High-resolution tracking of internal model formation using automated live-in training

Rifqi O. Affan (roaffan@bu.edu)
Graduate Program for Neuroscience, Boston University
Boston, MA

Benjamin B. Scott (bbs@bu.edu)
Department of Psychological and Brain Sciences and Center for Systems Neuroscience, Boston University
Boston, MA

Abstract:

The ability to acquire and use an internal world model to plan actions is a hallmark of general intelligence. How this adaptive mechanism develops throughout learning and which neural mechanisms underlie it are unknown. Here we used an automated training system to collect data 24 hours a day, 7 days a week from rats learning and performing a two-step decision-making task designed to dissociate model-based planning from model-free strategies. The live-in system allowed self-paced, around-the-clock training, providing high-temporal resolution assessment of learning in the two-step task. Furthermore, this system allowed behavioral data collection at high temporal resolution and revealed how model-based planning emerges during learning. These data indicate that rats rapidly adopted an internal model of the two-step task and exhibited a model-based planning strategy early in training.

Introduction

In sequential decision-making tasks where reward contingencies are probabilistic, humans employ an internal model of the task statistics to maximize rewards (Daw et al., 2011). Like humans, rats employ a model-based strategy to solve similar tasks, providing opportunities for studying the neurobiological mechanisms that underlie this behavior (Miller et al., 2017, 2022). However, whether such strategies are gradually developed in rats and whether they are robust against overtraining is an open question (Redish, 2016). Moreover, the neural circuitry and computations driving the emergence of this behavior are unknown and insights from experimental work may inform efforts in advancing artificial intelligence. Here, we implemented an automated live-in operant system that allowed rats to engage in self-paced training on the two-step decision-making task and provided behavioral measures across learning stages.

Methods

Rats were trained to perform a two-step task in a six-port operant chamber (Miller et al., 2017). In the first step of the task, a rat initiates a trial by poking its nose into the top-center port and then chooses one of two top-side ports (Figure 1a, i-ii). One top-side port commonly activates the LED in the bottom-left port (p=80%) and rarely activates the bottom-right port (p=20%) in the second step of the task, while the other top-side port has the opposite consequence. Once a decision is made, the rat must initiate the second step by poking into the bottom-center port (Figure 1a, iii). One of the bottom-side ports is then illuminated, depending on the rat's choice in the first step (Figure 1a, iv). The rat must enter this active port to obtain a sucrose reward. The reward probability at each of the bottom-side ports is either 80% or 20% and is randomly flipped at random intervals (Figure 1b-c).

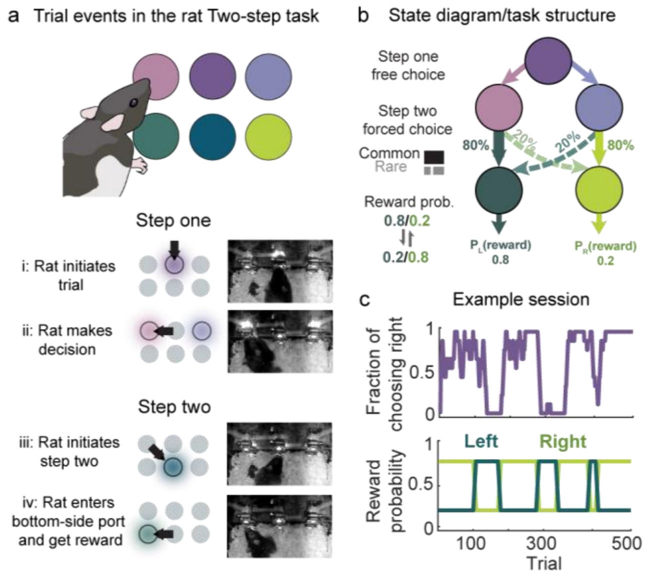

Figure 1: (a) Schematic of trials in two-step task. (b) State diagram describing task rule. (c) An example behavioral session from one example rat.

Results

Rats learn the two-step task through self-paced training in an automated live-in system.

We implemented the two-step task in a live-in operant system, which allowed for self-paced training. Rats performed trials throughout the day and night, exhibiting variable 24-hour cycles of task engagement (Figure 2a and b). Peak activity occurred during the dark period of the rats' light/dark cycle (median = 6th hour, interquartile range = 3.5 to 7.8 hrs into darkness).

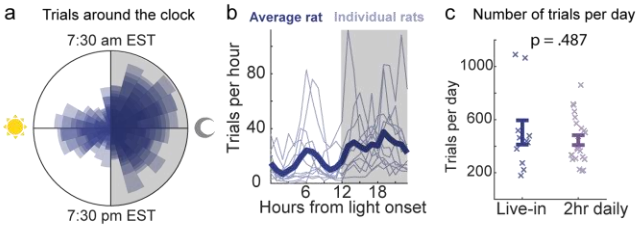

Figure 2: (a) Rats performed the task around the clock in the live-in system. (b) Number of trials rats performed at different hours of the day. (c) Rats in the live-in system performed similar numbers of trials per day (503 ± 303 trials/day) compared to those classically trained through daily 2-hour sessions (445 ± 178 trials/day).

Rat performance in the live-in training condition matched those exhibited by rats trained in 2-h daily sessions as assessed by several metrics (Figure 2c and Figure 3a). First, we observed similar rates of choice adjustment following reward probability flips. Second, we observed similar asymmetries in response time (RT) for rare and common transitions (Miller et al., 2017 Akam et al., 2021; Figure 3b). Finally, using a trial-history regression analysis and mixture-of-agent model fits (Miller et al., 2017, 2022), we show that rats trained in the self-paced condition adopted a model-based strategy like those trained in daily 2-h sessions (Figure 3c and d). Together these results suggest that rats exhibit similar planning strategies in the live-training system and conventional daily training systems.

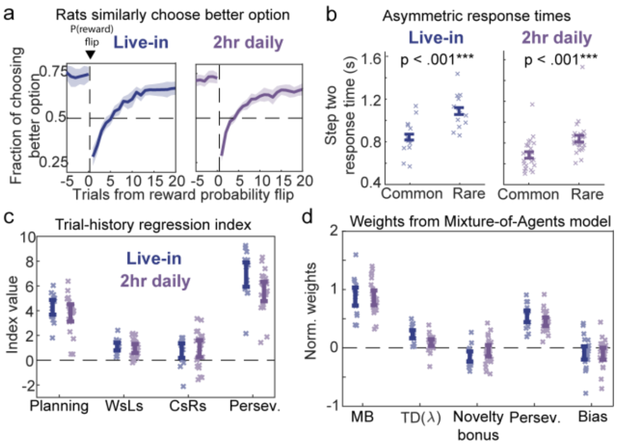

Figure 3: (a) Rate of choosing the better option following a reward probability flip. (b) RT in the second step following common and rare transitions. (c) Behavioral strategy index based on trial-history regression of rat choices. (d) Weight parameters from Mixture-of-Agents model fit to rat data.

Model-based decisions emerge rapidly.

Next, we used the high-temporal resolution behavioral data generated by the live-in system to evaluate learning in the two-step task. We first looked at how model-based planning emerged using trial-history regression. A rolling estimate of strategy indices indicated that model-based decisions increased over the first 1000 trials (Figure 4c), while other behavioral patterns, such as win-stay/lose-switch, perseveration, and bias, were stable throughout training.

We next evaluated the emergence of RT asymmetry for rare and common transitions (Akam et al., 2021; Castro-Rodrigues et al., 2022; Miller et al., 2017), which is thought to reflect knowledge of the transition structure. Surprisingly, RT asymmetry was observed during an earlier shaping stage of training (Figure 4d-e). In this stage, animals made no decision, but simply followed light sequences through trials with both rare and common transitions. This result indicates that rats learned action-outcome transitions (i.e., the internal model of the task structure) during the shaping period and readily used it to exploit rewards during training. This observation resembles previous findings from maze-based tasks where rats form an internal model of the environment through latent learning (Tolman, 1948). Interestingly, we also observed a transient increase in the asymmetries during the first 1000 trials of the training stage, which was also the period when model-based planning first emerged (Figure 4c-d).

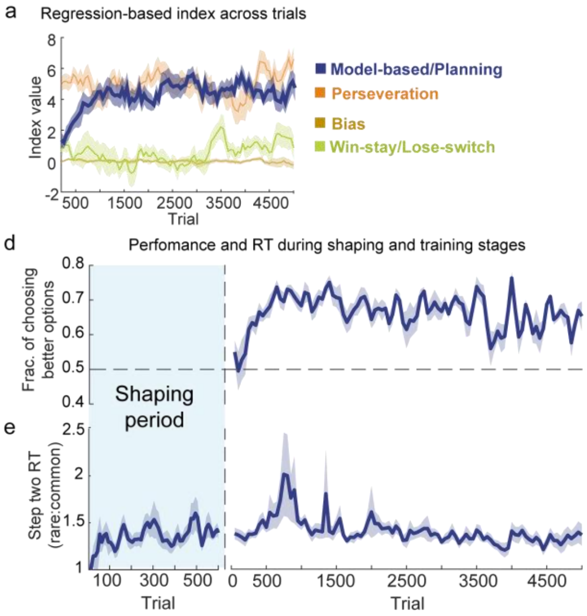

Figure 4: (a) Rolling estimate of behavioral indices during learning. (b) Rat performance during learning. (c) Response time ratio during shaping and training stages.

Conclusions

Our results demonstrate that a fully automated live-in operant system provided rich behavioral data across daily cycles and learning stages. Using this system, we find that rats develop a model-based strategy early during learning and continue to leverage it across thousands of trials to maximize rewards. Future studies could exploit this rapid form of learning to study the neural mechanisms of model formation.

Acknowledgments

We thank Kevin Miller for insightful discussions and advice and Joe McGuire for valuable feedback. We also thank Quan Do, Hongjie Xia, Gary Kane, and the Boston University Animal Science Center for helping develop and maintain the live-in training facility. ROA was supported by NIH Blueprint DSPAN Award no. F99NS130925.

References

Akam, T., Rodrigues-Vaz, I., Marcelo, I., Zhang, X., Pereira, M., Oliveira, R. F., ... & Costa, R. M. (2021). The anterior cingulate cortex predicts future states to mediate model-based action selection. *Neuron*, *109*(1), 149-163.

Castro-Rodrigues, P., Akam, T., Snorasson, I., Camacho, M., Paixão, V., Maia, A., ... & Oliveira-Maia, A. J. (2022). Explicit knowledge of task structure is a primary determinant of human model-based action. *Nature human behaviour*, *6*(8), 1126-1141.

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.

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature neuroscience*, *20*(9), 1269-1276.

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2022). Value representations in the rodent orbitofrontal cortex drive learning, not choice. *Elife*, *11*, e64575.

Redish, A. D. (2016). Vicarious trial and error. Nature Reviews Neuroscience, 17(3), 147-159.

Smies, C. W., Bodinayake, K. K., & Kwapis, J. L. (2022). Time to learn: The role of the molecular circadian clock in learning and memory. *Neurobiology of learning and memory*, *193*, 107651.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological review*, *55*(4), 189.