Dynamics and Structure of Generalization During Reinforcement Learning in Human Brains and Artificial Networks

Shany Grossman (shany.grossman@uni-hamburg.de)
Institute of Psychology, University of Hamburg,
Von-Melle-Park 5 20254 Hamburg, Germany
and
Max Planck Institute for Human Development,
Max Planck UCL Centre for Computational Psychiatry and Ageing Research,
Lentzeallee 94, 14195 Berlin, Germany

Noa Hedrich (noa.hedrich@uni-hamburg.de)
Institute of Psychology, University of Hamburg
Von-Melle-Park 5 20254 Hamburg, Germany.
and
Max Planck Institute for Human Development
Max Planck UCL Centre for Computational Psychiatry and Ageing Research,
Lentzeallee 94, 14195 Berlin, Germany

Andrew Saxe (a.saxe@ucl.ac.uk)
Gatsby Computational Neuroscience Unit & Sainsbury Wellcome Centre,
University College London, Howland 25, W1T 4JG London, UK
and
CIFAR Azrieli Global Scholar, CIFAR.

Nicolas W. Schuck (nicolas.schuck@uni-hamburg.de)
Max Planck Institute for Human Development
Max Planck UCL Centre for Computational Psychiatry and Ageing Research,
Lentzeallee 94, 14195 Berlin, Germany
and
Institute of Psychology, University of Hamburg,
Von-Melle-Park 5 20254 Hamburg, Germany.

Abstract

Goal-directed decision making amidst an overwhelming stream of sensory input requires learning internal representations that capture a task's underlying structure. Importantly, such internal abstractions enable generalization. Representing an object's shape but ignoring its color, for instance, means that anything learned about a green triangle will generalize to red triangles. Here, we investigate this dynamic interaction of task representation learning and generalization. Human participants and artificial neural networks were trained with the same contextual reinforcement learning task. Analyses of human data reveal that participants learned an abstract task structure, which becomes detectable in the orbitofrontal cortex (OFC) after learning. Recurrent neural networks trained on the same learning curriculum exhibit similar abstractions of task representations over time. Notably, we find that the similarity structure of the networks' internal task representations affects how weight updates after a single example alter network behavior and representations on other trials. The network's progressing context differentiation in its internal layers hence leads to generalization of single experiences to other events within the same context. Ongoing work aims to gain a mechanistic understanding of model observations and contrast them with learning dynamics in the human brain.

Keywords: decision making; reinforcement learning; fMRI; task states; task representation learning.

Introduction

Humans excel at making decisions to achieve task goals despite being faced with an overwhelming stream of information. This ability is commonly attributed to the brain's capacity to distil an abstract representation of task-relevant components and their relations, forming a "cognitive map" or a "task state-space" (Behrens et al., 2018; Niv, 2019). A key property of such representations is that they guide generalization of prior experiences to new, unexperienced, ones. For instance, a novice chess player may decide on moves in unfamiliar board configurations by generalizing from previous encounters according to how similar their internal representation of the current board is to previously seen boards.

The exact task properties that facilitate effective learning and decision making depend on the task. Hence, task representations must themselves be learned, a "learning to learn" process. Computationally, this process entails the challenge of rapidly finding a transformation of high-dimensional inputs onto a reward-predictive low-dimensional space. Learning a useful transformation is a challenge for biological and artificial neural networks alike (Lake & Baroni, 2023; Radulescu et al., 2025; Zhang et al., 2020), yet its underlying mechanisms remain poorly understood.

Here, we set out to investigate the dynamics of task representation learning and its interaction with generalization in the human brain and in artificial networks.

Task and experimental design

Participants (total N=62) completed the "Realtor task" across two fMRI sessions. On each learning trial, participants first viewed a client, followed by two house alternatives and then decided which house the client would prefer (Fig. 1, top row). Their decision was followed by numerical feedback (0-100) indicating the client's satisfaction. The clients fell into two types based on a certain feature (e.g., whether they had glasses). The defining feature of the client type therefore provided context for deciding between the subsequent house alternatives. Participants therefore learned the house preferences of client types through feedback. Solving the task is challenging since, as in real life scenarios, the features of the clients and houses varied in their relevance for making correct choices.

A critical component in the experimental design is the inclusion of estimation trials. During estimation trials, participants did not receive feedback but instead observed client-house pairs that are distinct from those shown in neighbouring learning trials. The same set of estimation trials was presented prior to and immediately after segments of learning trials (Fig. 1 middle and bottom row). This design enables us to track "mass representation learning": how activity patterns for task items co-shift in response to feedback on a single, distinct, task item.

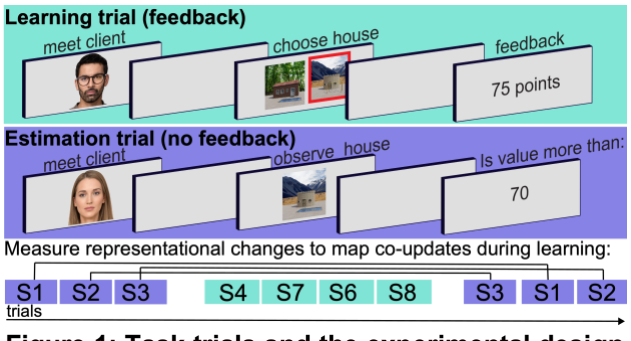

Figure 1: Task trials and the experimental design

Task abstractions in behavior and brain

Behavioral modelling. We modeled successful participants' behavior to test whether they learned the true underlying task structure, rather than memorizing individual input-output mappings. To do this, we fitted decisions using linear approximation models assuming two alternative representations of task states. In the *true model*, a one-hot vector with 8 nodes represented the true underlying task states. In contrast, the *full model* used a 32-node vector that preserved all possible client-house pairs, without merging irrelevant variations or splitting identical observations based on the prior context (i.e., client type). The models were optimized to minimize the negative log-likelihood between model predictions and participant choices. The true model significantly outperformed the full model (paired t-test, $p<10^{-8}$, Fig. 2a), demonstrating that participants learned an abstract representation of the task rather than an elaborated input-output mapping.

Preliminary fMRI results. Correlating single trial activity patterns from identical task states revealed that abstract state representations emerged in the OFC after learning (paired t-test, p=0.0005, Fig. 2b), in accordance with previous findings (Schuck et al., 2016). Further analyses will address anatomical specificity of the effect and its time-resolved evolution.

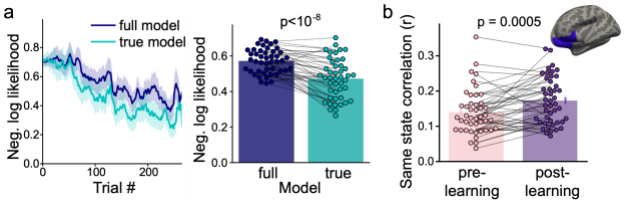

Figure 2. Behavioral modelling and preliminary fMRI results (a) Linear approximation model fits per trial (left) and averaged across trials (right, N=45 successful learners). (b) Mean correlations between single trial activity patterns in the OFC corresponding to the same task states prior versus after learning.

Dynamics of task abstractions in artificial networks

Neural network model. To contrast artificial with human learning, we first analyzed learning dynamics in a simple recurrent neural network composed of two hidden layers: a recurrent layer (64 units, tanh) followed by a fully connected layer (32 units, tanh). The network receives as input two binary vectors presented consecutively, the first indicating the client and the second indicating house features, and outputs a value estimation for each house. On each trial, the network receives feedback on the chosen house which is then used to update its weights via gradient descent optimization and back-propagation.

"Mass representation learning" analysis. To investigate how receiving feedback on a single item shifts representations of other task items, we recorded hidden representations for all possible items prior to and immediately after each learning trial. Initial results demonstrate that a single weights' update triggers a coordinated shift in all hidden representations pertaining to the same context (Fig. 3b). A control analysis ruled out the possibility that this effect was due to shared context-selectivity of hidden units. This dynamic was also observed in the network behavior, as increased performance for all task items in the same context that was given feedback (Fig. 3c).

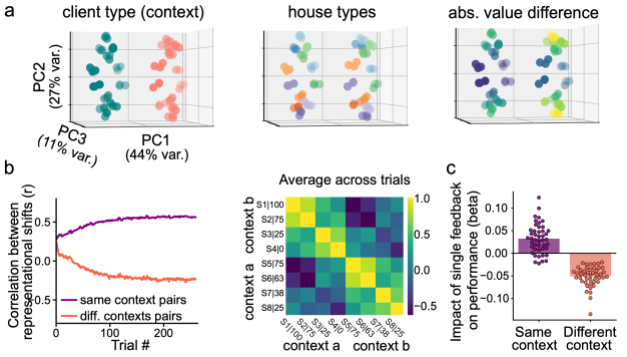

Figure 3. Neural networks simulations (a) Post learning hidden representations (recurrent layer) in a representative network. Each circle is a trial with a unique visual input. Different color-codes visualize encoded dimensions. (b) Mean pairwise correlations of change vectors induced by single feedback updates, averaged separately across pairs of task items from the same/different contexts (left), and the full pairwise matrix averaged across all learning trials (right). (c) GLM estimates of the impact single feedback has on performance in the same versus different context as the item given feedback. Individual circles mark the different networks, each trained with a curriculum of a specific participant.

Outlook

Ongoing work focuses on analyzing mass representation learning dynamics in fMRI data, comparing brain signals with neural network simulations and predictions derived from their analytical solutions. The results can advance our understanding of how efficient learning is achieved in the human brain, enabling generalization from single events to suitably similar future situations.

References

Anand, A., Racah, E., & Ozair, S., Bengio, Y., Côté, M.A., & Hjelm, R. D. (2019). *Unsupervised State Representation Learning in Atari*. *Advances in Neural Information Processing Systems*, 32. https://doi.org/10.48550/arXiv.1906.08226

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Kirk, R., Zhang, A., Research, M. A., & Rocktäschel, T. (2023). A Survey of Zero-shot Generalisation in Deep Reinforcement Learning. In *Journal of Artificial Intelligence Research* (Vol. 76) https://doi.org/10.1613/jair.1.14174

Lake, B. M., & Baroni, M. (2023). Human-like systematic generalization through a meta-learning neural network. *Nature*, *623*(7985), 115–121. https://doi.org/10.1038/s41586-023-06668-3

Li, D., Yang, Y., Song, Y.-Z., & Hospedales, T. M. (n.d.). *Learning to Generalize: Meta-Learning for Domain Generalization*. www.aaai.org

Niv, Y. (2019). Learning task-state representations. *Nature Neuroscience*, *22*(10), 1544–1553. https://doi.org/10.1038/s41593-019-0470-8

Radulescu, A., Shin, Y. S., & Niv, Y. (2025). *Human Representation Learning*. *28*, 26. https://doi.org/10.1146/annurev-neuro-092920

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005

Zhang, A., McAllister, R., Calandra, R., Gal, Y., & Levine, S. (2020). *Learning Invariant Representations for Reinforcement Learning without Reconstruction*. *International Conference on Learning Representations*. https://doi.org/10.48550/arXiv.2006.10742