



Learn and Teach Structured Knowledge by Balancing Curiosity and Competence

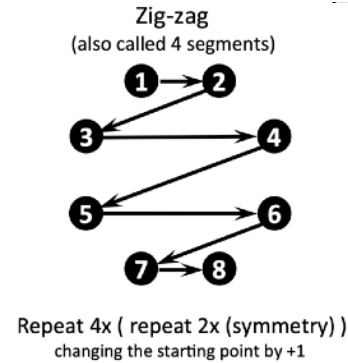
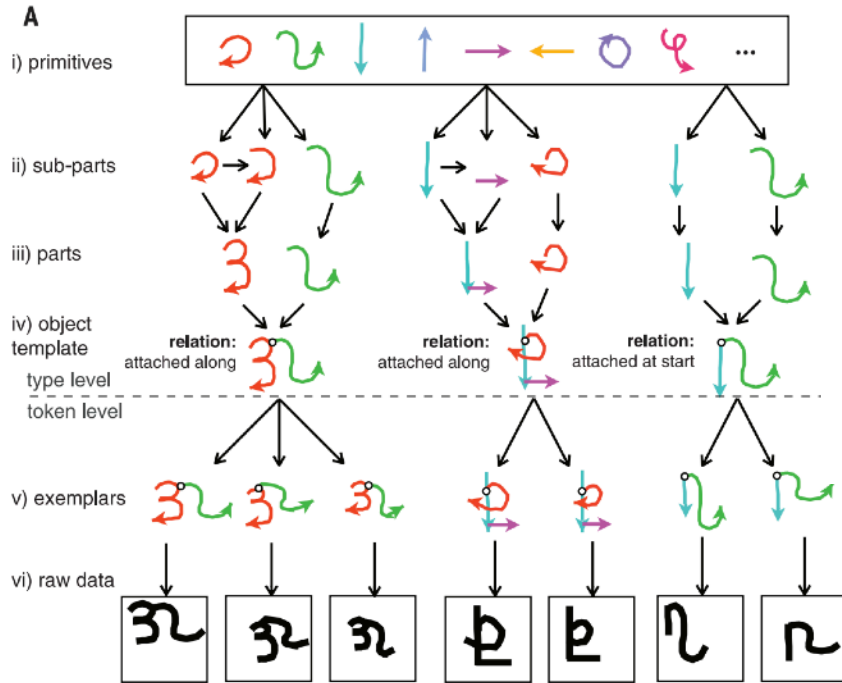



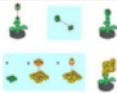
Hanqi Zhou

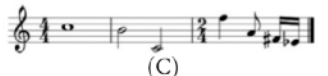
hanqi.zhou@uni-tuebingen.de


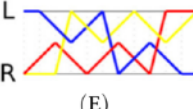
*Cluster of Excellence “Machine Learning for Science”, Tübingen AI Center,
University of Tübingen*

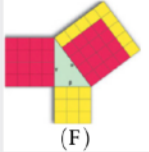
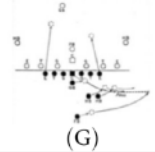
How human learn structured knowledge?



(A)  (B) 

(C) 

(D)  (E) 

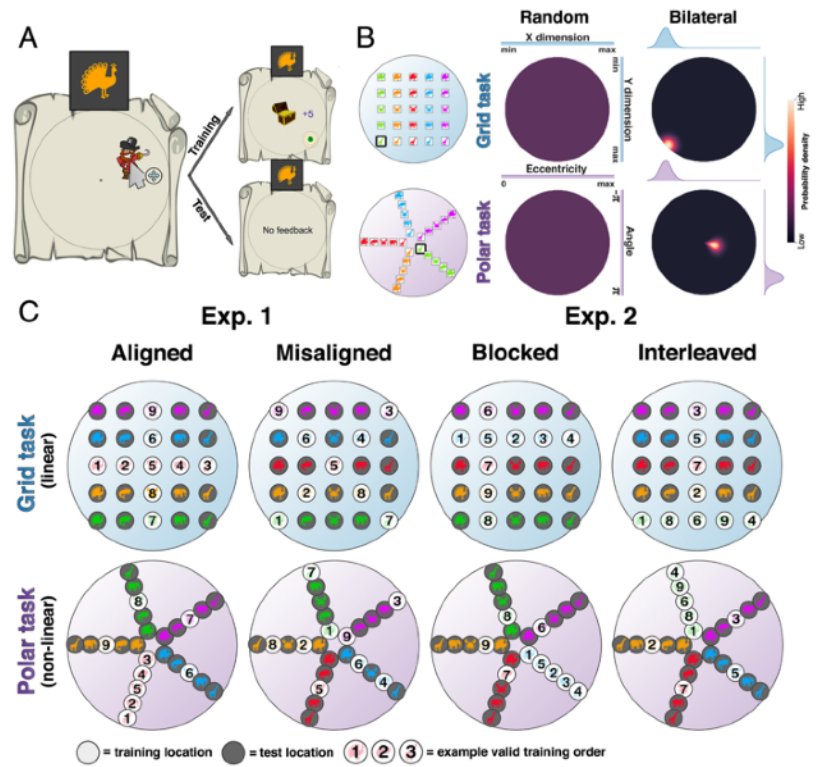
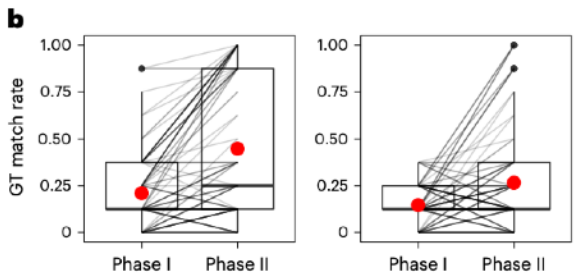
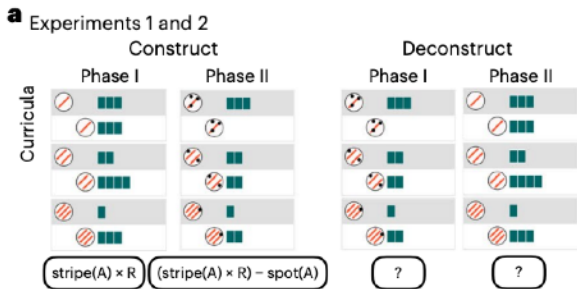
(F)  (G) 

Trends in Cognitive Sciences

Figure 1. Several Classes of Programs Expressed as Symbolic Images. (A) Blueprints, (B) assembly instructions, (C) musical notation, (D) knotting diagrams, (E) juggling patterns, (F) graphical proofs, and (G) football plays.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266), 1332-1338.
 Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S., & Sablé-Meyer, M. (2022). Symbols and mental programs: a hypothesis about human singularity. *Trends in Cognitive Sciences*, 26(9), 751-766.
 Rule, J. S., Tenenbaum, J. B., & Piantadosi, S. T. (2020). The child as hacker. *Trends in cognitive sciences*, 24(11), 900-915.

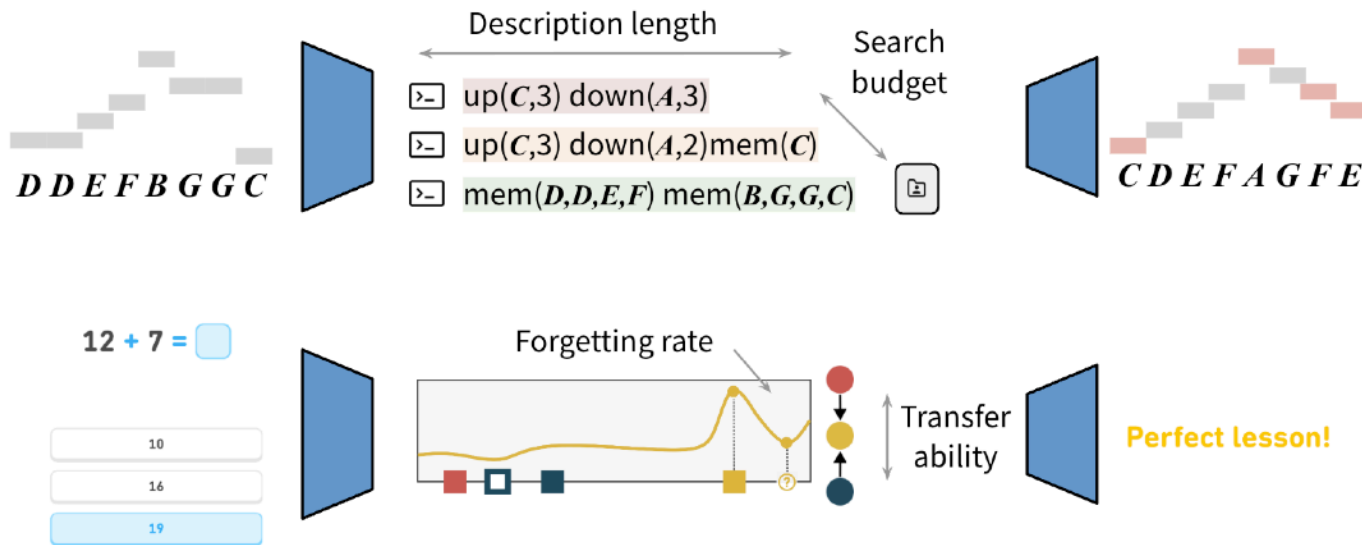
What is the implication for education practice?



Zhao, B., Lucas, C. G., & Bramley, N. R. (2024). A model of conceptual bootstrapping in human cognition. *Nature Human Behaviour*, 8(1), 125-136.

Dekker, R. B., Otto, F., & Summerfield, C. (2022). Curriculum learning for human compositional generalization. *Proceedings of the National Academy of Sciences*, 119(41), e2205582119.

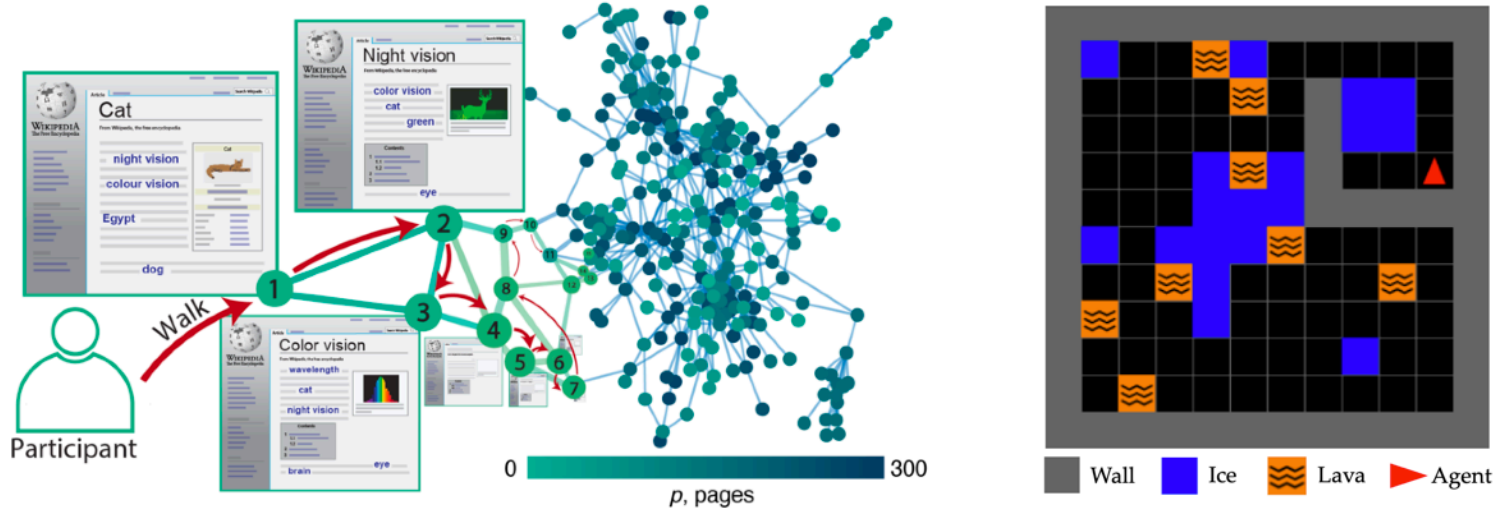
- Learner: learning under resource constraints
- Teacher: find optimal order of a fixed set of learning materials to **maximize performance**



Learning & teaching: motivating



- Learner: learning under resource constraints
- Teacher: maximize performance while **motivating** learners

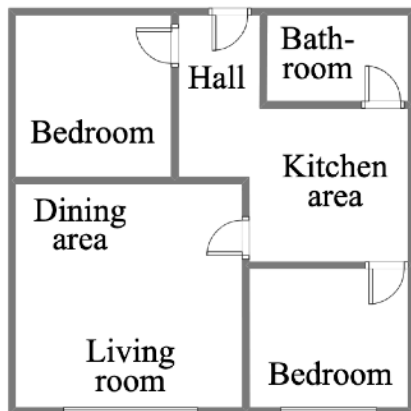


Patankar, S. P., Zhou, D., Lynn, C. W., Kim, J. Z., Ouellet, M., Ju, H., ... & Bassett, D. S. (2023). Curiosity as filling, compressing, and reconfiguring knowledge networks. *Collective Intelligence*. CogSci 2025

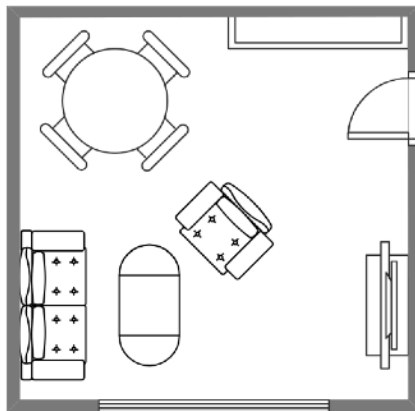
- Learner: learning under resource constraints
- Teacher: useful **interactions** for maximize performance while motivating learners

Granularity

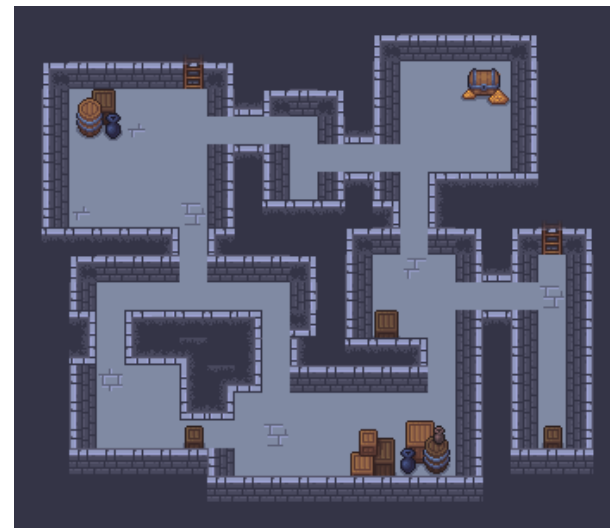
Global AI: assigns functional roles to rooms within a home



Local AI: optimises furniture arrangements within a room



Explanation



When extrinsic reward is sparse... in our real life



Knowledge-seeking (prediction)

- reduce uncertainty about their environment by seeking out new or surprising information

Competence-seeking (control)

- maximize their control and influence over their environment

Epistemic, perceptual, specific, diverse
(Berlin, 1966)

Temporary state or enduring personality
trait (Boyle, 1983)

Motivates for information seeking
(Lowenstein, 1994)

Wanting and liking, emotional-
motivational state (Litman, 2005)

A metacognitive feeling (Metcalf et al.,
2020)

...

Knowledge-seeking (prediction)

Entropy as ignorance (Lowenstein, 1994)

- Novelty (state-based): explore more states

$$H(s) = - \sum p(s) \log p(s)$$

Knowledge-seeking (prediction)

Entropy as ignorance (Lowenstein, 1994)

- Novelty (state-based): explore more states

$$H(s) = - \sum p(s) \log p(s)$$

- Information gain (dynamics-based): take actions that are maximally informative about the environment

$$H(S' | s, a) - H(S' | s, a, s')$$

Knowledge-seeking (prediction)

- Novelty (state-based): explore more states

$$H(s) = - \sum p(s) \log p(s)$$

- Information gain (dynamics-based): take actions that are maximally informative about the environment

$$H(S' | s, a) - H(S' | s, a, s')$$

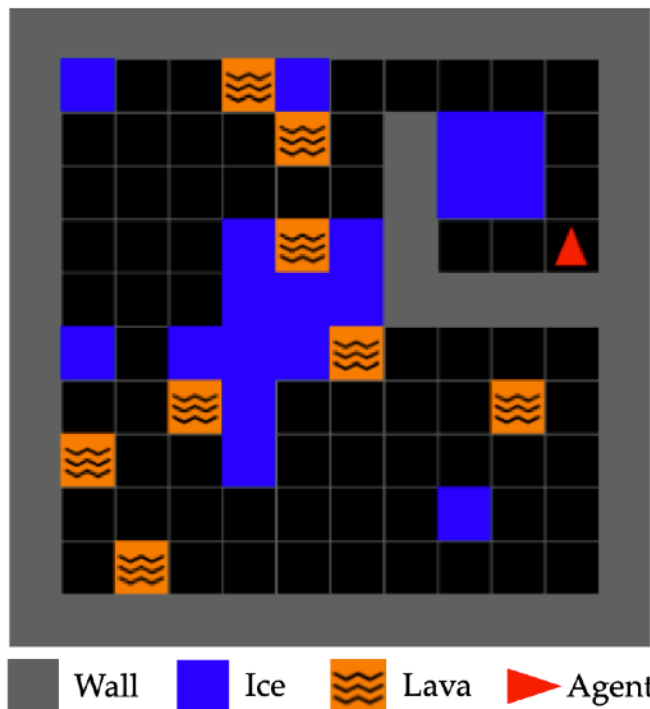
Competence-seeking (control)

- Empowerment: measurable influence over the environment

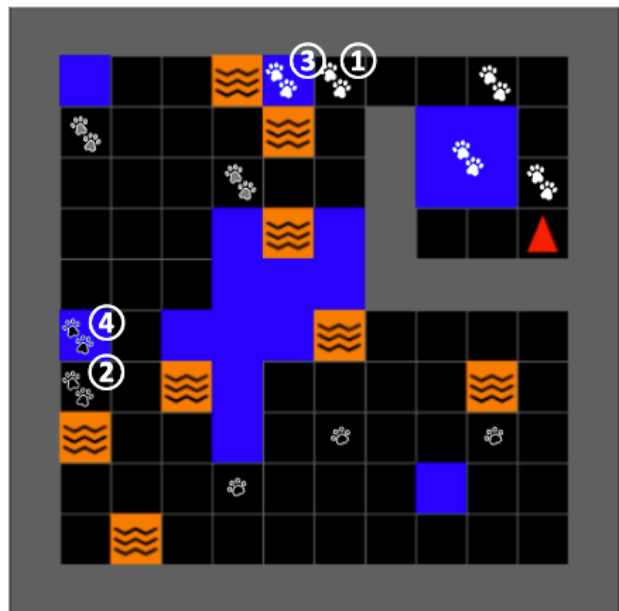
$$\mathfrak{E}_t := C \left(p(s' | s, a) \right)$$

$$= \max_{p(a)} I(S'; \mathcal{A} | s)$$

$$= \max_{p(a)} (H(s' | s) - H(s' | a, s))$$



Schematic illustration of each motivation

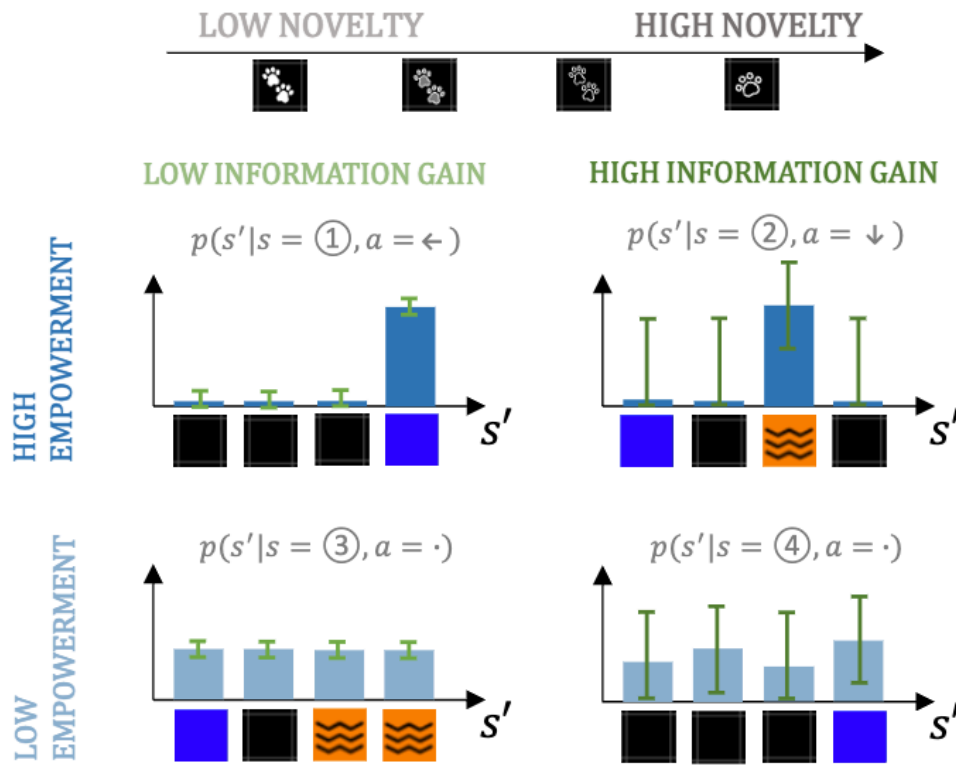
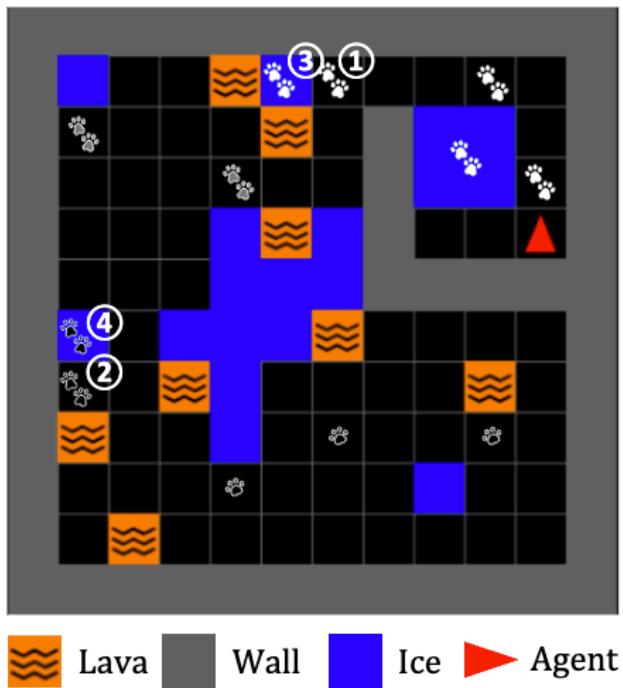


LOW NOVELTY

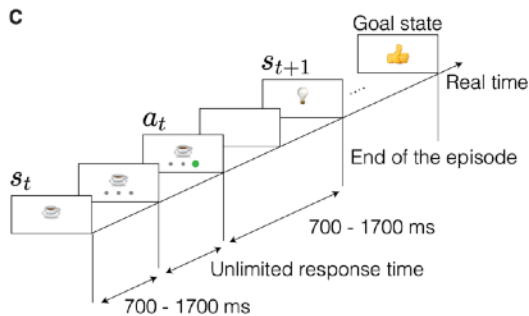
HIGH NOVELTY



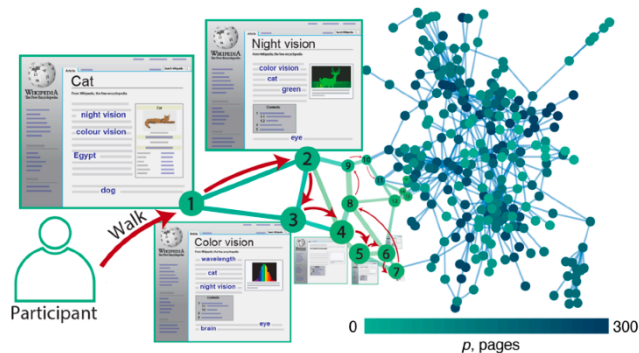
Schematic illustration of each motivation



How to match human exploration?



Game interface



Entropy, surprise, novelty, prediction error,
learning progress, information gain...

Motivation
choice

Approximation (variational, single step...)

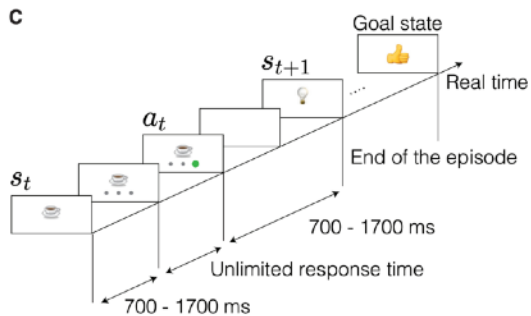
Motivation
implementation

Model

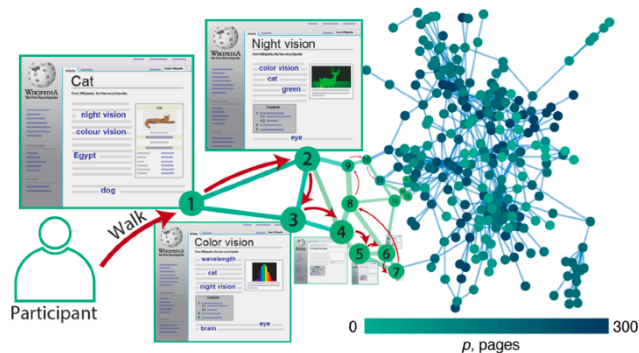
Task design

Xu, H. A., Modirshanechi, A., Lehmann, M. P., Gerstner, W., & Herzog, M. H. (2021). Brändle, F., Stocks, L. J., Tenenbaum, J. B., Gershman, S. J., & Schulz, E. (2023). Lidayan, A., Du, Y., Kosoy, E., Rufova, M., Abbeel, P., & Gopnik, A. (2025). Patankar, S. P., Zhou, D., Lynn, C. W., Kim, J. Z., Ouellet, M., Ju, H., ... & Bassett, D. S. (2023).

How to match human exploration?



Game interface



Entropy, surprise, novelty, prediction error,
learning progress, information gain...

Motivation
choice

Approximation (variational, single step...)

Motivation
implementation

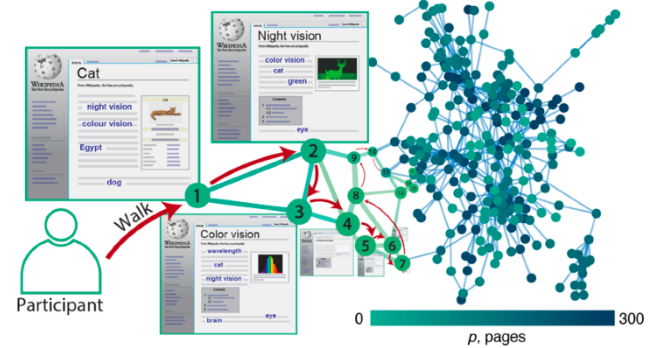
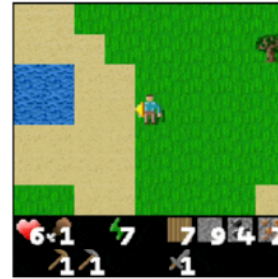
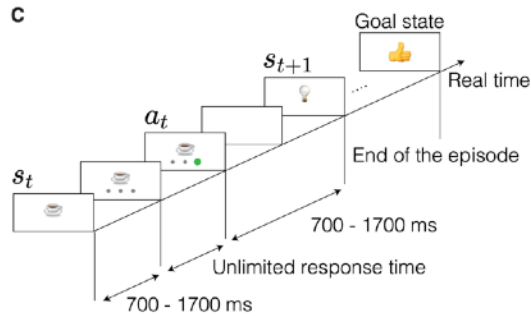
Model

Task design

Updating, world model, representation

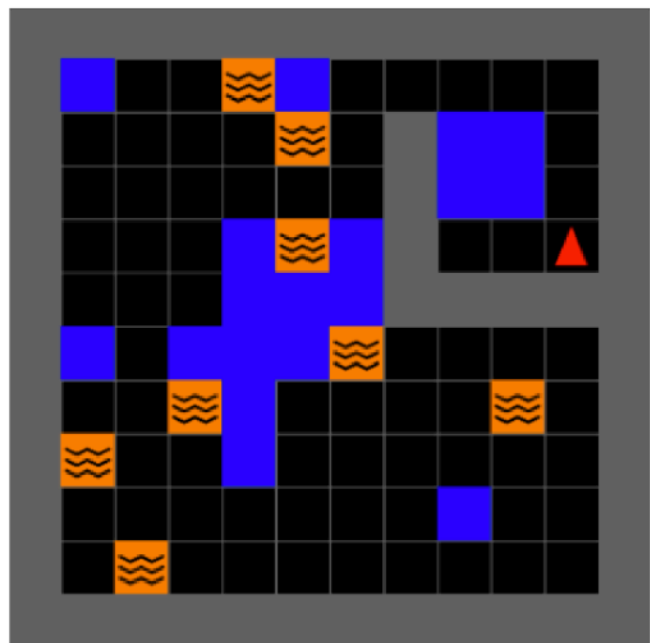
Agent
learning

How to match human exploration?



Synergy between knowledge-seeking and competence-seeking

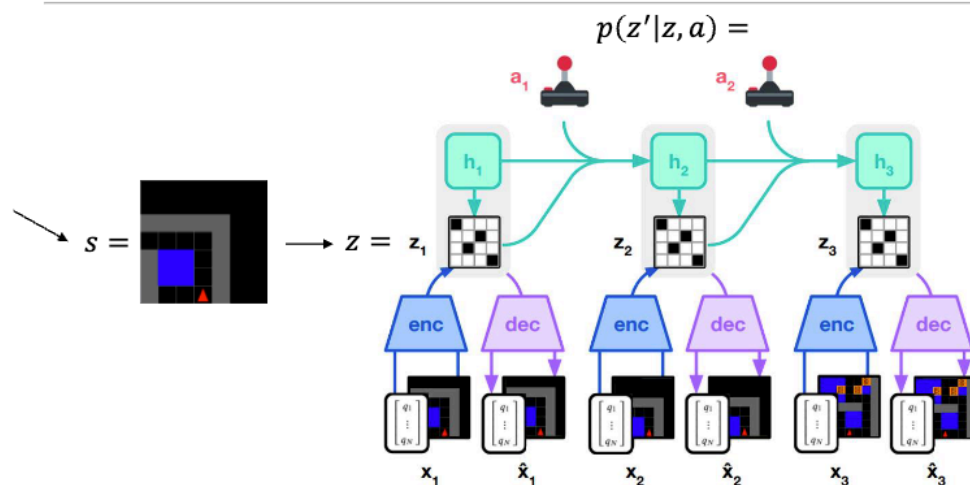
Action-perception loop in world model and exploration

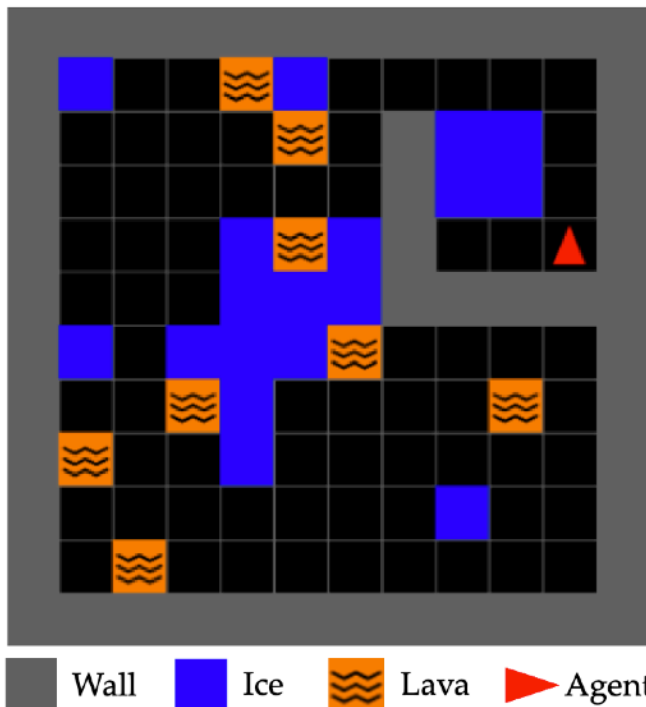


Wall
 Ice
 Lava
 Agent

$$s = (x, y, \varphi) \rightarrow z = (x, y, 0) \quad p(z'|z, a) = p(s'|s, a) = \frac{N(s, a, s')}{N(s, a)}$$

s'	(0, 0, 0)	...	(10, 3, 0)	...	(12, 12, 3)
$p(s' s = (10, 4, 0), a = \text{forward})$	0.0	0.0	1.0	0.0	0.0



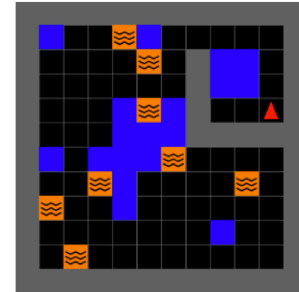
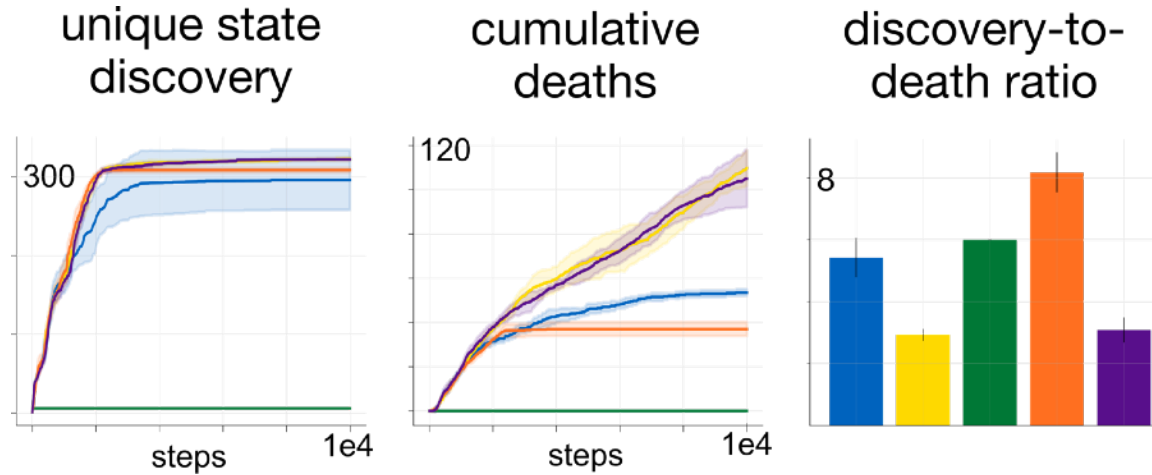


- Fixed state representation $s = (x, y, \theta) \in \mathbb{N}^3$
- Learn transition function and state's Q-value

$$p(s' | s, a) = \frac{\beta N(s, a, s') + 1}{\sum_{s' \in \mathcal{S}} \beta N(s, a, s') + 1}$$

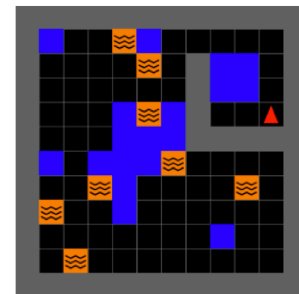
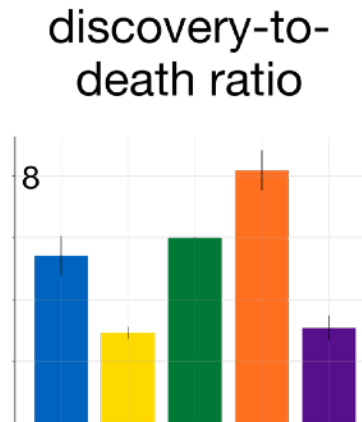
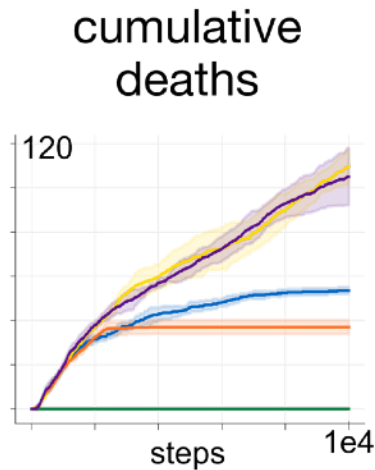
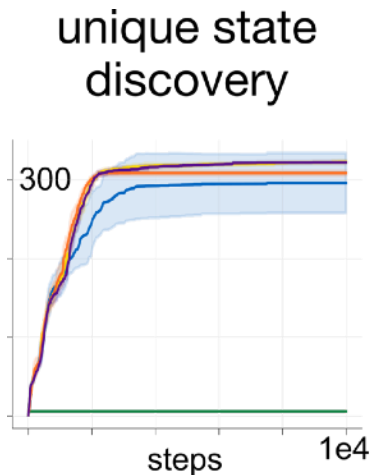
$$Q(s, a) = \sum_{s' \in \mathcal{S}} p(s' | s, a) [R(s, a, s') + \gamma \max_{a'} Q(s', a')]$$

Q-learning agent: exploration effectiveness



■ novelty ■ info gain ■ empowerment ■ info gain + empowerment ✱ empowerment

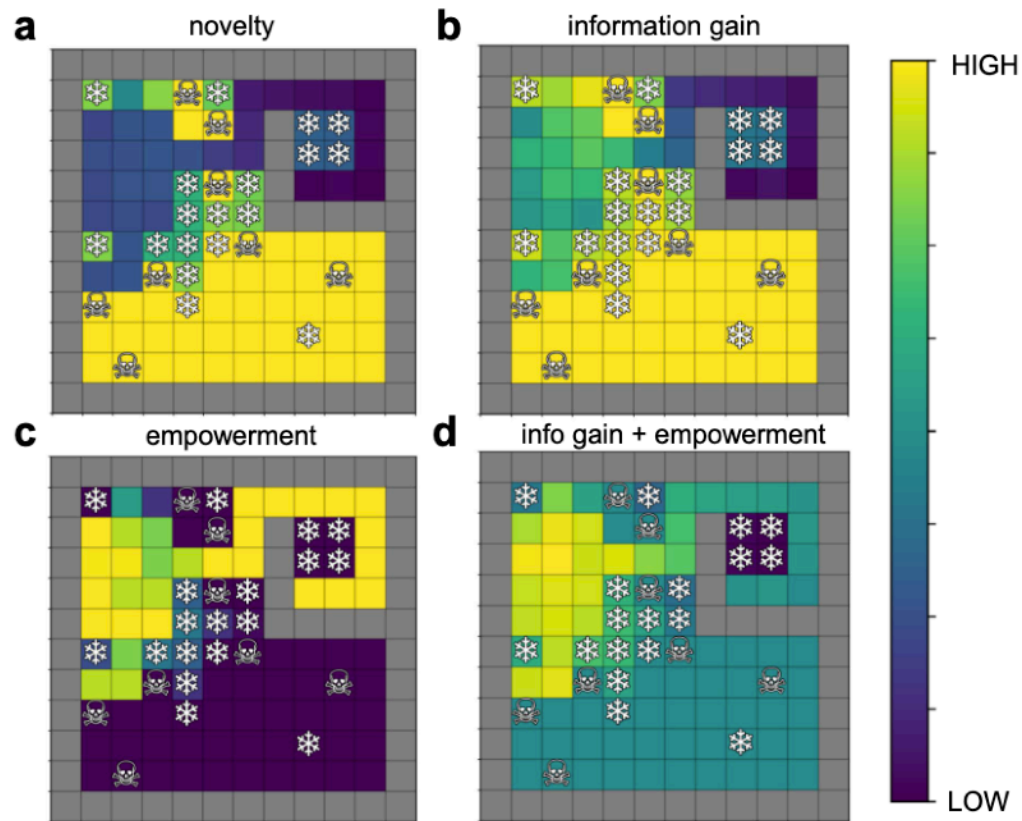
Q-learning agent: exploration effectiveness

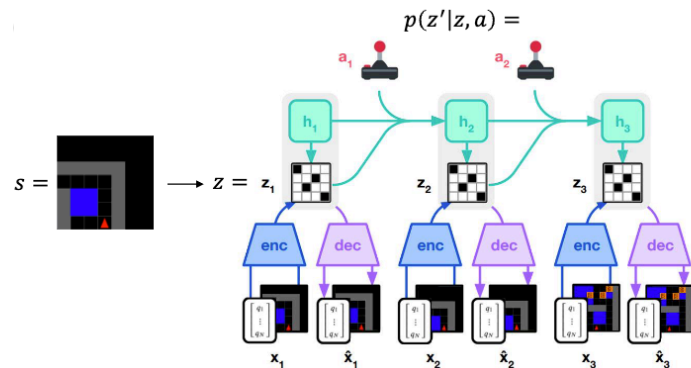
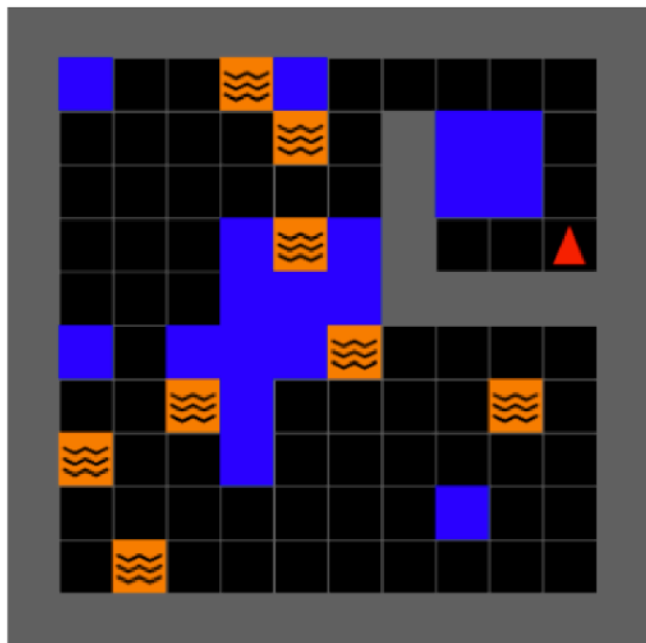


- Uniform prior
$$p(s' | s, a) = \frac{1}{|\mathcal{S}|}$$
- Maximal entropy $H(S' | s, a)$

novelty info gain empowerment info gain + empowerment

Q-learning agent: motivation heat maps

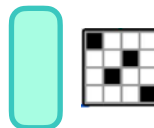


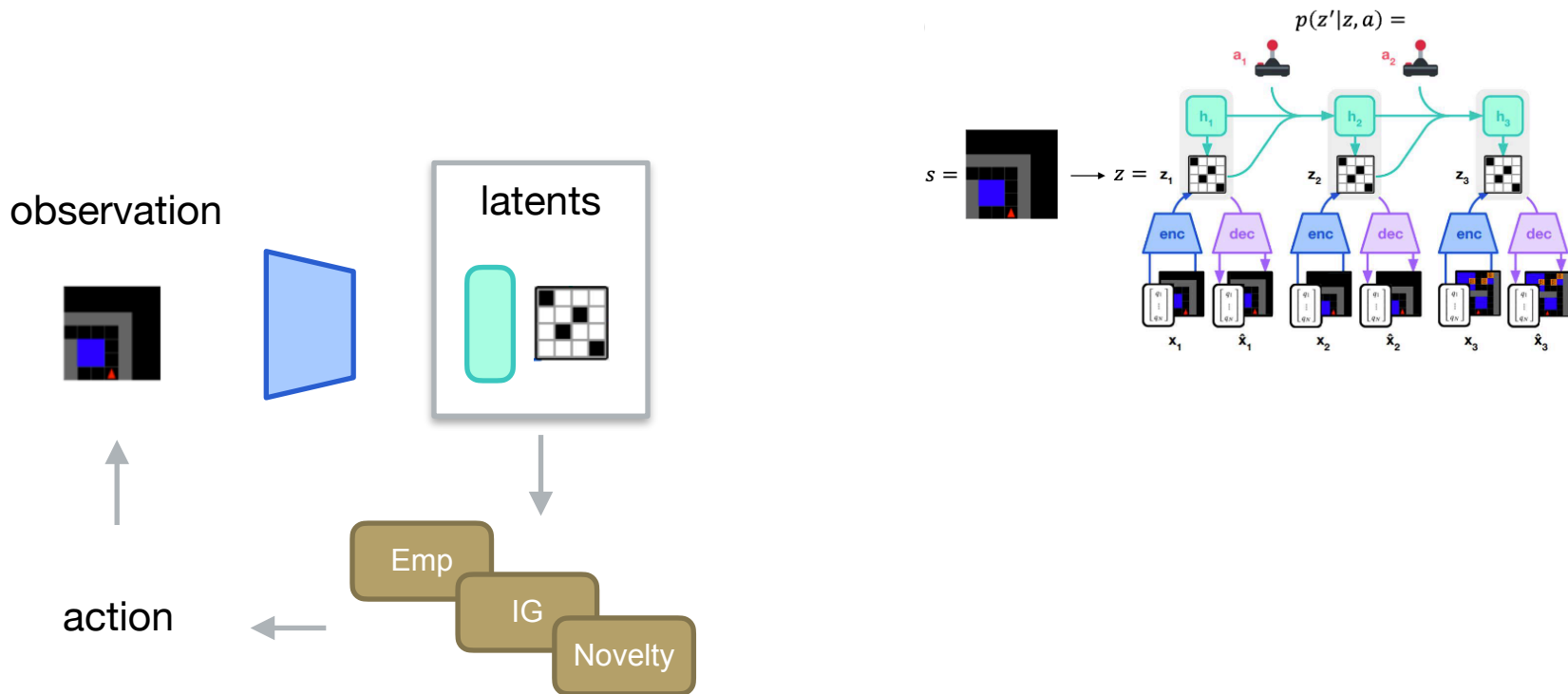


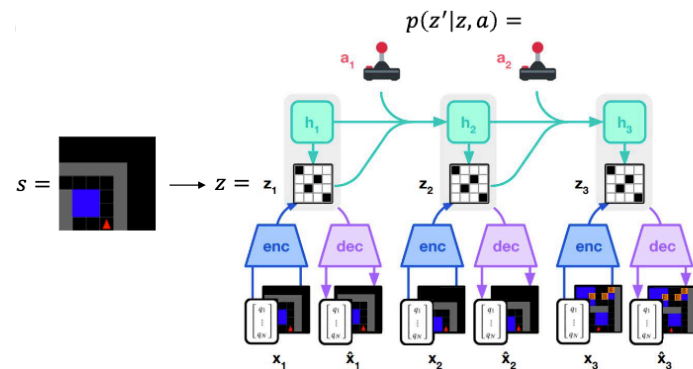
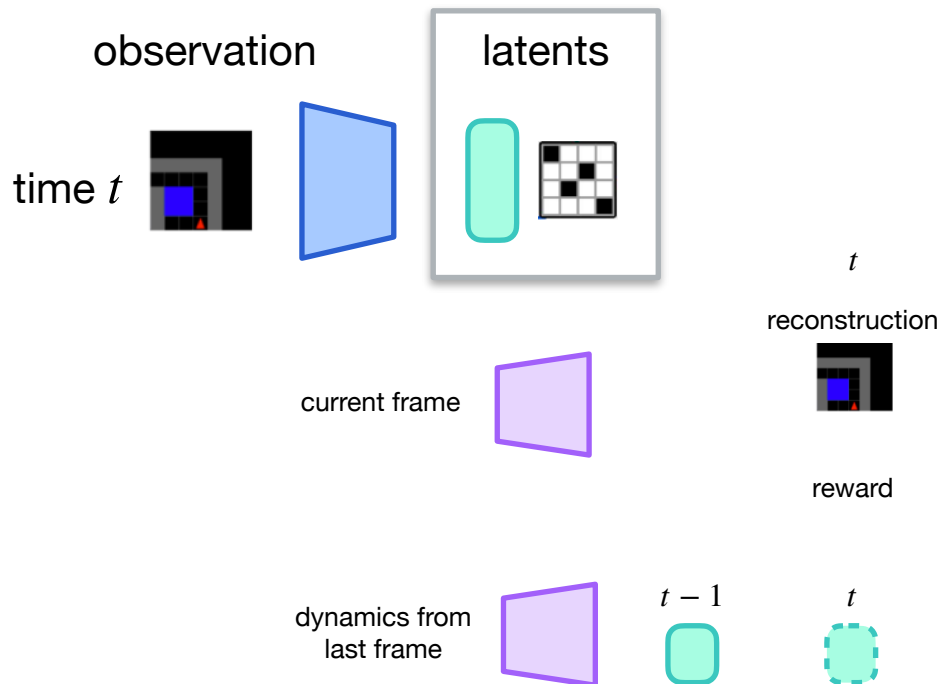
observation encoder

latents

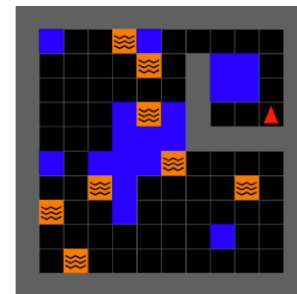
time t



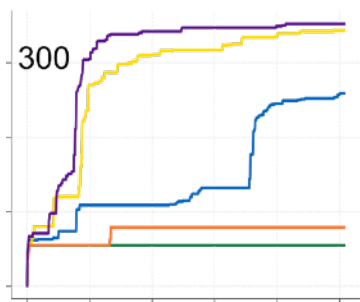




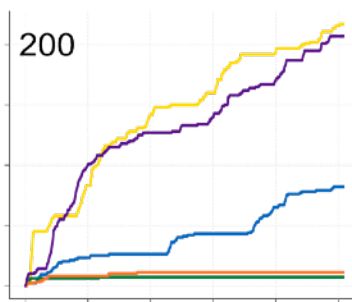
Dreamer agent: exploration effectiveness



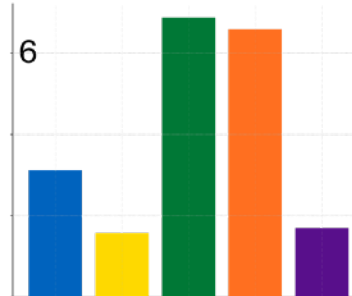
unique state discovery



cumulative deaths

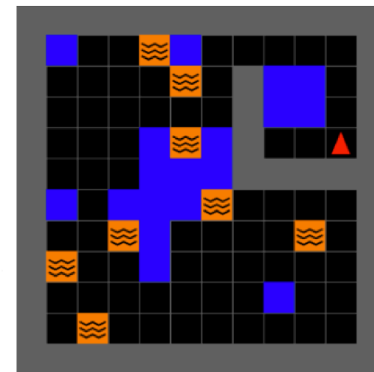
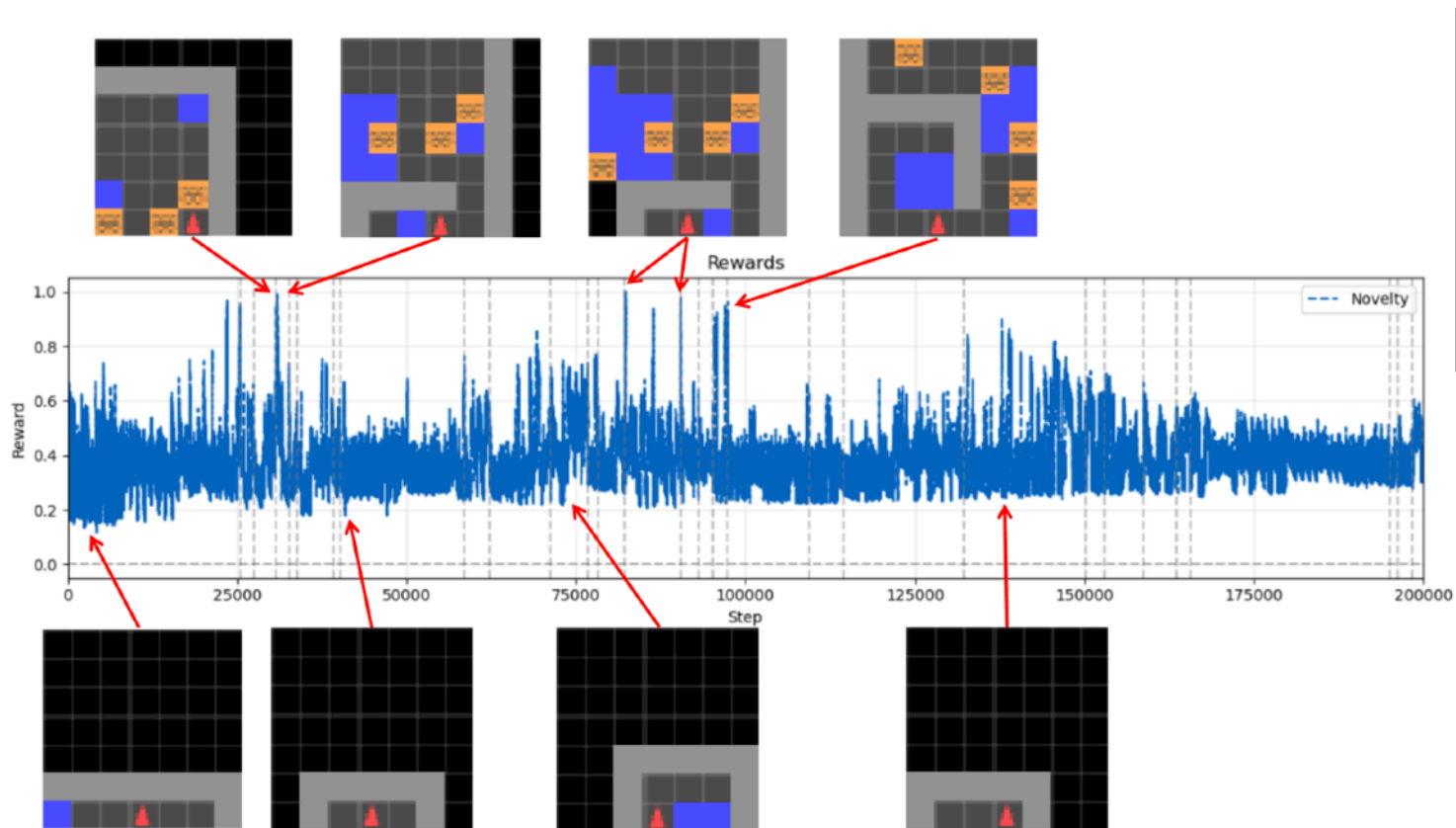


discovery-to-death ratio

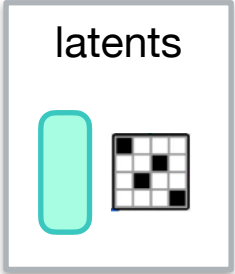
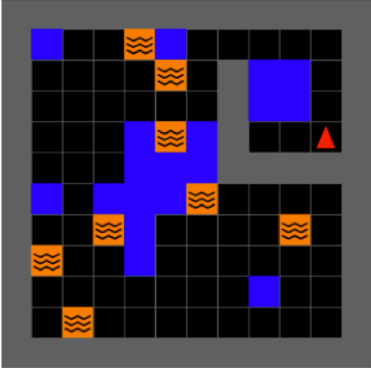


novelty info gain empowerment info gain + empowerment info gain * empowerment

Dreamer agent: novelty zoom in



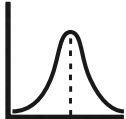
Dreamer agent: novelty gaming?



time t

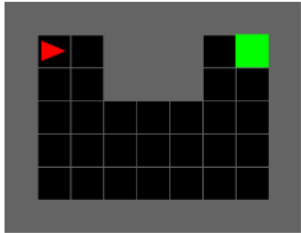


time $t - 1$

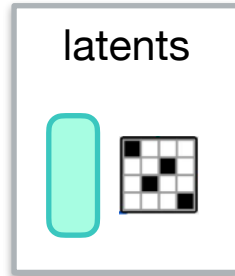
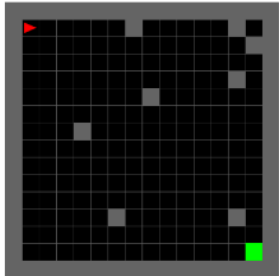


Dreamer agent: novelty gaming?

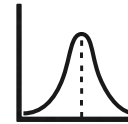
deterministic



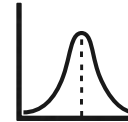
stochastic



time t

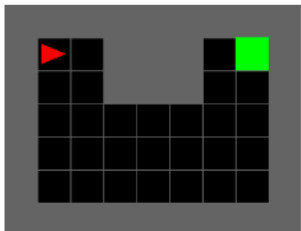


time $t - 1$

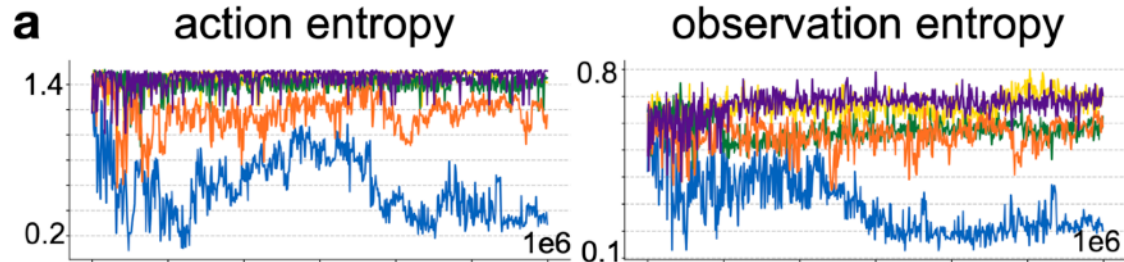
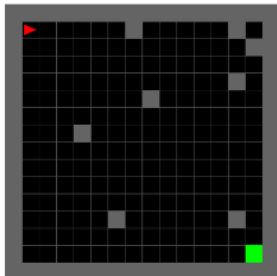


Dreamer agent: novelty gaming?

deterministic



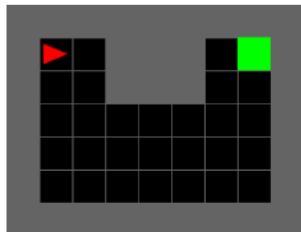
stochastic



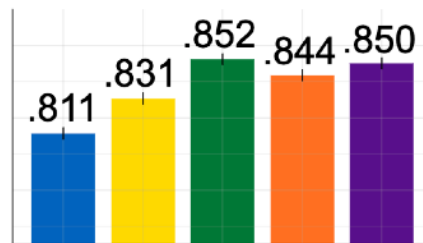
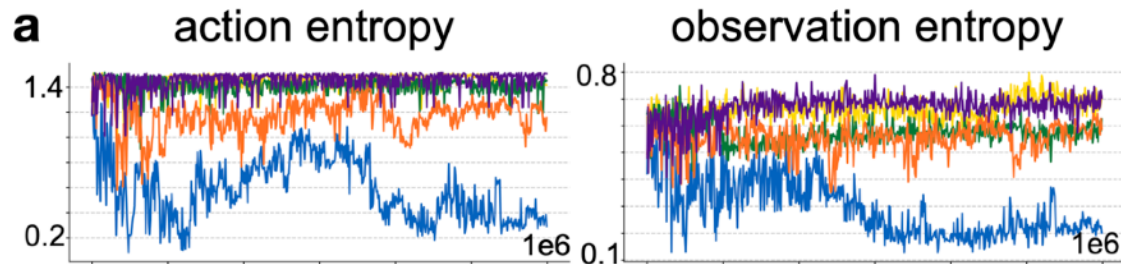
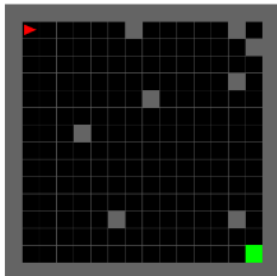
■ novelty ■ info gain ■ empowerment ■ info gain + empowerment ✖ empowerment

Dreamer agent: novelty gaming?

deterministic

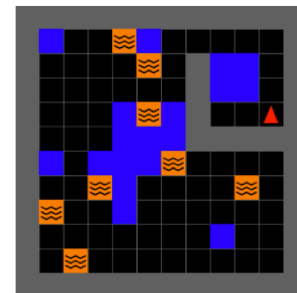


stochastic

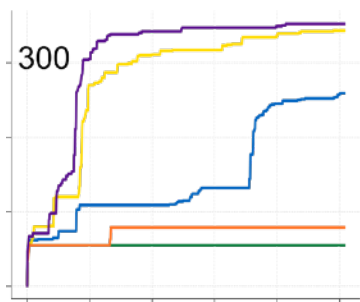


novelty info gain empowerment info gain + empowerment

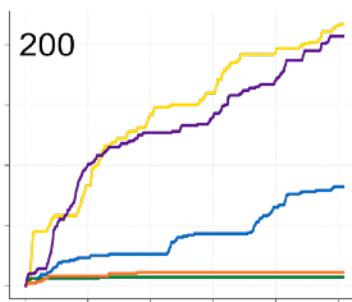
Dreamer agent: exploration effectiveness



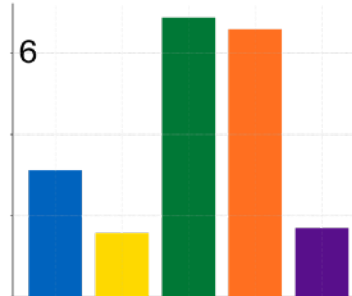
unique state
discovery



cumulative
deaths

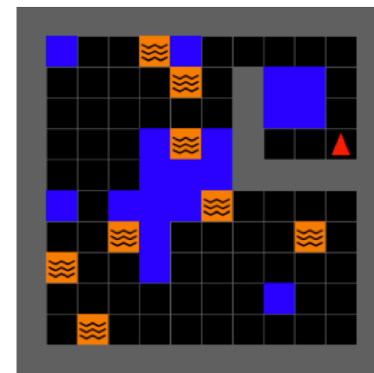
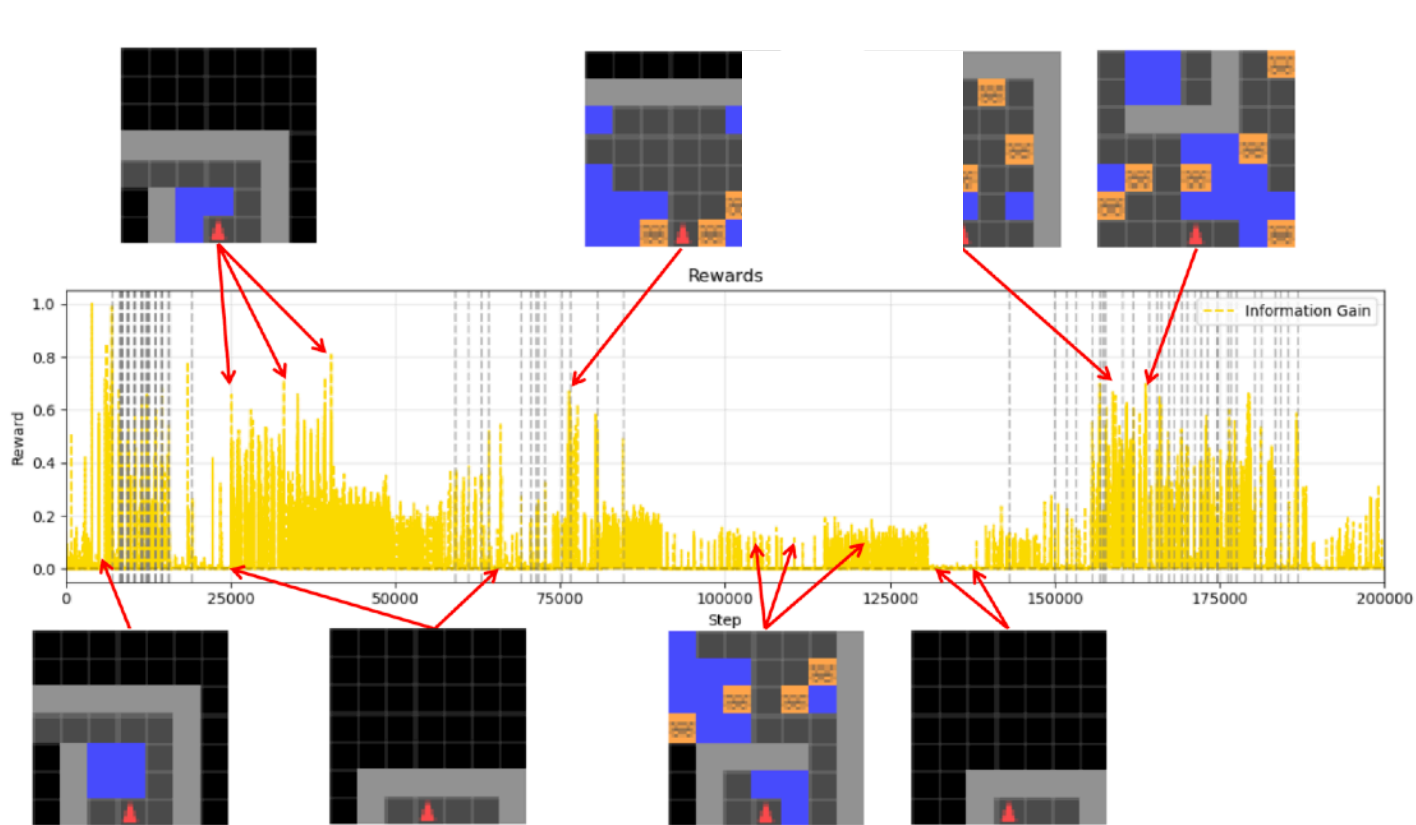


discovery-to-
death ratio

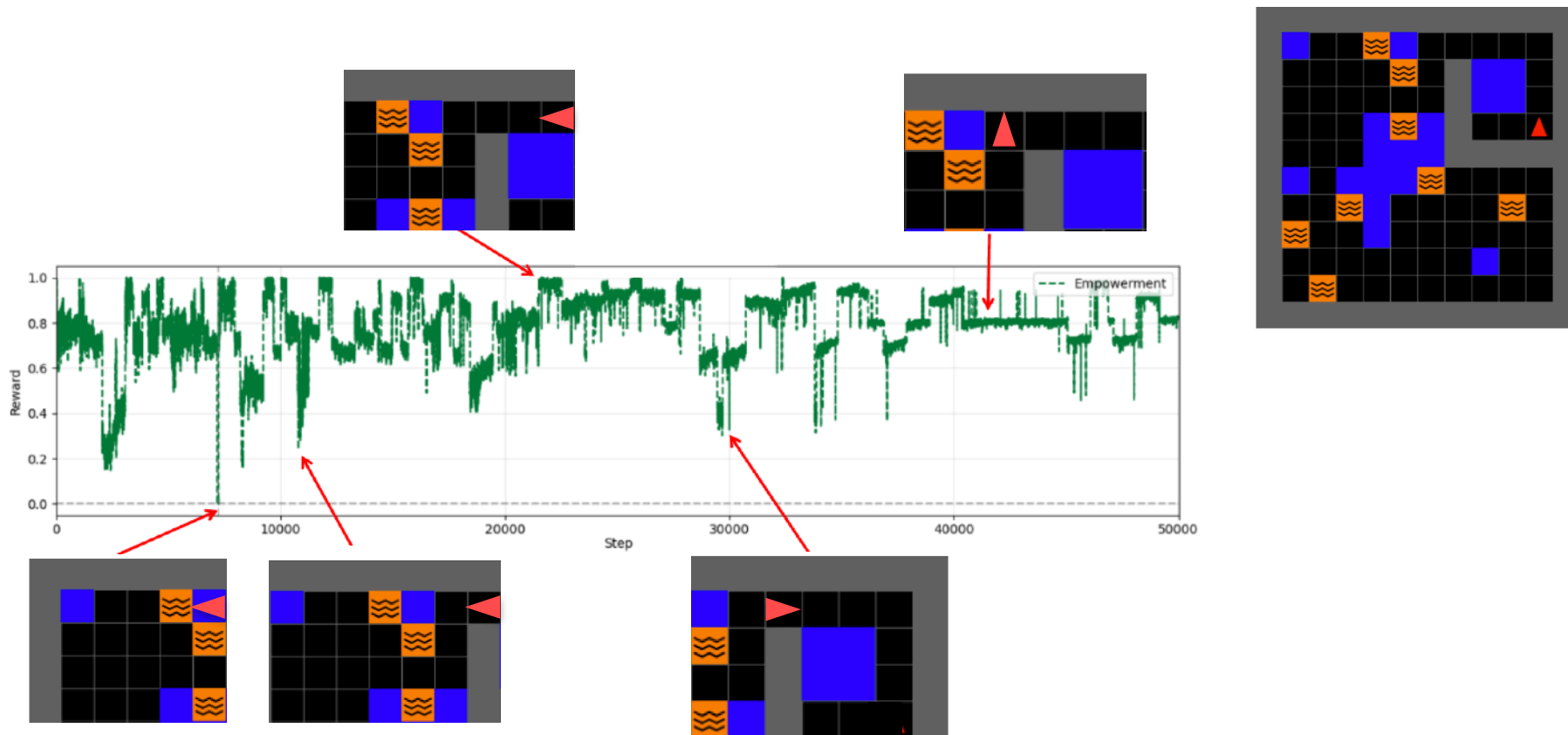


novelty info gain empowerment info gain + empowerment info gain * empowerment

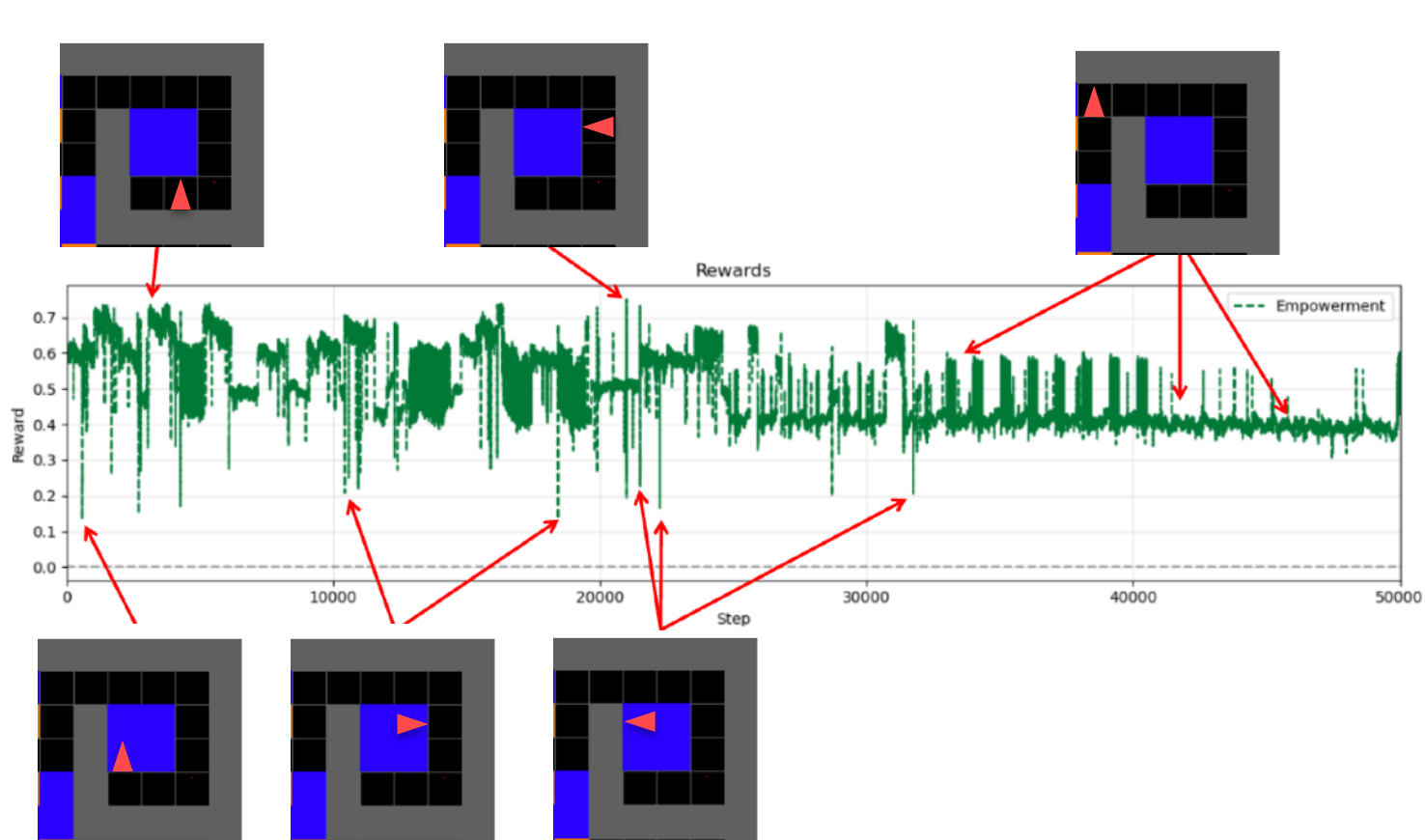
Dreamer agent: InfoGain zoom in



Dreamer agent: empowerment zoom in



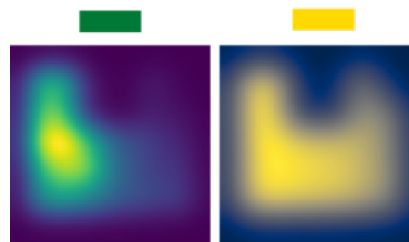
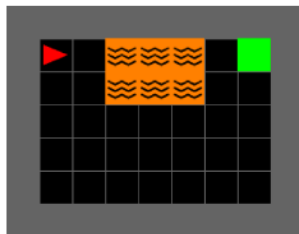
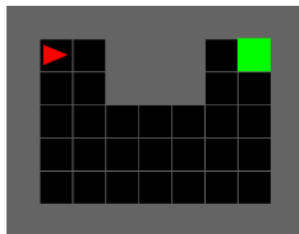
Dreamer agent: zoom in empowerment



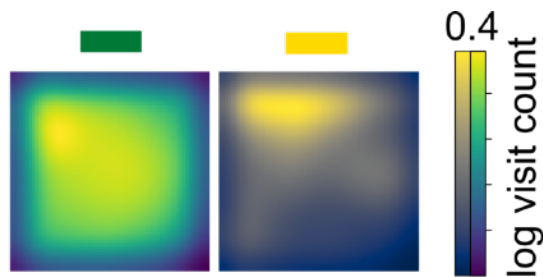
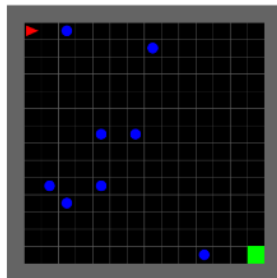
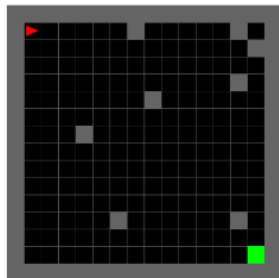
Dreamer agent: give empowerment a nudge



deterministic

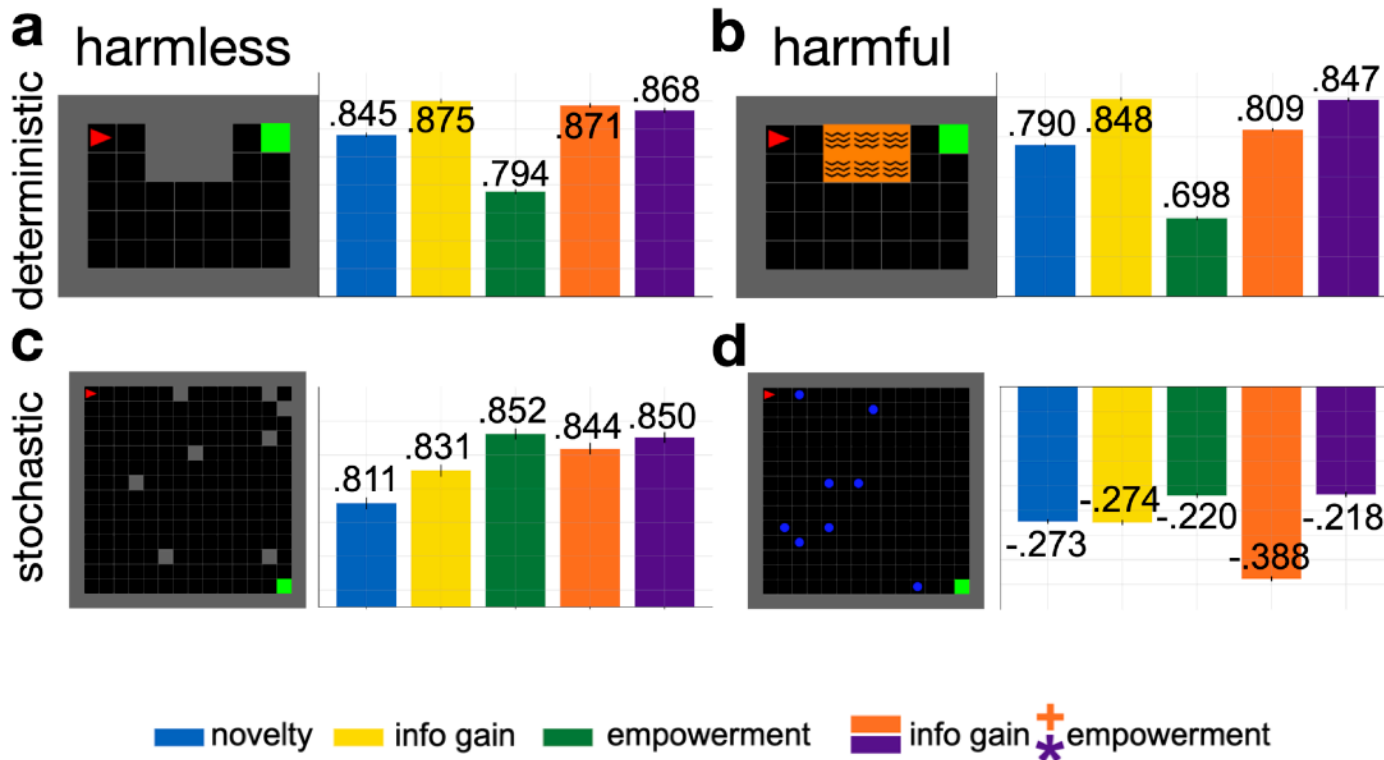


stochastic

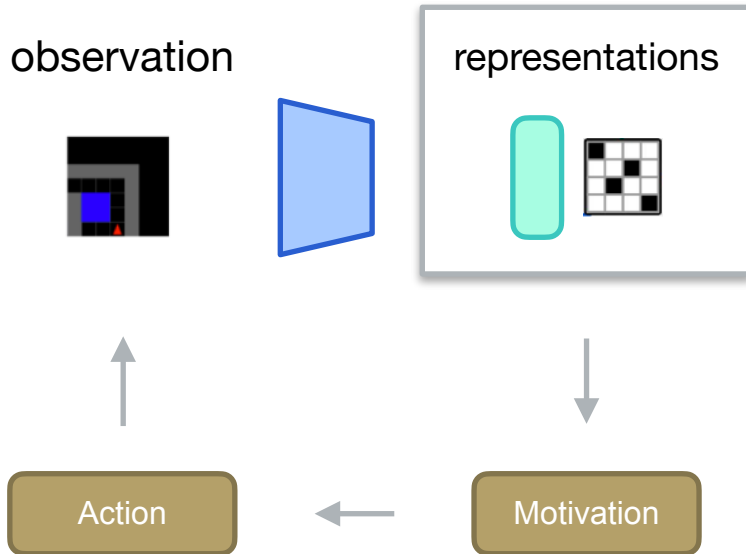


■ novelty ■ info gain ■ empowerment ■ info gain + empowerment

Dreamer agent: give empowerment a nudge

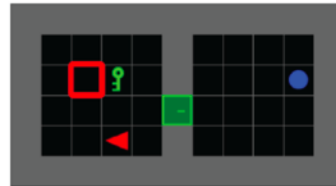
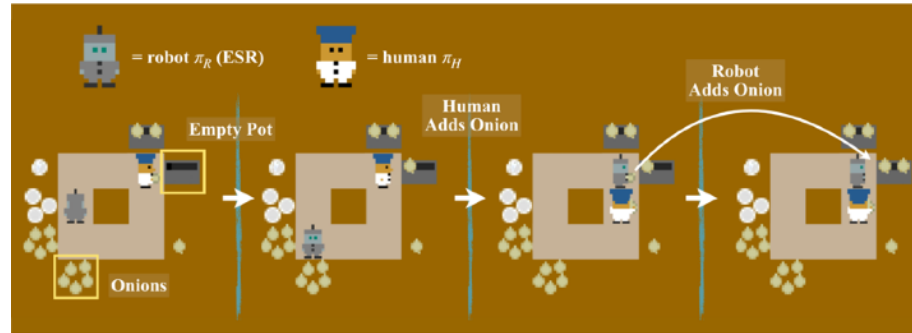
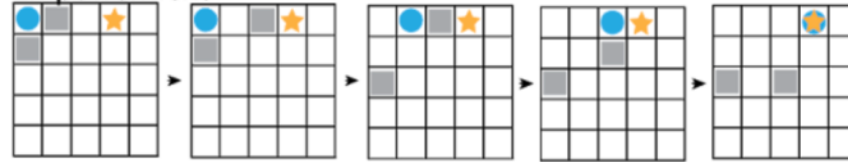


How to match human exploration?



- Each intrinsic motivation has limitations
- In the perception–action loop:
 - Agents can hack the motivation signal.
 - Yet humans don't seem to fall into these traps. Why not?
- Some possible explanations:
 - Humans operate in multi-goal, adaptive environments.
 - We switch between fast, intuitive and slow, deliberative systems.
 - We are scaffolded by teachers, core priors, and social communication.

Empowerment



Du, Y., Tiomkin, S., Kiciman, E., Polani, D., Abbeel, P., & Dragan, A. (2020). Ave: Assistance via empowerment. Advances in Neural Information Processing Systems, 33, 4560-4571.

Myers, V., Ellis, E., Levine, S., Eysenbach, B., & Dragan, A. (2024). Learning to assist humans without inferring rewards. arXiv preprint arXiv:2411.02623.

Campero, A., Raileanu, R., Küttler, H., Tenenbaum, J. B., Rocktäschel, T., & Grefenstette, E. (2020). Learning with amigo: Adversarially motivated intrinsic goals. arXiv preprint arXiv:2006.12122.