The Role of Neural Replay in Structure Learning and Value Generalization

Fabian M. Renz (renz@mpib-berlin.mpg.de)
Institute of Psychology, Universität Hamburg, Von-Melle-Park 5, 20254 Hamburg, Germany
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Shany Grossman (grossman@mpib-berlin.mpg.de)
Max Planck Research Group NeuroCode, Max Planck Institute for Human Development, Berlin, Germany

Nathaniel Daw (daw@princeton.edu)
Department of Psychology, Princeton University, Princeton, NJ, USA
Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA

Peter Dayan (dayan@tue.mpg.de)
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
University of Tübingen, Tübingen, Germany

Christian F. Doeller (doeller@cbs.mpg.de)
MPI for Human Cognitive and Brain Sciences, Leipzig, Germany
Kavli Institute for Systems Neuroscience, NTNU, Trondheim, Norway
Wilhelm Wundt Institute of Psychology, Leipzig University, Leipzig, Germany

Nicolas W. Schuck (nicolas.schuck@uni-hamburg.de)
Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany
Institute of Psychology, Universität Hamburg, Von-Melle-Park 5, 20254 Hamburg, Germany

Abstract

Humans can quickly learn and update latent task structures, and use them to guide value-based decision-making. In a functional magnetic resonance imaging study, 52 participants learned a latent graph structure requiring non-local value generalization upon reward reversals. Performance was best explained by a latent cause inference model that captures structure learning and admits value generalization. The functional imaging data offered a possible substrate for this generalization by demonstrating that the hippocampus tracked the underlying task structure and exhibited non-local reactivation of unobserved sequences sharing a reward.

Keywords: Cognitive maps; Latent Cause Inference; Representational change; Replay

Introduction

The ability to generalize knowledge to novel situations in a flexible manner is a hallmark of intelligent behavior. Generalization has been proposed to be enabled by the formation of abstract representations that capture the underlying structure of an environment, often conceptualized as a cognitive map. Recent work has provided compelling evidence that cognitive maps enable inferences about unobserved relationships and generalization (Garvert, Saanum, Schulz, Schuck, & Doeller, 2023; Liu, Mattar, Behrens, Daw, & Dolan, 2021; Gupta, Van Der Meer, Touretzky, & Redish, 2010). Although cognitive maps are widely implicated in cognitive phenomena in spatial and conceptual domains (Bellmund, Gärdenfors, Moser, & Doeller, 2018), much remains unknown about how they are acquired and updated (Garvert, Dolan, & Behrens, 2017; Moneta, Grossman, & Schuck, 2024; Whittington, McCaffary, Bakermans, & Behrens, 2022), and, in particular, how patterns and regularities are discovered across changing experiences. We sought to elucidate the learning processes that enable the formation and adaptation of cognitive maps, and to test whether neural replay leverages learned task structure to allow non-local value generalization.

Methods

Using a similar paradigm as Liu et al. (2021), we asked participants to learn a latent graph structure that contained a number of outcome nodes associated with fluctuating rewards (Figure 1A). By manipulating both the rewards and the underlying structure, we probed how cognitive maps are built, applied, and adapted. Fifty-two participants completed a four-session fMRI experiment across two days. The graph structure consisted of four distinct four-step sequences, each composed of brief 750-ms video clips of everyday objects (e.g., cars, animals), that led to an outcome node. Unbeknownst to participants, three sequences (paths A, A* and B) led to a shared reward on day one (i.e. the outcomes experienced after these sequences are drawn from the same reward distribution, R1), while one sequence (path C) led to a different reward (R2). Importantly, two of the three common sequences contained

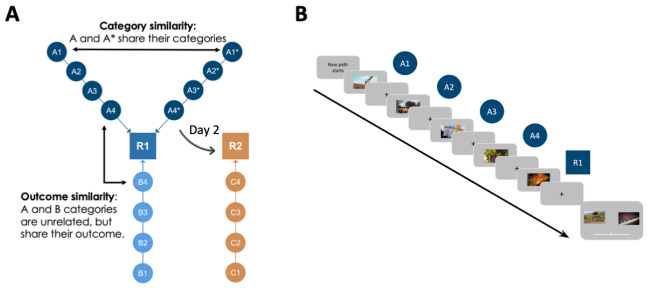

Figure 1: **A.** Latent task structure to be discovered. Four sequences lead to two rewards (R1 and R2). Three paths share the same reward (outcome similarity), enabling value generalization. Moreover, two of these paths share their categories (category similarity). On day two, one shared path (A*) switches from R1 to R2, necessitating relearning and an update in value generalization. **B.** Example trial: Participants passively view 750-ms video sequences before receiving a reward (displayed as a point total ranging from 0 to 100). R1 and R2 differ only in point value and are visually identical, participants must subsequently choose between two nodes to identify the more rewarding option at that moment. After reward reversals, probing unobserved nodes serves as the primary measure of structure knowledge.

semantically similar items (Paths A and A*), while the remaining sequence contained unrelated items (Path B). The images shown when traversing the sole path that led to R2 (path C) were unrelated to all other items.

On day 1, participants learned the task structure by observing the shared reward fluctuation of the paths leading to R1. The underlying reward distributions R1 and R2 were reversed every 15 trials (on average), and we tested participants' knowledge of the graph structure and the changing outcomes by periodically asking them to make value-based decisions between two nodes from two different paths. Crucially, after reward reversals, participants were immediately probed on nodes from unobserved paths, requiring them to generalize value from observed to unseen paths. This one-shot generalization served as our measure of structure knowledge.

On day 2, participants were first explicitly informed about the task structure, and later faced a structural change that required adaptation of their previously learned generalization strategy. FMRI data was acquired on both days and used for representational similarity analysis and classifier-based sequential reactivation analyses (Wittkuhn & Schuck, 2021; Schuck & Niv, 2019). To model the computational processes underlying participants' inference about the reward generating processes and how they changed over time, we developed a Bayesian nonparametric latent cause inference (LCI) model based on a Chinese Restaurant Process (Gershman & Blei, 2012; Gershman, Norman, & Niv, 2015). This allowed an unbounded number of latent causes to be inferred from the data, capturing how participants discover the underlying structure that governs the rewards. The model maintains multiple hypothesized latent structures and learns to generalize by associating multiple sequences with a shared latent cause.

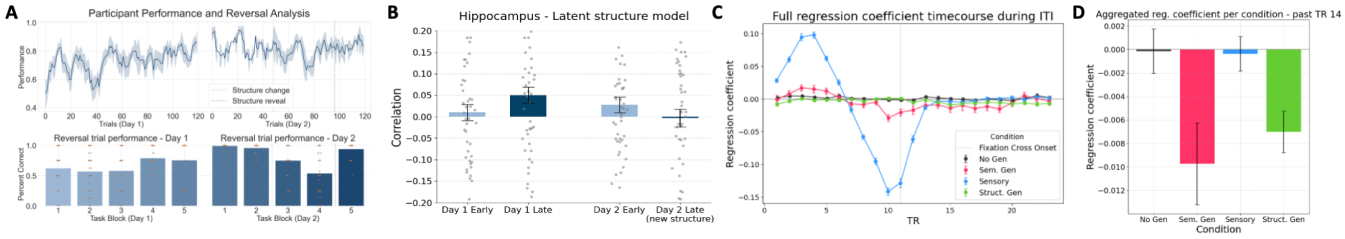

Figure 2: **A.** Participants' overall accuracy improved across the two days (top row) and during reversal trials (bottom row), demonstrating effective value generalization. **B.** Representational similarity analysis revealed that the hippocampus tracked the task structure on day one, with this effect diminishing on day two following a change in the latent structure. **C.** Averaged regression coefficient time courses are shown. The blue line represents the sequential ordering of classifiers for the presented sequence. The green (non-shared category) and red (shared category) lines correspond to coefficients for paths sharing the reward with the presented sequence, while the black line indicates the unrelated sequence. Grouping is relative to the path presented on each trial. Notably, the blue curve displays a sinusoidal pattern—with an early positive phase reflecting stronger activation of early sequence elements, followed by a later negative phase indicating stronger activation of late sequence elements. **D.** During the prolonged inter-stimulus interval, aggregated reactivation evidence showed distinct reactivation for both non-presented sequences that share a reward, compared to the presented and unrelated sequences.

Results

Participants demonstrated successful acquisition and application of the task structure, with the percentage of correct answers in choice trials improving from chance level (50%) to 87% (Fig. 2 panel A top row; $\beta = 0.0012$, $p < 0.001$). Notably, this was also true for performance on one-shot generalization across blocks (Fig. 2 panel A bottom row; $\beta = 0.049$, $p = 0.004$), suggesting that participants correctly learned to exploit the latent structure for value generalization. On day 2, with full knowledge of the task structure, participants performed at near-ceiling levels from the outset. Following the structural change, they quickly adapted their generalization strategy, although there were individual differences in the rate of adaptation. By the end of the task, participants had successfully acquired the new structure, achieving high performance levels. The LCI model provides an account of how participants discovered and utilized the task structure. Through integrating a Chinese Restaurant Process prior with a likelihood formed from the observed category sequence and Gaussian rewards, the model identifies latent causes and over time learns to map multiple sequences to a shared latent cause. This mapping allows resolution of the problem of non-local value updating, as unobserved sequences access a shared value estimate. Model comparison confirmed that the LCI model outperformed alternative approaches, capturing both behavior on reward reversal and structure relearning (LCI mean AICc = 44.6; alternative Temporal Difference agent AICc = 64). Representational similarity analysis revealed that the underlying task structure was tracked in the hippocampus, demonstrating an increased representation over the course of learning on the first day and a decrease in representation after structure change on the second day (tested using a linear mixed-effects model that showed a significant three-way interaction between the latent structure captured by a model similarity matrix incorporating similarity of shared reward nodes, day, and phase, $p = 0.026$; Fig. 2 panel B). To test for replay, classifiers of activity in visual cortex were trained on independent data, and were applied to prolonged interstimulus intervals after the presentation of the reward preceding the decisions. Fast sequences generate a signature monotonic ordering of classifier probabilities as reflected by regression coefficents (positive indicating a forward sequentiality, negative a backward sequentiality). We observed increased reactivation evidence for unobserved sequences sharing a reward compared to unrelated sequences, following the stimulus-evoked neural activation (Fig. 2 panels C and D). This effect was more pronounced in moments of reversing rewards necessitating value generalization ($p = 0.004$) and is predictive of behavioral performance in the following trials ($p = 0.017$), both tested using linear mixed-effect models.

Conclusion

Our findings demonstrate that humans can rapidly learn and flexibly update latent task structures to guide value-based decision-making. The observed patterns of non-local neural replay are in line with previous findings by Liu et al (2021) and provide a potential neural explanation for how unobserved values are updated. This aligns with the emerging representation of task structure in the hippocampus. The latent cause inference model offers one potential computational account for how participants learn the underlying task structure and adapt it flexibly upon change. Together, these results present a potential explanation for how we build, leverage, and adapt our representations of the environment flexibly.

Acknowledgments

N.W.S. is funded by a Starting Grant from the European Union (ERC-2019-17 StG REPLAY-852669) and the Federal Ministry of Education and Research (BMBF) and the Free and Hanseatic City of Hamburg under the Excellence Strategy of the Federal Government and the Länder. FMR was partially funded by the Max Planck School of Cognition. SG is supported by Zuckerman-CHE program. CFD's research is supported by the Max Planck Society, the European Research Council (ERC-CoG GEOCOG724836), the Kavli Foundation, the Jebsen Foundation, Helse Midt Norge and The Research Council of Norway (223262/F50; 197467/F50). PD's research is supported by the and the Alexander von Humboldt Foundation and the Max Planck Society.

References

Bellmund, J. L., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*(6415), eaat6766.

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *elife*, *6*, e17086.

Garvert, M. M., Saanum, T., Schulz, E., Schuck, N. W., & Doeller, C. F. (2023). Hippocampal spatio-predictive cognitive maps adaptively guide reward generalization. *Nature Neuroscience*, *26*(4), 615–626.

Gershman, S. J., & Blei, D. M. (2012). A tutorial on bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1–12.

Gershman, S. J., Norman, K. A., & Niv, Y. (2015). Discovering latent causes in reinforcement learning. *Current Opinion in Behavioral Sciences*, *5*, 43–50.

Gupta, A. S., Van Der Meer, M. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron*, *65*(5), 695–705.

Liu, Y., Mattar, M. G., Behrens, T. E., Daw, N. D., & Dolan, R. J. (2021). Experience replay is associated with efficient nonlocal learning. *Science*, *372*(6544), eabf1357.

Moneta, N., Grossman, S., & Schuck, N. W. (2024). Representational spaces in orbitofrontal and ventromedial prefrontal cortex: task states, values, and beyond. *Trends in Neurosciences*.

Schuck, N. W., & Niv, Y. (2019). Sequential replay of nonspatial task states in the human hippocampus. *Science*, *364*(6447), eaaw5181.

Whittington, J. C., McCaffary, D., Bakermans, J. J., & Behrens, T. E. (2022). How to build a cognitive map. *Nature neuroscience*, *25*(10), 1257–1272.

Wittkuhn, L., & Schuck, N. W. (2021). Dynamics of fmri patterns reflect sub-second activation sequences and reveal replay in human visual cortex. *Nature communications*, *12*(1), 1795.