Adaptive planning and policy mixture explain the naturalistic foraging in 3D virtual reality.

Jae Young Jeon (jaeyoungjeon94@gmail.com)

Center for Neuroscience Imaging Research, Institute for Basic Science,
Suwon, 16419, Republic of Korea
Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University,
Suwon, 16419, Republic of Korea

Won Mok Shim (wonmokshim@skku.edu)

Center for Neuroscience Imaging Research, Institute for Basic Science,
Suwon, 16419, Republic of Korea
Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University,
Suwon, 16419, Republic of Korea
Department of Biomedical Engineering, Sungkyunkwan University,
Suwon, 16419, Republic of Korea

Sung Bum Michael Yoo (sbyoo@skku.edu)

Center for Neuroscience Imaging Research, Institute for Basic Science,
Suwon, 16419, Republic of Korea
Department of Biomedical Engineering, Sungkyunkwan University,
Suwon, 16419, Republic of Korea

Abstract

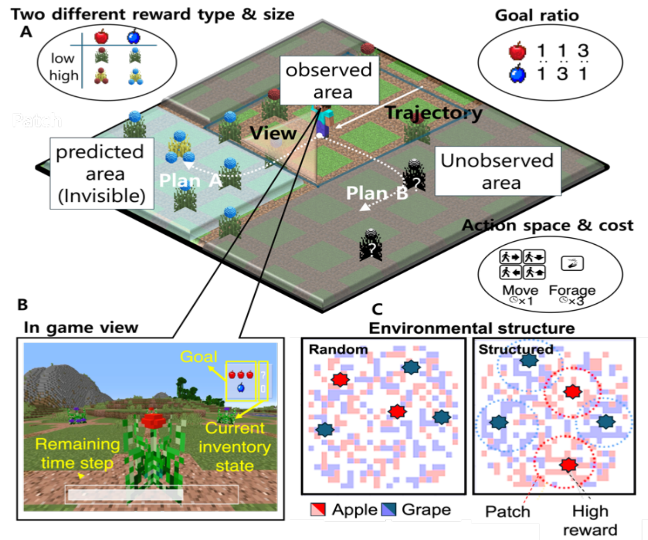
Animals and humans often reject immediate rewards by using mental models to plan for future outcomes. However, proactive rejection—intentionally skipping favorable immediate options—has been understudied due to difficulties distinguishing it from forced or reactive rejections. Using a custom-designed Minecraft-based 3D foraging task paired with a sequential Bayesian inference model, we systematically identified and characterized proactive rejection behaviors. Participants strategically increased rejection of immediate rewards as spatial regularity became more apparent, resulting in enhanced overall foraging outcomes. Our computational modeling revealed that planning depth and preference for information gathering significantly predicted rejection frequency. Crucially, proactive rejection behaviors—unlike reactive rejections—were best explained by adaptive modulation of planning depth and information prioritization based on participants' confidence in spatial regularity. These findings provide mechanistic insights into proactive rejection, highlighting its potential as a behavioral marker for goal-directed planning processes.

Keywords: Decision making, Foraging, Planning, Compound policy, Reward rejection

Task description

Naturalistic 3D foraging task

We designed a task within a 3D grid world, where rewards consisting of two different types (apples and grapes) and varying quantities (high and low) were placed at intersections. Participants were asked to maximize reward collection within a limited number of decision-making steps, while maintaining a specified collection ratio between the two reward types. Prior to reward collection, participants freely explored the environment to determine if rewards of the same type were spatially clustered (structured) or if different types were randomly dispersed (random). Structured environments contained patches (Constantino et al., 2015) with dominantly clustered rewards, including a high-value reward located centrally within these patches, which was crucial for optimal foraging outcomes. Participants were thus encouraged to develop strategies informed by the spatial regularity of rewards.

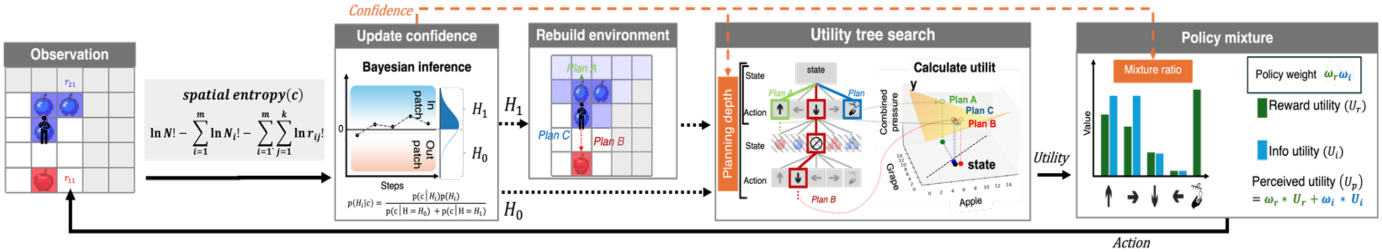


[Figure 1] (A) Task overview. Players freely navigated a 3D map using five actions (up, down, left, right, forage), with visibility limited to nine adjacent reward positions at each step. An interface displayed the current reward status and compositional goal (B) Task in game view (C) The grid map remained constant, but reward arrangement classified environments as random or structured.

Agent based generative model

Model structure

We developed an agent-based generative model with fourteen parameters to explore latent processes underlying complex decision-making during reward collection. The model takes reward maps, inventory status, and remaining steps as inputs and outputs one of five actions through a hierarchical five-step decision process. First, the agent calculates spatial regularity as spatial entropy (Zhao, 2019). Second, it updates beliefs about environmental structure using Bayesian inference (Sarafyazd et al., 2019) based on the spatial entropy calculated in the first step. Third, if the belief exceeds an individually fitted threshold and the current position is determined to be within or at the boundary of a reward patch, the agent reconstructs unseen areas. Fourth, the agent conducts multiple planning simulations via tree search within the reconstructed cognitive reward map, computing the utility of each potential plan in terms of information and reward value. Finally, actions are selected by integrating these utilities based on beliefs about environmental structure, effectively balancing immediate rewards with exploratory actions.

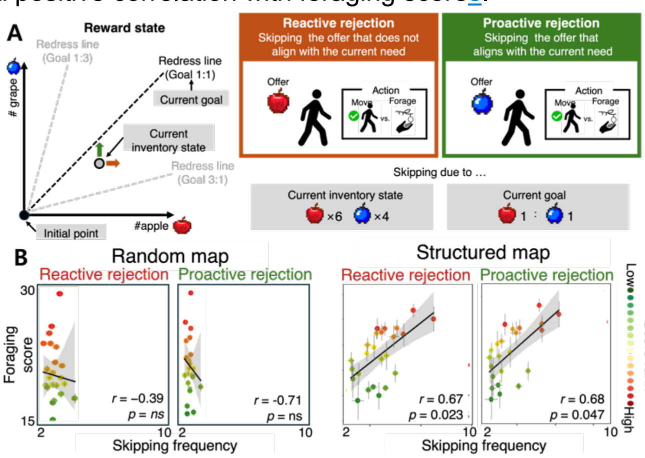


[Figure 2] Hierarchical generative model. The agent's confidence in environmental regularity modulates planning depth during tree search and policy mixture ratio between reward and information-seeking strategies.

Results

Human performance (N=24)

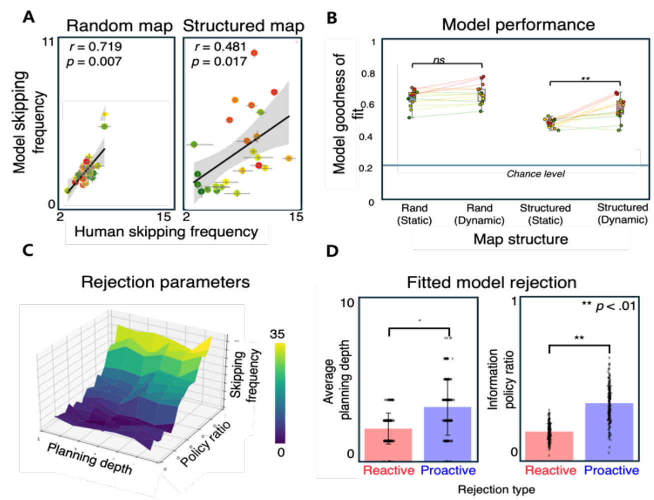
We categorized rejection behaviors into two types—reactive rejection and proactive rejection—based on the current reward status (Juechems et al., 2019) and the types of rewards encountered (Figure 3A). Reactive rejection involves skipping a surplus reward type to restore balance in the reward inventory. In contrast, proactive rejection occurs when participants intentionally skip a beneficial reward—one that could alleviate current redress pressure—to continue exploring. Participants skipped rewards more frequently in structured environments, where future rewards were spatially predictable (Fig 3B). Notably, proactive rejection showed a positive correlation with foraging scores.



[Figure 3] (A) The type of rejection is determined based on whether the given offer aligns with the goal in the current state. (B) Skipping frequency and rejection type across environmental conditions

Model performance

We assessed whether the model accurately replicated human planning behavior, which would be revealed in the rejection pattern. We found that the trajectory generated by the generative model on the same map as the human showed a high correlation in skipping frequency (Figure 4A). The generative model's performance in predicting human behavior improved only under structured environments when the planning depth and policy mixture ratio were dynamically modulated based on confidence in environmental regularity (Figure 4B). Overall skipping frequency was positively correlated with deeper planning depth and greater information policy prioritization (Figure 4C), and notably, proactive rejection showed a strong association with both latent features (Figure 4D).



[Figure 4] (A) Reward skipping comparison between humans and fitted generative model (B) Model performance across environmental conditions (C) Rejection frequency across model parameters (D) Model parameters across rejection type

Conclusion

This study demonstrates that skipping behaviors arise from flexible adjustments in planning depth and targeted prioritization of information-seeking, particularly as spatial configurations become more predictable. Proactive rejection highlights the essential role of these latent features in decision-making. The performance gain associated with these adaptive adjustments suggests that modulating planning depth and behavioral policies is advantageous in environments with varying degrees of regularity.

Acknowledgements

This work was supported by IBS-R015-D1, 24-BR-03-04, the Fourth Stage of Brain Korea 21 Project in Department of Intelligent Precision Healthcare, Sungkyunkwan University, and National Research Foundation of Korea funded by the Ministry of Education (RS-2024-00412481) & (RS-2023-00211018)

References

Constantino, S. M., & Daw, N. D. (2015). Learning the opportunity cost of time in a patch-foraging task. Cognitive, Affective, & Behavioral Neuroscience, 15, 837-853.

Zhao, Y., & Zhang, X. (2019). Calculating spatial configurational entropy of a landscape mosaic based on the Wasserstein metric. Landscape Ecology, 34, 1849-1858.

Sarafyazd, M., & Jazayeri, M. (2019). Hierarchical reasoning by neural circuits in the frontal cortex. Science, 364(6441), eaav8911.

Bermudez-Contreras, E., Clark, B. J., & Wilber, A. (2020). The neuroscience of spatial navigation and the relationship to artificial intelligence. Frontiers in Computational Neuroscience, 14, 523744.

Juechems, K., Balaguer, J., Castañón, S. H., Ruz, M., O'Reilly, J. X., & Summerfield, C. (2019). A network for computing value equilibrium in the human medial prefrontal cortex. Neuron, 101(5), 977-987.