Spatial regularity of rewards modulates reward skipping during naturalistic 3D foraging task

J. JEON[1,2], W. SHIM[1,2,3†], S.B.M. YOO[1,2,3†]

[1]Ctr. for Neurosci. Imaging Res., Inst. for Basic Sci. (IBS), Suwon, Korea, Republic of
[2]Dept. of Intelligent Precision Healthcare Convergence,
[3]Dept. of Biomed. Engin., Sungkyunkwan Univ., Suwon, Korea, Republic of

Abstract:

While economic decision-making theories typically assume that choices aimed at maximizing rewards are made between available options, foragers often reject an immediate option (i.e., skipping). We hypothesize that foragers exhibit skipping behavior when they prioritize information regarding future rewards and strategically plan ahead to encounter potentially more advantageous options. This behavior is likely to occur in environments with a level of regularity that enables predictability. To test this hypothesis, we manipulated the spatial distribution of rewards in a naturalistic harvesting task to make future rewards more predictable. Our findings show that skipping becomes more frequent as the predictability of future rewards increases. Crucially, a generative model that effectively captures human behavior suggests that skipping results from a compound action policy that prioritizes information-seeking and allows for the flexible adjustment of planning depth. The increased frequency of skipping, which leads to better performance selectively in predictable environments with spatial regularity, indicates that skipping behavior is an adaptive decision to maximize rewards in predictable environments.

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

Introduction

Skipping imminent rewards is evidence of the ability to exploit the environmental structure (Hayden, 2014). However, it remains unclear what underlying cognitive processes contribute to skipping behaviors. In this study, using an environment with spatial regularity, we designed a generative model that aligns with human foraging behaviors and examined the effect of its parameters, particularly planning depth and information-seeking policy, on skipping.

Task description

Naturalistic 3D orchard harvest task

We constructed a task within a 3D grid world, where rewards, comprising two different types (apples and grapes) and quantities (high and low), were positioned at intersections. Participants were asked to maximize reward collection within a limited number of decision-making steps while adhering to a specified collection ratio of each reward type. Prior to collecting rewards, participants determined whether rewards of the same type were spatially clustered (structured) or if different types of rewards were randomly dispersed (random) through free observation. In the structured environment, the center of a patch contained high-value rewards, which were crucial for maximizing foraging outcomes, and participants were encouraged to develop strategies aligned with the spatial regularity of rewards.

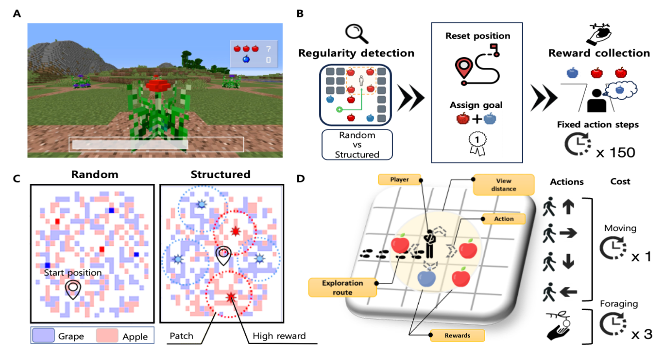

Fig 1 (A) Example view of the task. (B) Two-step task structure: regularity detection phase and reward collection phase. (C) Example of environmental structures. (D) Task overview. The player's abstract action space consisted of five actions (moving up, down, left, right, and foraging) at different temporal costs. Players were able to observe the nine (3x3) adjacent reward positions at each state.

Agent based foraging model

Model structure

We designed an agent-based model with ten free parameters that exhibit adaptive foraging behavior similar to that of participants (Fig 2). The model took the observed map, remaining action steps, and the status of reward collection as its input. Then, it generated hierarchical actions at each choice: initially deciding to forage or move, then determining the direction of action if the agent decided to move. The process involves four sequential modules: 1) reward reconstruction based on state belief, 2) tree search simulation, 3) value calculation, and 4) policy blending. At every step, the regularity of rewards was updated and transformed into a measure of clusteredness (Fig 2), which served as an indicator of the spatial regularity of rewards. We quantified it as the spatial configurational entropy based on Wasserstein matrices (Zhao et al., 2019). A high clusteredness indicated an increased predictability of future states, prompting the model 1) to extend the depth of tree search simulations and 2) to prioritize policies that aim to maximize environmental information over those that seek to maximize reward.

Reward representation

In the model, the action in each step was determined by the reward value and information value. The reward value was defined by the extent to which it reduced the combined pressure of reward compositionality and goal. It was represented as a weighted sum of the Manhattan distance between the current state within the reward map (Summerfield, 2019) and the redress line formed by the given goal of the reward ratio and the optimal goal point.

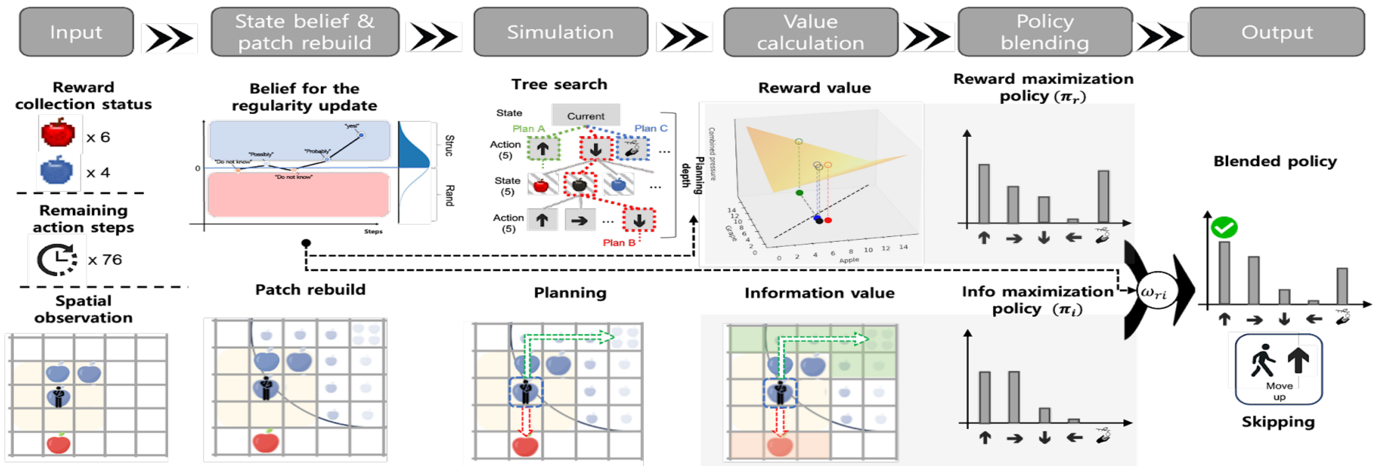

Fig 2 Model structure. The model computes beliefs about the environment through external states and subsequently extracts multiple plans inside the rebuilt map via simulation. In the final step, it blends information and internal reward policies with weights based on these beliefs to derive an action that maximizes internal value.

On the other hand, the information value increased proportionally to the area of newly explored regions (Bermudez-Contreras, 2020) and the increase in clusteredness of the current goal reward type achieved through the action.

Results

Participants exhibited a higher frequency of skipping in environments with a structured layout, where future rewards were more predictable (Fig 3A). Notably, there was a positive correlation between the frequency of skipping and foraging scores selectively in the spatially structured environment (Fig 3B).

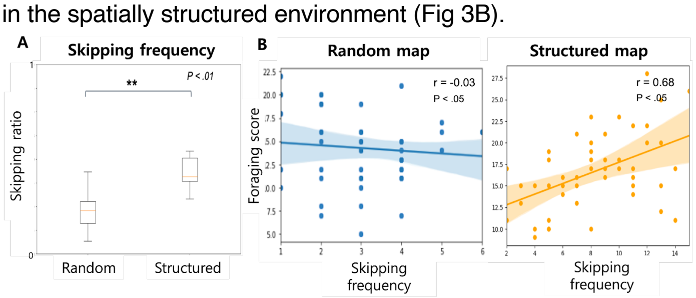

Fig 3 Participants' skipping frequency and foraging scores in random and structured maps.

To construct an agent that mimics the foraging behavior and generates skipping action patterns observed in human subjects, we fitted the model's ten parameters to each individual subject (N=14).

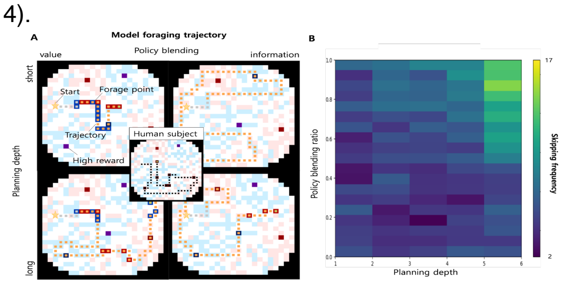

Fig 4 (A) Example foraging trajectory of a fitted model agent compared to that of a human participant. (B) Skipping frequency across model parameters.

The model agent generated different foraging trajectories depending on the planning depth (i.e., tree simulation depth) and policy blending ratio, which is a weight assigned to each policy, and variations in the skipping pattern were observed accordingly (Fig 4A). A deeper planning depth and a stronger tendency towards information seeking were associated with more proactive reward-skipping behaviors (Fig 4B).

Human subjects did not merely increase their reward skipping frequency; instead, they engaged in strategic skipping to achieve efficient foraging. This behavior was accomplished by adjusting planning depth and policy blending ratio, which were manipulated based on their beliefs about the spatial regularity of rewards. These beliefs were calculated through Bayesian inference that takes perceived clusteredness from a reward map as input, and continually updates the participant's beliefs about the environmental structure based on recent observations. Indeed, the model that adjusted latent behavioral variables based on spatial regularity demonstrated high predictive accuracy for human decision-making, not only in random maps but also in structured maps (Fig 5A).

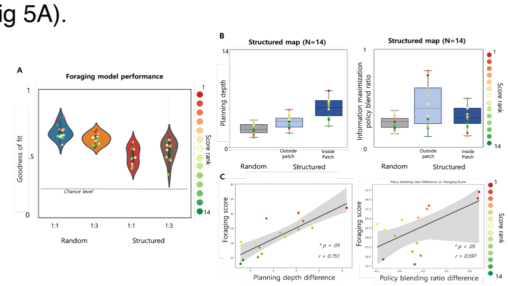

Fig 5 (A) Model performance showing the goodness of fit between the blended policy and human behavior. (B) Planning depth and policy blending ratio modulated through updated belief of spatial regularity of rewards. (C) Correlation between degree of adjustment for latent behavior variables and foraging score

Analyzing the dynamics of parameters fitted to each individual subject at every step revealed that subjects who obtained higher foraging scores tended to adjust their planning depth and policy blending ratio more sensitively in response to the spatial regularity of rewards. In contrast, subjects with lower foraging scores consistently adhered to a greedy strategy, irrespective of environmental beliefs (Fig 5B). This suggests that flexible strategy shifting based on spatial information is linked to strategic reward-skipping actions and higher foraging outcomes (Fig 5C).

Conclusion

This study reveals that skipping behaviors arise from the flexible adjustment of planning depth and prioritization of information-seeking when the spatial configuration becomes more predictable. The increased performance associated with the flexible adjustment of these parameters, which results in skipping, suggests that adjusting planning depth and behavioral policy is an adaptive strategy in an environment with varying degrees of regularity.

Acknowledgments

This work was supported by IBS-R015-D1 (W.M.S. and S.B.M.Y.), RS-2023-00211018 (S.B.M.Y.), 24-BR-03-04 (W.M.S.) and the Fourth Stage of Brain Korea 21 Project in Department of Intelligent Precision Healthcare, Sungkyunkwan University (W.M.S.)

References

Hayden, B. Y., & Walton, M. E. (2014). Neuroscience of foraging. Frontiers in neuroscience, 8, 87581.

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.

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.

Hill, J. A. C. (1983). A computational model of language acquisition in the two-year old. *Cognition and Brain Theory*, *6*, 287-317.

Matlock, T. (2001). *How real is fictive motion?* Doctoral dissertation, Psychology Department, University of California, Santa Cruz.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Ohlsson, S., & Langley, P. (1985). *Identifying solution paths in cognitive diagnosis* (Tech. Rep. CMU-RI-TR-85-2). Pittsburgh, PA: Carnegie Mellon University, The Robotics Institute.

Shrager, J., & Langley, P. (Eds.) (1990). *Computational models of scientific discovery and theory formation*. San Mateo, CA: Morgan Kaufmann.