Novelty drives **human exploration** even when it is **suboptimal**

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

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

😍 **preprint** →

1. Introduction

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

Recent models in computational and behavioral neuroscience:

- **Intrinsically motived RL algorithms**: incorporating intrinsic reward signals into traditional RL models. [1]
- However, **different choices of intrinsic reward** result in fundamentally **different exploration strategies**. [2]

A new question: **Which intrinsic reward explains human behavior best?**

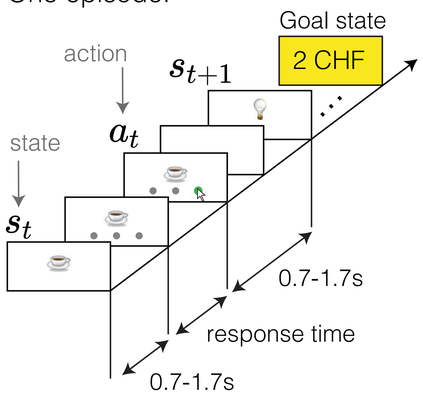
Our contribution:

Inspired by the so-called **"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 the behavior of **human participants**.

[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. Experiment

One episode:



2.1. Instruction given to participants:

There are 3 goal states with different reward values:

CHF: Swiss Franc — 2 CHF and 3 CHF and 4 CHF

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

2.2. Reward manipulation:

To motivate exploration, there was only one goal state:

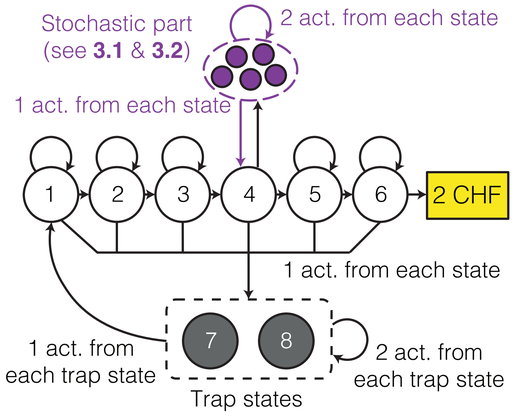
2 CHF **OR** 3 CHF **OR** 4 CHF

We focus on the group of participants with **lowest** reward: 2 CHF

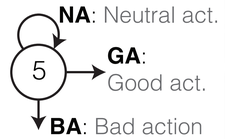
Highly motivated to explore in episodes 2-5 to find the larger rewards!

3. Underlying map (unknown to the participants)

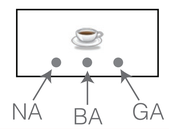
- An environment with **58 states** + **3 actions per state**.



An examplar state:

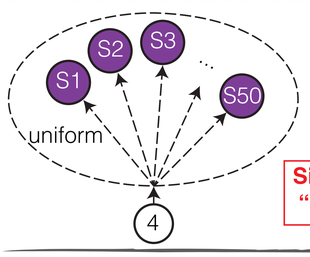


Representation in the experiment:

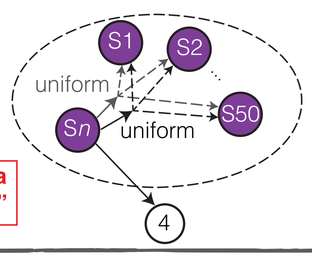


randomly assigned

3.1. Transitions to stoch. states



3.2. Transitions from stoch. states



Similar to a "noisy TV"

4. Results

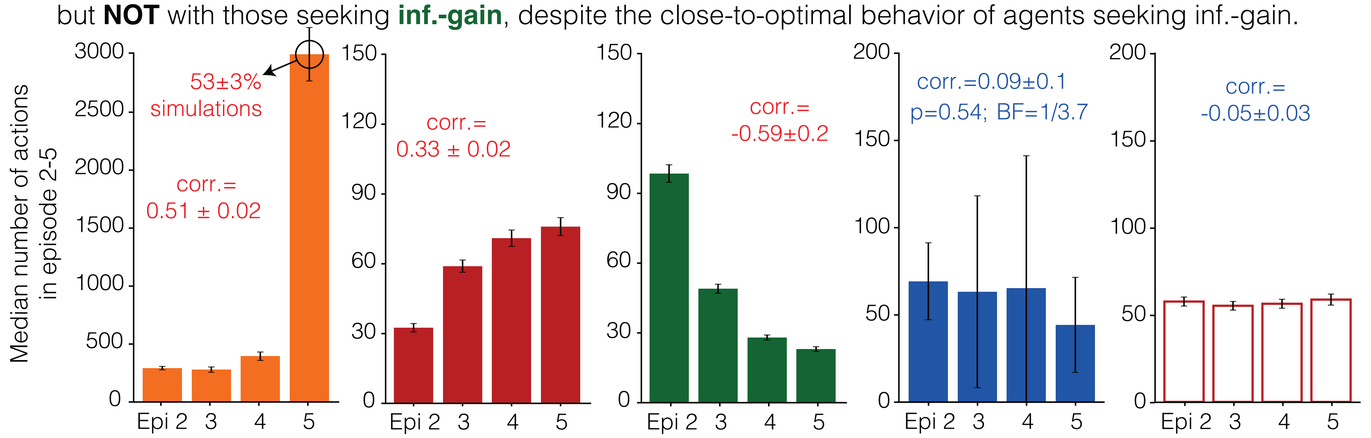
Predictions before data:

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

Analyses with data:

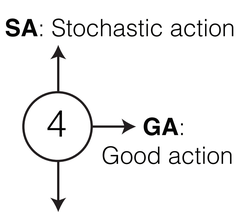
- human behavior (N=21)
- fitted novelty-seeking (N=500) (posterior pred. checks)

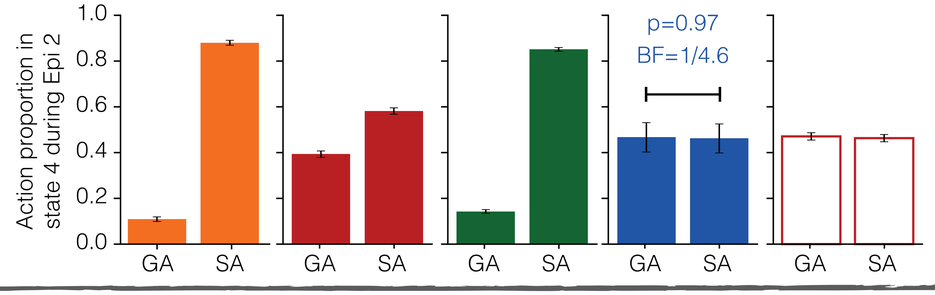
4.1. Human persistent attraction to the stochasticity is consistent with the behavior of agents seeking novelty but NOT with those seeking inf.-gain, despite the close-to-optimal behavior of agents seeking inf.-gain.



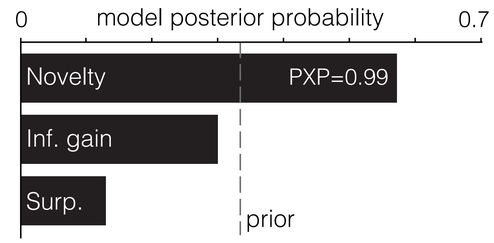
4.2. Humans and novelty-seeking agents show a similar preference for GA and SA during Epi 2

SA: Stochastic action





4.3. Model-selection:



5. Conclusions

5.1. Simulated agents

1. Simulated agents driven by **inf.-gain** eventually **lose their interest** in stochasticity when they realize that there is **no information to gain**.
2. Agents driven by **novelty** exhibit a **persistent attraction** to stochasticity.
3. Agents driven by **surprise** exhibit a **detrimentally increasing attraction**.

5.2. Human participants

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 both qualitatively and quantitatively **consistent with** that of **novelty-driven** agents and **NOT** with those driven by **inf.-gain (≈ optimal behavior)** or **surprise.**

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).