Representation learning and choice in naturalistically complex environments

Dale Zhou[1,2,3] **Shuheng Guo**[3] **Michael A. Yassa**[1,2] **Aaron M. Bornstein**[2,3]

[1] Neurobiology and Behavior, University of California, Irvine
[2] Center for the Neurobiology of Learning and Memory, University of California, Irvine
[3] Department of Cognitive Sciences, University of California, Irvine

dale.zhou@uci.edu, jianleg@uci.edu, myassa@uci.edu, aaron.bornstein@uci.edu

Abstract

Balancing exploration of uncertain options with exploiting past rewards is challenging when seemingly similar situations produce very different outcomes. An open question is how individuals pursue rewards when the reward-maximizing strategy requires implausibly detailed memory for the landscape of potential choice options. Here, we examined how humans navigate such rugged reward landscapes under limits on representational resources. Using an information-theoretic framework, we quantified *policy complexity* (mutual information between states and actions) as a measure of representational cost. Participants (n = 49) performed a naturalistic foraging task in a modified *Super Mario* environment, learning rewards across multidimensional, nonlinearly interdependent features. Behavior ranged from simple heuristics to complex state-dependent policies. Exploration induced dimensionality reduction in state representations, while greater policy complexity over selected features predicted higher reward, better memory, and improved generalization to novel stimuli. Together, these results suggest that humans dynamically balance reward maximization with representational efficiency in complex environments.

Introduction

In many decisions, we balance exploring novel options with recalling past outcomes. Efficiently doing so requires generalizing prior experience to new, similar contexts (Sims, 2018; Wu et al., 2018; Zhou & Bornstein, 2024). This is challenging in multidimensional, naturalistic environments where similar situations can lead to either high reward or costly failure (Wise et al., 2024). For example, a field mushroom and a destroying angel both look pale and unassuming, with domed caps and slender stalks, but one feeds while the other poisons. Navigating such jagged reward landscapes requires solving a trade-off between richly detailed but resource-intensive representations, or resource-efficient but partially distorted representations (Zhou et al., 2025). Humans mitigate these costs through selective attention (Mack et al., 2020), episodic memory (Nicholas & Mattar, 2026), and representation learning (Niv et al., 2015), distilling task-relevant information through dimensionality reduction.

From an information theory view, this resource allocation is captured by *policy complexity*, the information needed to specify actions in a situation (Lai & Gershman, 2021). However, it is unclear what to represent not only in *complicated* environments with many dimensions, but in *complex* ones where dimensions interact to determine reward. To address this, we develop a naturalistic reinforcement learning task that requires learning complex state–action–reward mappings (**Figure 1**).

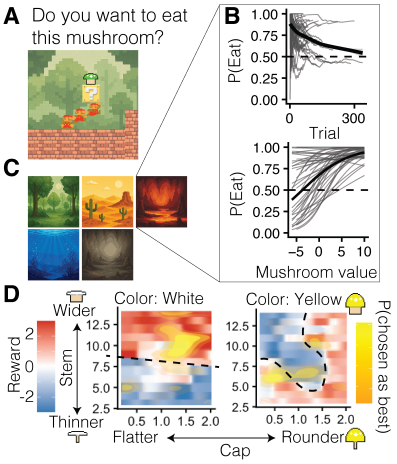

Figure 1: **(A)** Sampling a mushroom from a mystery box. **(B)** Eat vs. leave decisions over time in one world (thick line=average, thin line=individual). **(C)** Worlds contain different mushrooms (clockwise: forest, desert, lava, cave, ocean). **(D)** Reward as linear (left) vs. non-linear (right) function of stem and cap shape. Yellow heatmap of participants' estimations of the most rewarding shape.

Methods

Participants (n=49, 22 male, 27 female, mean age 39.59 $\pm$ 12.02 years) were recruited from Prolific ($20 compensation with performance-based bonus from a randomly selected part of the task). They completed a modified *Super Mario* game, a real-time 2D side-scroller with enemies and navigation-based win/lose conditions removed. Our task centered on exploration and information sampling, learning latent reward structure to seek rewarding mushrooms and avoid punishing ones. Participants revealed mushrooms by opening "mystery boxes"

Extended abstract presented at the *9th Conference on Cognitive Computational Neuroscience*, New York, NY, USA, 2026.

across worlds and had 5 seconds to decide whether to eat or leave each mushroom.

Stimulus state-reward functions. Mushrooms (n=3,200) varied along perceptual cues of taste and toxicity. Features differ across 8 colors, stem widths, and cap shapes, emphasizing color given its salience to humans when evaluating food (Macario, 1991). Each color had a distinct reward distribution (means: -2.89 to 5.06; SDs: 1.77 to 3.19). Decision boundaries vary by color, including linear (e.g. reward increases with stem width), nonlinear (intermediate widths highest), and interdependent (e.g. intermediate width rewarded only for round yellow caps; **Figure 1D**) reward functions.

Exploration across worlds. Self-paced exploration involved sampling mushrooms across five worlds (585 ± 118 trials/person), differing in average reward (0.09–1.35). All worlds contain 60% rewarding and 40% poisonous mushrooms, ensuring stationary rewards (even "eat everything" yields rewarded, albeit inefficiently). Explorers begin in each world with 20 stamina, decreasing by 0.1 per second and updated by the value of eaten mushrooms. If stamina falls to 0, there is a 5 second time-out. Reaching 30 stamina allows them to either stay at or switch worlds via three doors.

Action policy complexity. Following prior work (Gershman, 2020), we quantified representational cost via policy complexity: the mutual information $I(S;A)$ between mushroom states S and actions A during exploration. Higher $I(S;A)$ reflects more detailed and complex state-dependent mappings, whereas lower $I(S;A)$ reflects more simple and compressed state-agnostic priors or heuristics that reduce representational cost.

Memory test and similarity judgments. A two-choice task evaluates value learning and generalization by having participants pick the more rewarding mushroom (60 trials). To test generalization, one mushroom may be novel (24/60 trials). Afterward, participants reconstruct the most rewarding mushroom shape and rank colors and worlds. Participants also made similarity judgments, picking the "odd one out" among three randomly sampled mushrooms (100 trials each pre- and post-exploration).

Inferring task representations. We used methods that infer a representation best predicting similarity judgment choices via a distance-based choice function (Muttenthaler et al., 2022; Roads & Mozer, 2019).

Results

Foraging policies evolved with exploration **(Figure 1B)**. Greater policy complexity predicted higher net reward ($r = .36, p = .01$), whereas policy compression reflected an indiscriminate "eat everything" strategy ($r = .91, p < .0001$). Reward memory was more accurate by color than

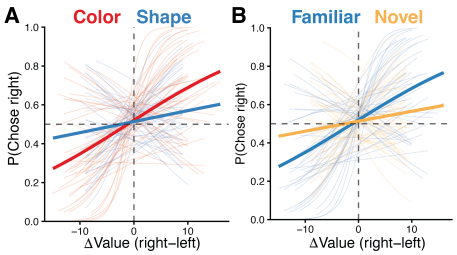

Figure 2: Choice functions from the memory test for **(A)** color vs. shape (main effect $\beta = 0.07, p < .0001$, interaction with shape $\beta = -0.05, p = .002$) and **(B)** familiar vs. novel mushrooms (main effect $\beta = 0.08, p < .0001$, interaction with novelty $\beta = -0.05, p = .003$).

shape and for familiar than novel mushrooms (**Figure 2**), with policy complexity supporting accuracy ($r = .44, p = .003$) and generalization ($r = .46, p = .002$). Participants' post-task estimates of the most rewarding shape had net positive reward (mean of 1.4 out of 5.5 maximum; one-sample $t = 11.8, p < .0001$; **Figure 1D**), color rankings matched reward (Spearman's $\rho = .41, p < .0001$), and world rankings modestly tracked reward ($\rho = .25, p < .0001$). From 9,712 similarity judgments pooled across participants (16.3% of 59,640 possible combinations), simulations showed recoverable structure $r = .56$–$.87$). Mushroom embeddings indicate sparse, interpretable color and shape dimensions (**Figure 3A**). Consistent with representation learning, task representations show reduced dimensionality after exploration (**Figure 3B**).

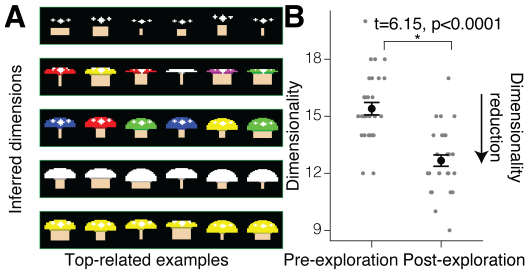

Figure 3: **(A)** Example inferred dimensions (rows) of mushroom color and shape. **(B)** Dimensionality reduction with learning. Gray circles (n=60) show inferred dimensions from bootstrap resamples (80% of trials) for pre- vs. post-exploration similarity judgments.

Discussion

Together, our findings suggest that state features are compressed while policy complexity is selectively allocated to state-action mappings that support reward, memory, and generalization. Future work will test how memory discrimination, intrinsic motivation, and different choice models influence representation learning.

Acknowledgements

The authors acknowledge funding from the George E. Hewitt Foundation for Biomedical Research (DZ) and NIMH R01MH128306 (AMB).

References

Gershman, S. J. (2020). Origin of perseveration in the trade-off between reward and complexity. *Cognition*, *204*, 104394.

Lai, L., & Gershman, S. J. (2021). Policy compression: An information bottleneck in action selection. In *Psychology of learning and motivation* (Vol. 74). Elsevier.

Macario, J. F. (1991). Young children's use of color in classification: Foods and canonically colored objects. *Cognitive Development*, *6*(1), 17–46.

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nature communications*, *11*(1), 46.

Muttenthaler, L., Zheng, C. Y., McClure, P., Vandermeulen, R. A., Hebart, M. N., & Pereira, F. (2022). Vice: Variational interpretable concept embeddings. *Advances in Neural Information Processing Systems*, *35*, 33661–33675.

Nicholas, J., & Mattar, M. G. (2026). Episodic memory facilitates flexible decision-making via access to detailed events. *Nature Human Behaviour*, 1–17.

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *Journal of Neuroscience*, *35*(21), 8145–8157.

Roads, B. D., & Mozer, M. C. (2019). Obtaining psychological embeddings through joint kernel and metric learning. *Behavior research methods*, *51*(5), 2180–2193.

Sims, C. R. (2018). Efficient coding explains the universal law of generalization in human perception. *Science*, *360*(6389), 652–656.

Wise, T., Emery, K., & Radulescu, A. (2024). Naturalistic reinforcement learning. *Trends in Cognitive Sciences*, *28*(2), 144–158.

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2018). Generalization guides human exploration in vast decision spaces. *Nature human behaviour*, *2*(12), 915–924.

Zhou, D., & Bornstein, A. (2024). Expanding horizons in reinforcement learning for curious exploration and creative planning. *The Behavioral and Brain Sciences*, *47*, e118–e118.

Zhou, D., Noh, S. M., Harhen, N. C., Banavar, N. V., Kirwan, C. B., Yassa, M. A., & Bornstein, A. M. (2025). A compressed code for memory discrimination. *bioRxiv*.