One-shot reinforcement learning decisions engage the hippocampus

Christopher S. Iyer (c.iyer@columbia.edu)
Mortimer B. Zuckerman Mind, Brain, Behavior Institute & Department of Psychology, Columbia University

Raphael T. Gerraty (rtg2116@columbia.edu)
Mortimer B. Zuckerman Mind, Brain, Behavior Institute, Columbia University

Katherine D. Duncan (katherine.duncan@utoronto.ca)
Department of Psychology, University of Toronto

Nathaniel D. Daw (ndaw@princeton.edu)
Princeton Neuroscience Institute & Department of Psychology, Princeton University

Daphna Shohamy (ds2619@columbia.edu)
Mortimer B. Zuckerman Mind, Brain, Behavior Institute, Kavli Institute for Brain Science, & Department of Psychology, Columbia University

Abstract

Decisions about value can be based on multiple sources of information from memory. Classic reinforcement learning models describe how value estimates are incrementally learned over many trials, while decisions can also be guided by "one-shot" episodic memories for single experiences. The goal of this study is to better understand the contribution of these two processes—incremental learning and episodic memory—to value-based decisions. Human participants were scanned with fMRI while performing a decision task in which choices could be guided by either incremental and episodic information. Choices based on episodic information were associated with increased BOLD activity in the hippocampus. Intriguingly, hippocampal activity was also associated with incrementally-learned value information, derived from reinforcement learning models. Finally, we observed reward-related reinstatement of patterns during episodic decisions in the ventromedial prefrontal cortex. These findings reveal both shared and distinct markers of incremental and episodic memory during value-based decisions.

Keywords: episodic memory, reinforcement learning, hippocampus, striatum

Introduction

We rely on memory of past experiences to guide value-based decisions. In scenarios with repeatedly encountered options, decisions can be guided by value estimates averaged over many previous experiences. These value estimates are computed and updated through an incremental learning process captured by reinforcement learning models (Sutton and Barto 1998; Daw et al. 2006; Daw 2011), which mirror patterns of neural firing in midbrain dopamine neurons (Schultz et al. 1997; O'Doherty et al. 2003). However, incremental averaging is insufficient for decisions based on sparse or single episodes, where episodic memory is necessary (Gershman and Daw 2017). Few behavioral paradigms combine incremental learning and episodic memory on a trial-by-trial basis, leaving open questions about how they trade off to guide decisions and, in particular, about when and how participants use episodic memory to evaluate choice options.

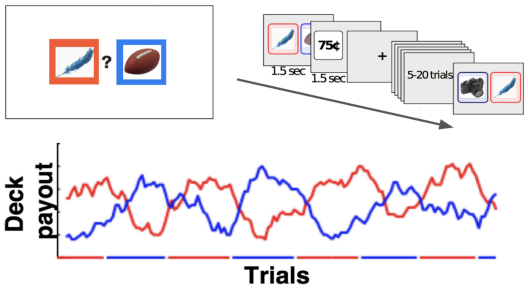

Figure 1: task design

Here, 31 human participants performed a hybrid incremental-episodic decision task while undergoing functional magnetic resonance imaging. On each trial, they chose between two decks with cards depicting unique objects. The average deck payouts fluctuated and reversed every 16-24 trials, allowing participants to choose based on incrementally-learned deck averages. Some objects were presented again 10-30 trials later, and participants were told that old cards would be worth their previous value, allowing them to also select based on one-shot episodic memories of a card's value. Old cards were assigned to decks pseudorandomly, to decorrelate deck value and old card value. Participants could thus rely independently on estimates of each deck's average value or on episodic memory for the value of a previously seen object. We investigated when participants are most likely to use episodic information, and what neural systems support this process.

Results

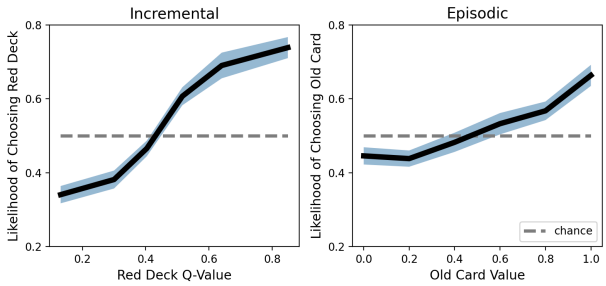

Figure 2: choices reflect both old card value and deck Q-value

Behaviorally, participants used both incremental and episodic information to guide their decisions. They were more likely to select decks with higher estimated value, indicating successful incremental learning (β=3.49, SE=0.46, $p<0.001$; deck value was computed by fitting Q-learning models to subject choices). They

were also more likely to select cards previously shown with higher values, indicating successful episodic retrieval (β=0.79; SE=0.16; p<0.001).

When were participants likely to use episodic memory? As more trials pass after a deck reversal, participants may grow more certain about average deck values, and may be less likely to use episodic information. Indeed, we found that the effect of object value on choice was strongest shortly after a reversal, and decreased in subsequent trials (i.e., interaction between trials since reversal and object value; β=-0.04, SE=0.02, p<0.05).

To identify trials in which participants engaged episodic memory, we estimated each trial's "episodic likelihood," computed as the log probability of each choice, conditioned on a logistic choice model including only object value (and ignoring deck value). This metric estimates the degree to which each choice is consistent with episodic retrieval of object value, rather than incrementally-learned deck value. Parametrically regressing this likelihood on BOLD data during choice revealed higher activity in the hippocampus on trials with higher episodic likelihood.

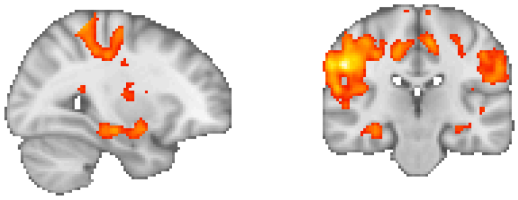

Figure 3: brain regions correlating with episodic likelihood (voxelwise p < 0.001; cluster forming p < 0.05)

Unexpectedly, hippocampal BOLD activity also correlated with the Q-value of the chosen deck, derived from reinforcement learning models including both deck and object value. These results suggest that overlapping neural processes may underlie decisions using episodic and incremental information (Bornstein et al. 2017).

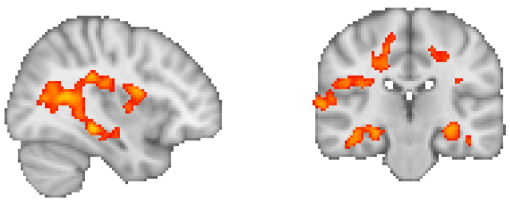

Figure 4: brain regions correlating with Q-value of chosen deck (voxelwise p < 0.001; cluster forming p < 0.05)

To assess encoding-related activity, we shifted our measure of episodic likelihood onto the trial during which object value was initially encoded, quantifying the likelihood that a single item would later be used to guide retrieval choice. A whole-brain interaction of this encoding-shifted likelihood with the number of trials since reversal yielded two clusters in the bilateral hippocampus, neither of which survived whole-brain FWE correction (uncorrected p<0.01).

Lastly, we asked if detailed patterns from reward feedback encoding were reinstated during retrieval choice. We extracted multivoxel trial patterns from two a priori ROIs: hippocampus and ventromedial prefrontal cortex (vmPFC). vmPFC patterns during reward feedback periods contained information about value, as indicated by above-chance decoding of reward value (t(30)=5.72, p<0.001). vmPFC patterns also showed reinstatement of episode-specific value information from encoding, as indicated by higher correlation to the reward feedback period of their encoding trials than other trials with the same reward magnitude (t(30)=4.35, p<0.001). These results show that vmPFC patterns contain information both about general reward value and trial-specific episodic information. These effects were not observed in the hippocampus.

Conclusion

Our findings show that participants use both incremental and episodic information to guide value-based decisions, and that hippocampal activity is associated with decisions engaging episodic retrieval of single-shot value. Encoding of episodic information is greatest when participants are less certain about incremental value estimates. We find that hippocampal and medial temporal cortex activity also correlates with incremental value, suggesting shared neural circuitry underlying these two processes. Finally, we show that distributed patterns of activity in vmPFC encode reward value during feedback, and reinstate episode-specific information during subsequent retrieval choice.

References

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature communications, 8*, 15958. https://doi.org/10.1038/ncomms15958

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876–879. https://doi.org/10.1038/nature04766

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron, 69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: An integrative framework. *Annual Review of Psychology, 68*, 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 38*(2), 329–337. https://doi.org/10.1016/s0896-6273(03)00169-7

Sutton, R.S. and Barto, A.G. (1998). Reinforcement Learning: An Introduction. Vol. 1, MIT press, Cambridge.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science (New York, N.Y.), 275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593