State-dependent Online Reactivations for Different Learning Strategies in Virtual Foraging

Sangkyu Son (ss.sangkyu.son@gmail.com)
Department of Biomedical Engineering, Sungkyunkwan University
Suwon, Republic of Korea

Maya Zhe Wang (mayawangz@gmail.com)
National Institute of Mental Health
Bethesda, Maryland, U.S.

Benjamin Y. Hayden (benhayden@gmail.com)
Department of Neurosurgery, Baylor College of Medicine
Houston, Texas, U.S.

Seng Bum Michael Yoo (sbyoo@g.skku.edu)
Department of Biomedical Engineering, Sungkyunkwan University
Suwon, Republic of Korea

Abstract

Learners often seek multiple objectives paired with different strategies, and the brain needs to (re)activate representation fit for each objective. We questioned if neural reactivation, typically viewed as stereotyped during the learning process, is the subject of control to promote different types of navigational learning. We trained macaques to forage in a first-person virtual maze. Two behavioral repertories emerged from the low-level features; one seeks reward (explore-like) and the other for information (exploit-like). While alternating among two objectives, the orbitofrontal (OFC) and retrosplenial cortices (RSC) preplayed the future optimal path and goal itself, specifically when prioritizing reward. When prioritizing information, both cortices strategically devalued the uninformative paths with reduced reactivation. Meanwhile, the reactivation of the fastest path that leads to the goal was reinforced when prioritizing reward. The artificial agent foraging in the identical maze confirmed that RSC and OFC devaluating the uninformative path and reinforcing the reward-optimal path promotes the reward rate. These results highlight that neural ensemble adaptively aids the learning process as per the need of each moment.

Keywords: virtual-reality navigation, retrosplenial cortex, orbitofrontal cortex, neural reactivation, preplay, multi-learning strategies

Results

We trained two macaques to navigate the first-person virtual reality maze using a joystick (**Figure 1A**, top). The monkey began each trial in a random location and was rewarded if reached the jackpot goal location (which was fixed within each session). For each choice point (i.e., junctions), twelve observable low-level behavioral features were quantified (**Figure 1A**, bottom; head turn speed, head turn consistency, residence time at the choice point, speed at, 500 ms before, and after the choice point, joystick press strength at, 500 ms before, and after the choice point, and acceleration at, 500 ms before, and after the choice point).

Two latent learning strategies

While encountering 4,549 choice points per subject on average, we found that the naturalistic navigational behavior fell into two clusters, visually in t-distributed stochastic neighbor embedding (tSNE) space (**Figure 1B**) and quantitatively in K-mean clustering algorithm performance (**Figure 1C**; maximum performance at K=2). In state 1 (referred to as *exploit-like*), the subjects' choice resulted in the more optimal path, meaning it led closer to the goal location (**Figure 1D**). In state 2 (referred to as *explore-like*), on the other hand, the choice was on the less visited route (i.e., the more informative path). Over trials, the portion of exploit-like behavior increased while the explore-like behavior decreased (**Figure 1E**). This together indicates that while subjects learned the maze, they naturally alternated between two strategies – prioritizing immediate reward and broadening the knowledge of spatial structure.

Preplay of goal and goal-directed path

To understand the neural basis of each learning strategy, we recorded the retrosplenial cortex (RSC; 655/738 neurons for each subject, respectively) and the orbitofrontal cortex (OFC; 581/928 neurons for each subject), simultaneously. We investigated whether the neuronal ensemble encoding the goal location was reactivated state-conditionally (goal reactivation) as well as reactivation of the path between the goal location and the current choice point (path reactivation). The goal reactivation was measured by the correlation

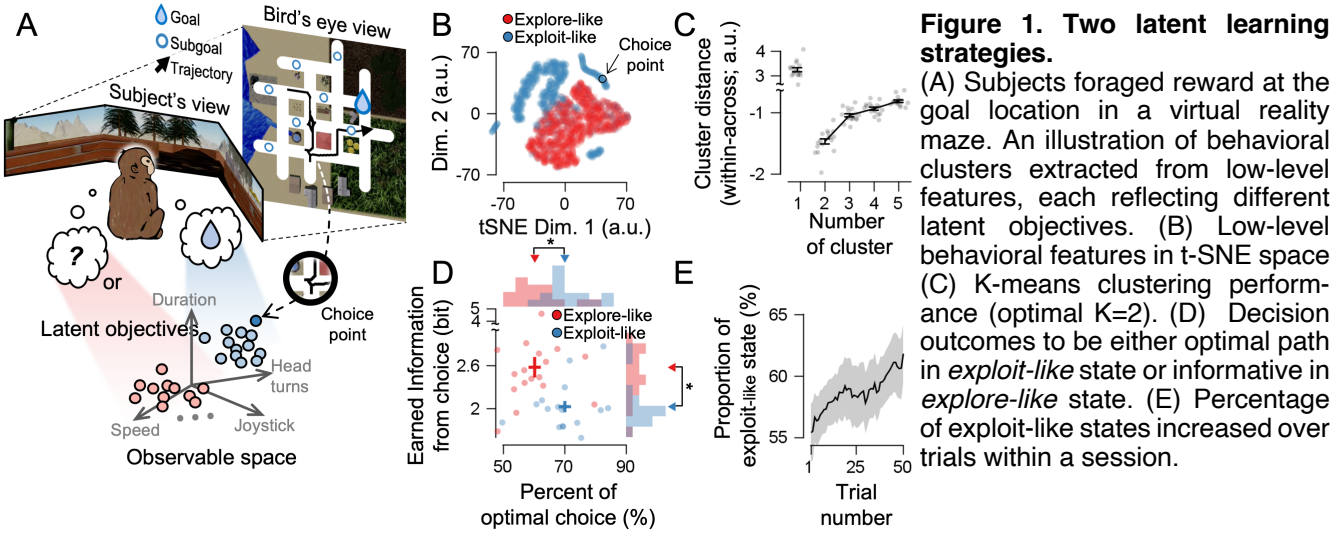

Figure 1. Two latent learning strategies.
(A) Subjects foraged reward at the goal location in a virtual reality maze. An illustration of behavioral clusters extracted from low-level features, each reflecting different latent objectives. (B) Low-level behavioral features in t-SNE space (C) K-means clustering performance (optimal K=2). (D) Decision outcomes to be either optimal path in *exploit-like* state or informative in *explore-like* state. (E) Percentage of exploit-like states increased over trials within a session.

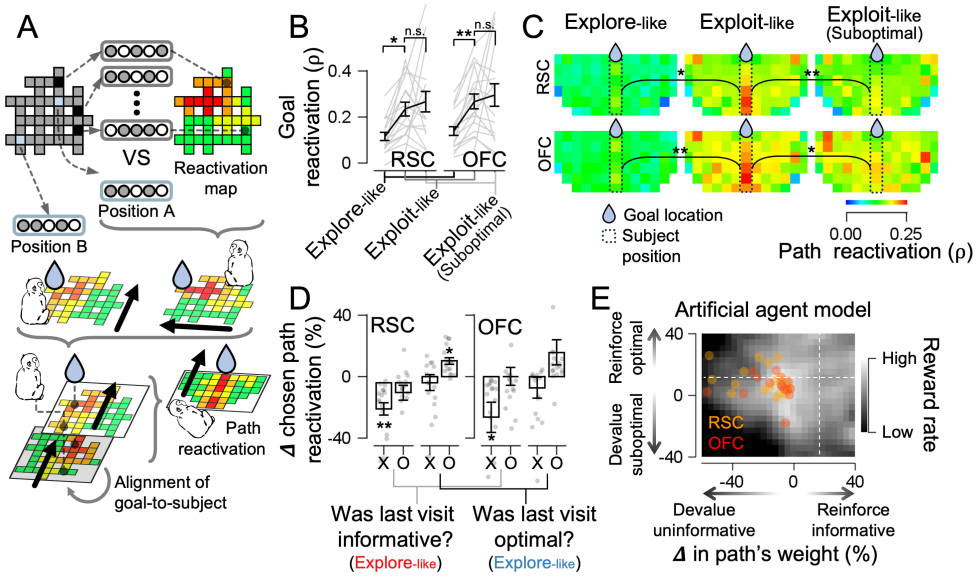

Figure 2. Reactivation of RSC and OFC supports distinct learning strategies. (A) The reactivation maps (firing pattern comparison between one position and the others) were rotated to align the goal location to different positions. (B) Reactivation of goal location (C) Reactivation of paths from current corner to goal (D) The change of path reactivation, conditioned by informativeness or optimality accomplished from the previous visit. (E) The reward rate of artificial agents with various combinations of strategies in the same maze.

of the firing rate at the goal location with those at choice points. To compute path reactivation, the firing pattern at the current position and all the other locations were compared and aligned into one direction, anchoring the goal location (**Figure 2A**).

We found that in RSC and OFC, the neural pattern of the goal and the path to it were reactivated far stronger when the subject's strategy was on prioritizing the reward (exploit-like state) (**Figure 2B** and **C**; left two columns). Moreover, the degree of reactivation reflects the subject's choice behavior; suboptimal choices—selecting paths that do not direct to rewards—exhibited goal reactivation levels comparable to those during optimal choices, yet they showed diminished path reactivation (**Figure 2B** and **C**; right two columns). Together, preplay-like results demonstrate that different representations can be triggered and recapitulated contingent upon the specific learning objectives.

Distinct learning strategies in reactivation

We further investigated if the extent to which each learning objective is achieved alters the reactivation strength. Indeed, we observed that the path reactivation changed depending on the informativeness and optimality of the last visit's choice (**Figure 2D**); the reactivation of the previously visited (uninformative) path was decreased after explore-like learning in both RSC and OFC, while reactivation of the optimal path was increased after exploit-like learning in RSC. The differential change contrasts the devaluation-based learning in the explore-like state with the conventional reinforcement-based learning in the exploit-like state.

Finally, we simulated an artificial agent foraging in the identical maze with path choices among options based on the weight updates (**Figure 2E**). When enforcing various combinations of devaluation (decreasing path's weight) and reinforcement strategies (increasing path's weight), we found that the changes in RSC and OFC's largely overlapped with the highest reward rate combinations. The simulation confirms that the distinctive two learning patterns reflected in RSC and OFC's reactivation were tightly associated with reward optimization.

Discussion

The current study supports the hypothesis that each learning process controls the distinct types of neural reactivations in service of the need of the moment. This challenges the traditional notions of neural reactivation associated with learning as a fixed, stereotyped form into a multi-faceted one for distinctive learning objectives, which may include curiosity (Kidd & Hayden, 2015; Poli et al., 2024) and empowerment (Brändle et al., 2023; Klyubin et al., 2005).

Moreover, although OFC's function is typically confined to value-based decision-making, the preplay of the spatial layout shown in the current result supports the notion of OFC's involvement in navigational planning (Basu et al., 2021; Maisson et al., 2023; Wikenheiser et al., 2021).

As a growing number of studies discover repertories even in complex naturalistic behaviors (Berman et al., 2016; Pereira et al., 2019; Voloh, Maisson, et al., 2023), the current study points out that even unconstrained behaviors, outside the simple laboratory tasks, can be bottom-up ethogrammed in line with various cognitive process, opening up the richness of naturalistic cognition (Yoo et al., 2021, 2020).

Acknowledgments

This research was supported by IBS-R015-D1, RS-2023-00211018, MH129439, and R01 MH125377.

References

Basu, R., Gebauer, R., Herfurth, T., Kolb, S., Golipour, Z., Tchumatchenko, T., & Ito, H. T. (2021). The orbitofrontal cortex maps future navigational goals. Nature, 599(7885), 449–452. https://doi.org/10.1038/s41586-021-04042-9

Berman, G. J., Bialek, W., & Shaevitz, J. W. (2016). Predictability and hierarchy in Drosophila behavior. Proceedings of the National Academy of Sciences of the United States of America, 113(42), 11943–11948. https://doi.org/10.1073/pnas.1607601113

Brändle, F., Stocks, L. J., Tenenbaum, J. B., Gershman, S. J., & Schulz, E. (2023). Empowerment contributes to exploration behaviour in a creative video game. Nature Human Behaviour. https://doi.org/10.1038/s41562-023-01661-2

Kidd, C., & Hayden, B. Y. (2015). The Psychology and Neuroscience of Curiosity. Neuron, 88(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Klyubin, A. S., Polani, D., & Nehaniv, C. L. (2005). Empowerment: a universal agent-centric measure of control. 2005 IEEE Congress on Evolutionary Computation, 1, 128-135 Vol.1. https://doi.org/10.1109/CEC.2005.1554676

Maisson, D. J.-N., Cervera, R. L., Voloh, B., Conover, I., Zambre, M., Zimmermann, J., & Hayden, B. Y. (2023). Widespread coding of navigational variables in prefrontal cortex. Current Biology: CB, 33(16), 3478-3488.e3. https://doi.org/10.1016/j.cub.2023.07.024

Manea, A. M. G., Maisson, D. J.-N., Voloh, B., Zilverstand, A., Hayden, B., & Zimmermann, J. (2024). Neural timescales reflect behavioral demands in freely moving rhesus macaques. Nature Communications, 15(1), 2151. https://doi.org/10.1038/s41467-024-46488-1

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., & Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. Nature Methods, 16(1), 117–125. https://doi.org/10.1038/s41592-018-0234-5

Poli, F., O'Reilly, J. X., Mars, R. B., & Hunnius, S. (2024). Curiosity and the dynamics of optimal exploration. Trends in Cognitive Sciences. https://doi.org/10.1016/j.tics.2024.02.001

Voloh, B., Eisenreich, B. R., Maisson, D. J.-N., Ebitz, R. B., Park, H. S., Hayden, B. Y., & Zimmermann, J. (2023). Hierarchical organization of rhesus macaque behavior. Oxford Open Neuroscience, 2. https://doi.org/10.1093/oons/kvad006

Voloh, B., Maisson, D. J.-N., Cervera, R. L., Conover, I., Zambre, M., Hayden, B., & Zimmermann, J. (2023). Hierarchical action encoding in prefrontal cortex of freely moving macaques. Cell Reports, 42(9), 113091. https://doi.org/10.1016/j.celrep.2023.113091

Wikenheiser, A. M., Gardner, M. P. H., Mueller, L. E., & Schoenbaum, G. (2021). Spatial Representations in Rat Orbitofrontal Cortex. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 41(32), 6933–6945. https://doi.org/10.1523/JNEUROSCI.0830-21.2021

Yoo, S. B. M., Hayden, B. Y., & Pearson, J. M. (2021). Continuous decisions. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 376(1819), 20190664. https://doi.org/10.1098/rstb.2019.0664

Yoo, S. B. M., Tu, J. C., Piantadosi, S. T., & Hayden, B. Y. (2020). The neural basis of predictive pursuit. Nature Neuroscience, 23(2), 252–259. https://doi.org/10.1038/s41593-019-0561-6