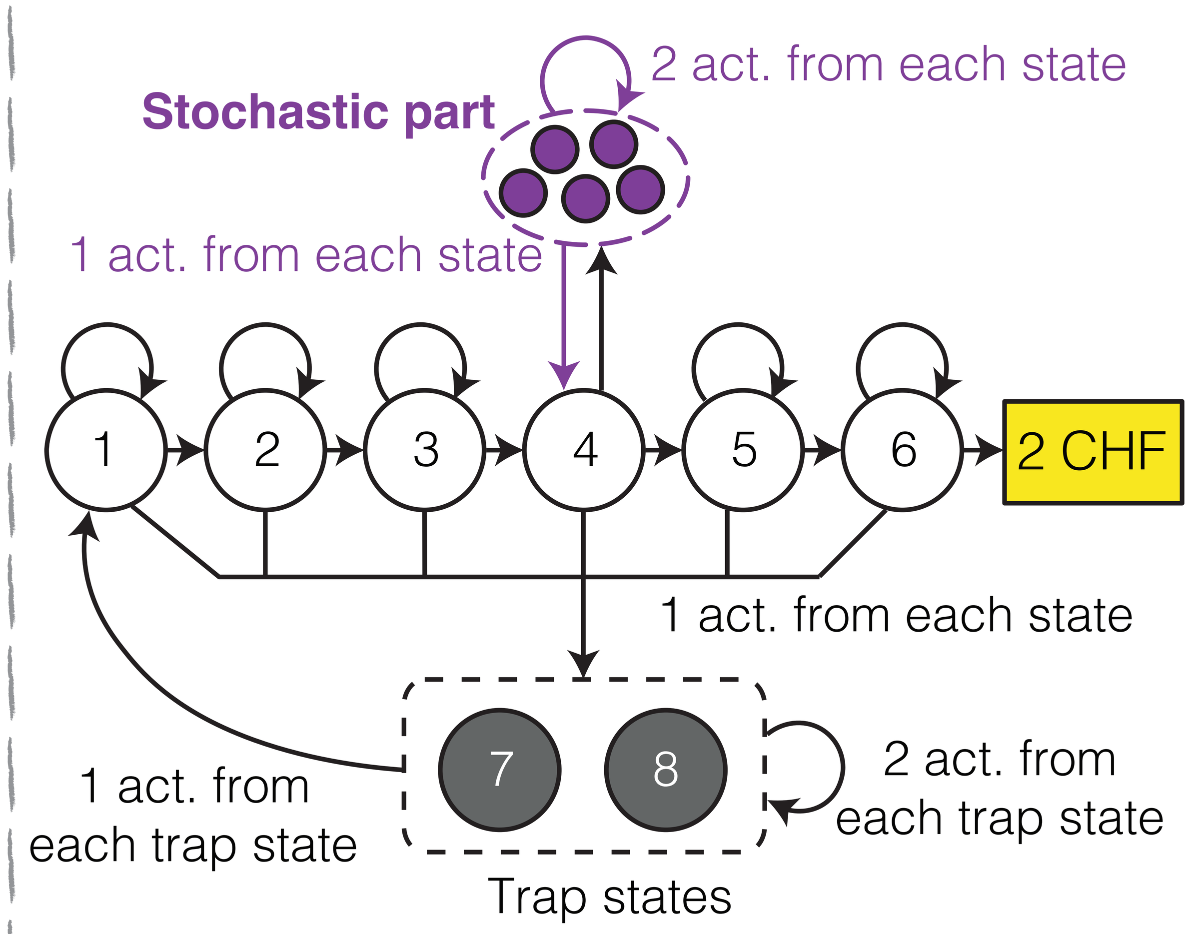
2. Experimental paradigm:

Multistep decision-making:

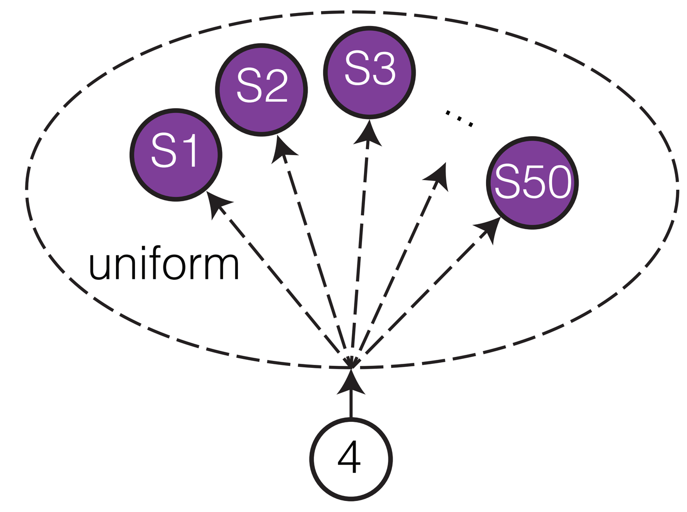


2.1. Underlying map (unknown to the participants)

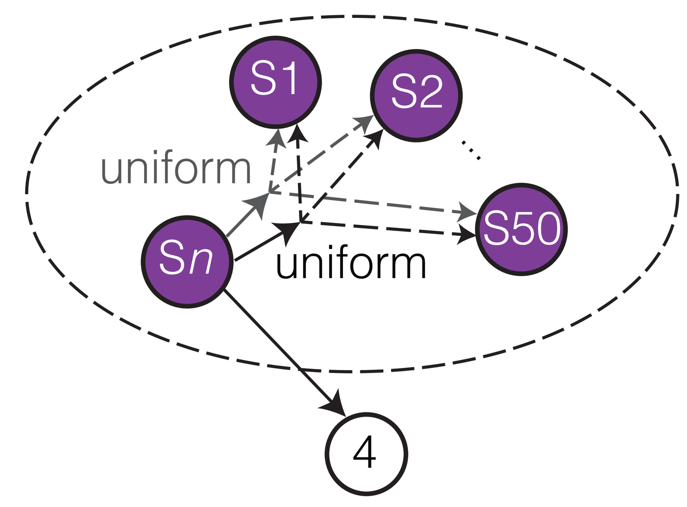
58 states + 3 actions per state:



2.1.1. Transitions to stoch. states

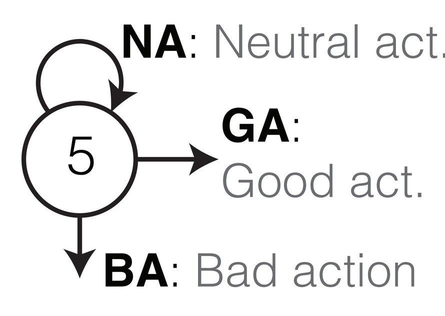


2.1.2. Transitions from stoch. states

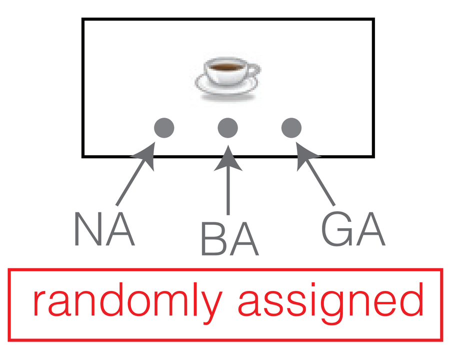


2.1.3. State representation

An examplar state:



Representation in the experiment:



2.2. Instruction given to participants:

There are **3 goal states** with different reward values: 2 CHF and 3 CHF and 4 CHF — CHF: Swiss Franc

- Participants were instructed to move to **any** of the three goal states **5 times** (= 5 episodes).

2.3. Reward manipulation:

There was **only one goal state** (unknown to the participants): 2 CHF **OR** 3 CHF **OR** 4 CHF — CHF: Swiss Franc

- We focus on the group of participants with **lowest** reward: 2 CHF — Highly motivated to explore in episodes 2-5 to find the larger rewards!
  (see our **preprint** for the other groups)

3. Results

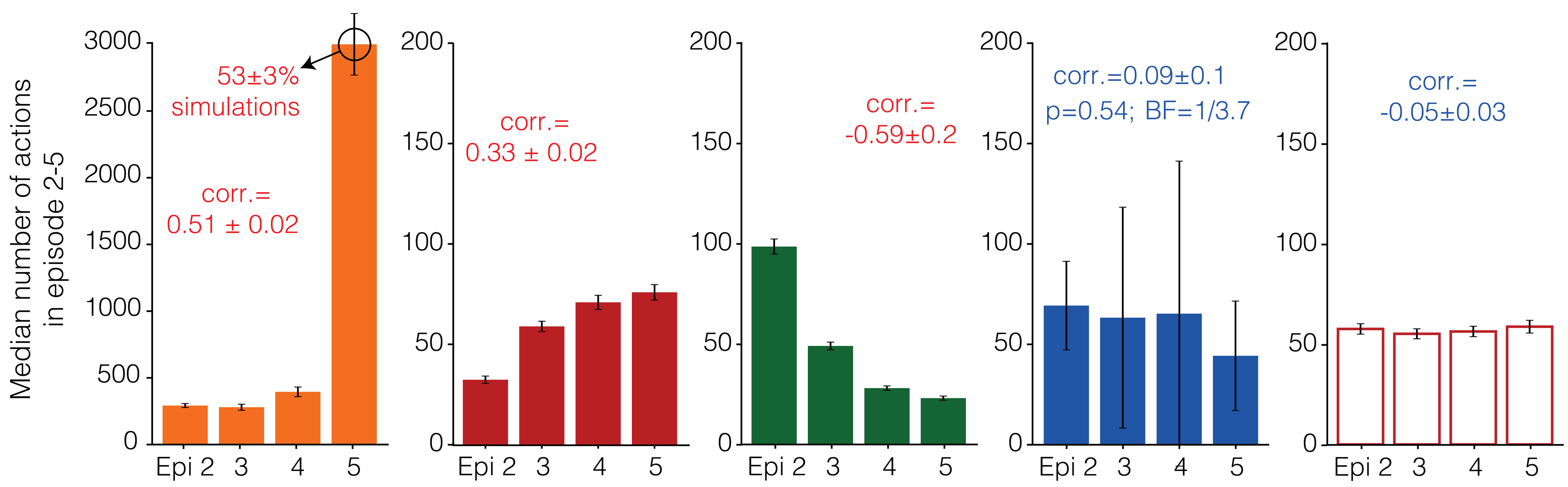
Predictions before data:

- purely surprise-seeking (N=500)
- purely novelty-seeking (N=500)
- purely inf.-gain-seeking (N=500)

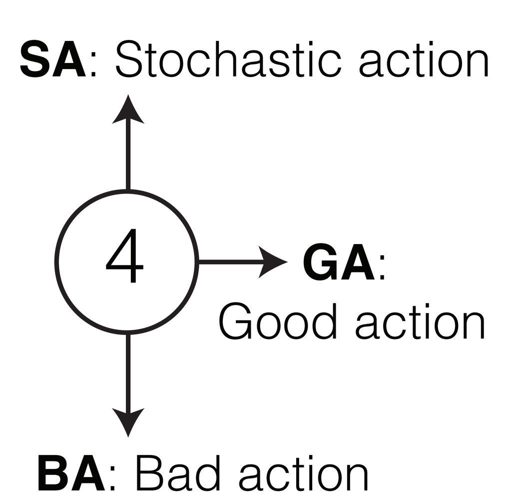
Analyses with data:

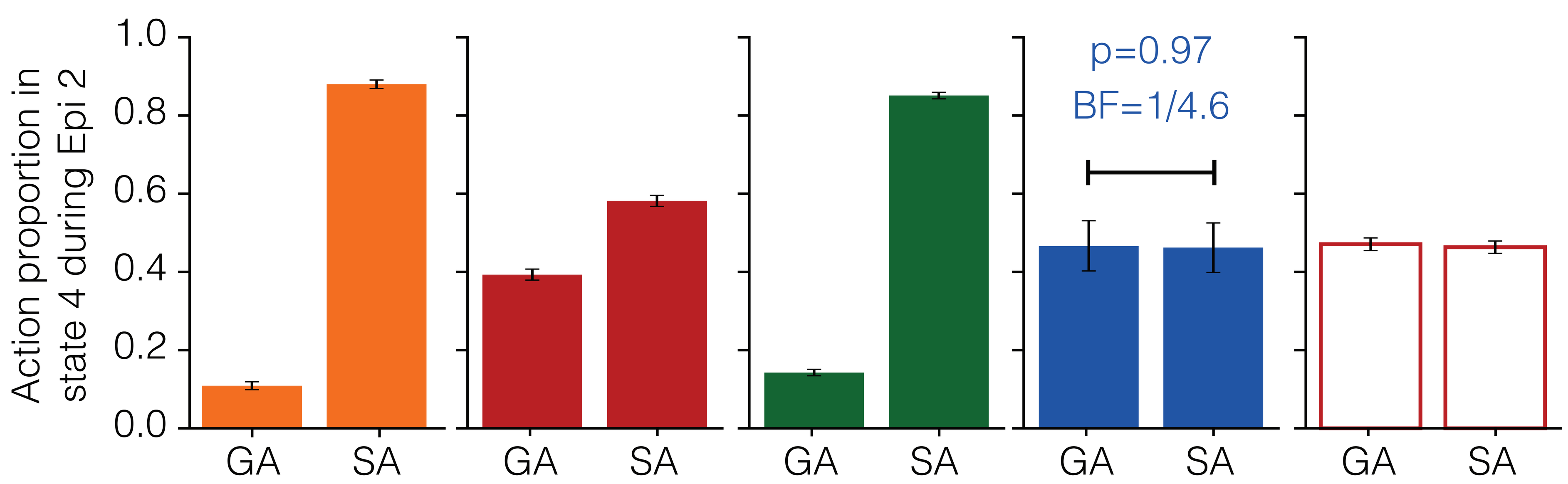
- human behavior (N=21)
- fitted novelty-seeking (N=500) (posterior pred. checks)

3.1. Humans and novelty-seeking agents exhibit a persistent attraction to stochasticity during eipsodes 2-5.

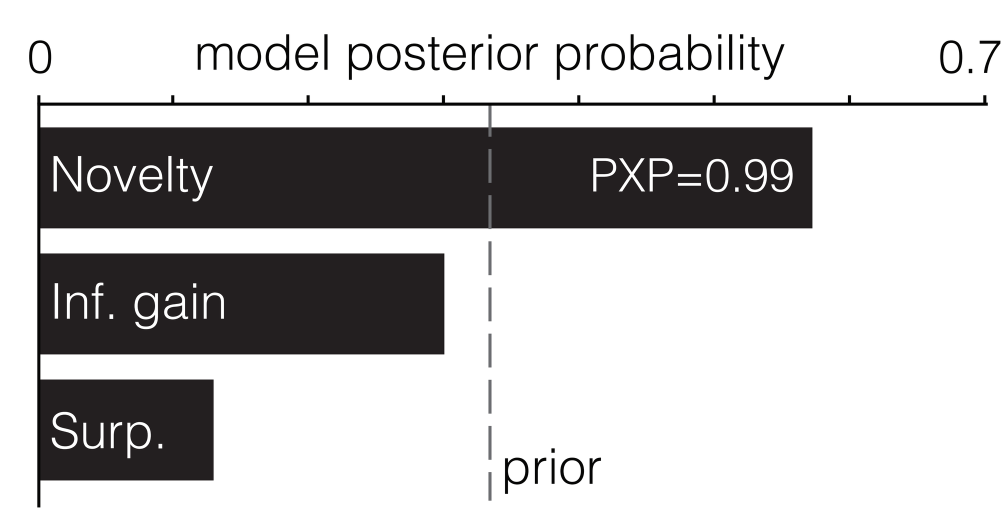


3.2. Humans and novelty-seeking agents show a similar preference for GA and SA during Epi 2





3.3. Bayesian model-selection:



4. Conclusions

1. Human participants who are **optimistic** about the availability of goal states of higher value than those already known exhibit **a persistent attraction** to stochasticity.
2. This behavior is **consistent with** that of **novelty-driven** agents but **NOT** with those driven by **information-gain (≈ optimal behavior)** or **surprise.**
3. Our work suggests that humans use **suboptimal** but **computationally cheap** policies (such as novelty-seeking) for exploration in **complex environments**.

This research was supported by Swiss National Science Foundation No. CRSII2 147636 (Sinergia), No. 200020_184615, and No. 200020_207426 and by the European Union Horizon 2020 Framework Program under grant agreement No. 785907 (Human Brain Project, SGA2).