Modeling the emergence of instrumental learning in an odor-based 2AFC task

Juliana Chase* **(j.chase@berkeley.edu)**
Department of Psychology
University of California, Berkeley

Jing-Jing Li* **(jl3676@berkeley.edu)**
Helen Wills Neuroscience Institute
University of California, Berkeley

Anne G. E. Collins (annecollins@berkeley.edu)
Department of Psychology
Helen Wills Neuroscience Institute
University of California, Berkeley

Linda Wilbrecht (wilbrecht@berkeley.edu)
Department of Psychology
Helen Wills Neuroscience Institute
University of California, Berkeley
* These authors contributed equally to this work

Abstract

Non-human animals and humans are able to learn policies rapidly with few exposures. However, the earliest moments of learning are difficult to capture and model and are thus understudied. In part, this is due to high levels of variability in learning trajectory across individuals. Here we train adolescent mice in an odor-based two-alternative forced choice (2AFC) task and then extend a recently developed latent state cognitive modeling framework to fit our behavioral data. This framework dynamically estimates decision policies on a trial-by-trial basis, capturing an animal's likelihood to remain in either of two latent decision states: reinforcement learning (RL) and biased for some action. We found that our hybrid model was a better fit than the RL policy alone, and that it successfully explained individual learning trajectories in a way that the RL model could not. All together, our task and model provide novel insight into the earliest moments of learning.

Keywords: decision-making; hidden Markov model; mice behavior; learning

Introduction

Human and non-human animals learn and adapt to changes in their environment without overtraining to specific situations. However, despite their critical relevance, the earliest moments of instrumental learning are excluded and rarely studied or modeled in neuroscience, perhaps due to unstructured variability in an animal's performance. Here, we study early learning in developing mice in an odor-based variant of the commonly used 2AFC task (Figure 1A). Within their first odor learning session, habituated mice show odor stimulus-action learning. But which strategies are employed to drive learning? For example, it is possible that animals initially ignore odor cues and instead use place (biased) strategies before understanding task structure and showing an increase in performance. Likewise, this shift towards engaged learning could be transient and precede animals transitioning towards a biased strategy within the same session (e.g., if the animal becomes demotivated after reaching satiation).

Recently, researchers have used increasingly complex models to better capture an array of possible strategies an animal may use to solve decision-making tasks. Hidden Markov models (HMM) that use latent states to describe hybrid policies outperform fully observable models (Ashwood et al., 2022; Bolkan et al., 2022). However, existing HMM approaches are constrained to descriptive policies, such as generalized linear models (Ashwood et al., 2022), and cannot be trivially extended to policies that describe generative processes, such as reinforcement learning. To develop a modeling framework that can capture both latent learning strategies and processes, we extend our recently developed HMM-based modeling framework (Li, Shi, Li, & Collins, 2024) to dynamically estimate decision policies underlying behavior on a trial-by-trial basis (Figure 1B).

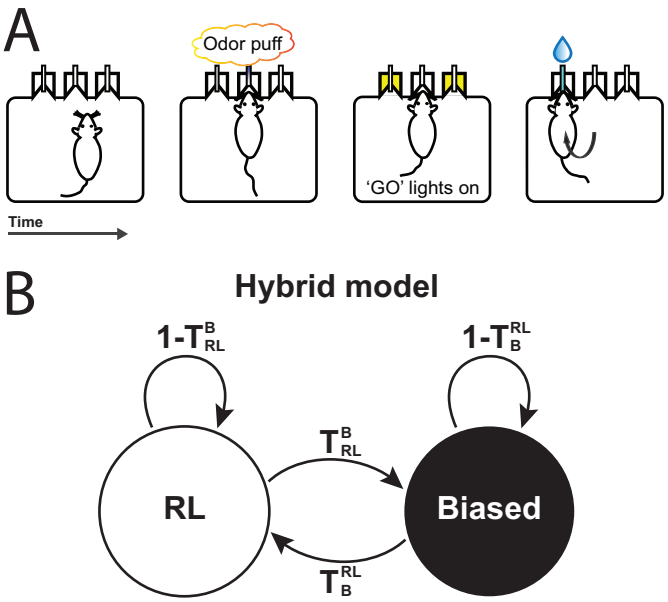

Figure 1: A: Behavioral task schematic. Animals self-initiate a trial by poking into the center port, releasing one of two odors. Following 'go' lights, animals make a choice of two lateral side ports and receive a water reward if correct. B: Structure of the hybrid HMM combining the RL policy with a biased policy for some action.

Methods

Task

Naive, water-restricted C57B/L6 male mice (n=31) at postnatal day (P) 27 were first habituated to the operant chambers and taught to receive 2µL water rewards from lateral nosepoke ports and to initiate trials via nosepoke to the central port (see Figure 1 A). Following this brief period, animals were exposed to two odors, odor A (cinnamon) & B (vanilla) which were presented pseudo-randomly and deterministically predicted water reward on either the left or right ports, respectively.

Modeling

We evaluated an RL model against a hybrid model between the same RL policy and a biased policy (Figure 1). On trial t, the RL model samples an action according to π_t, a softmax policy over the action values in the current state s_t:

$$\pi_t(s_t) = \text{softmax}\big(\beta \cdot (Q_t(s_t) + \text{stickiness})\big),$$

where β is the inverse temperature parameter. "Stickiness" is a set of parameters to model sticking to the same action conditioned on whether the state has changed from the previous state and whether the previous action was rewarded. When the chosen action a_t is rewarded, the action value is updated:

$$Q_{t+1}(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \alpha_+\big(1 - Q_t(s_t, a_t)\big),$$

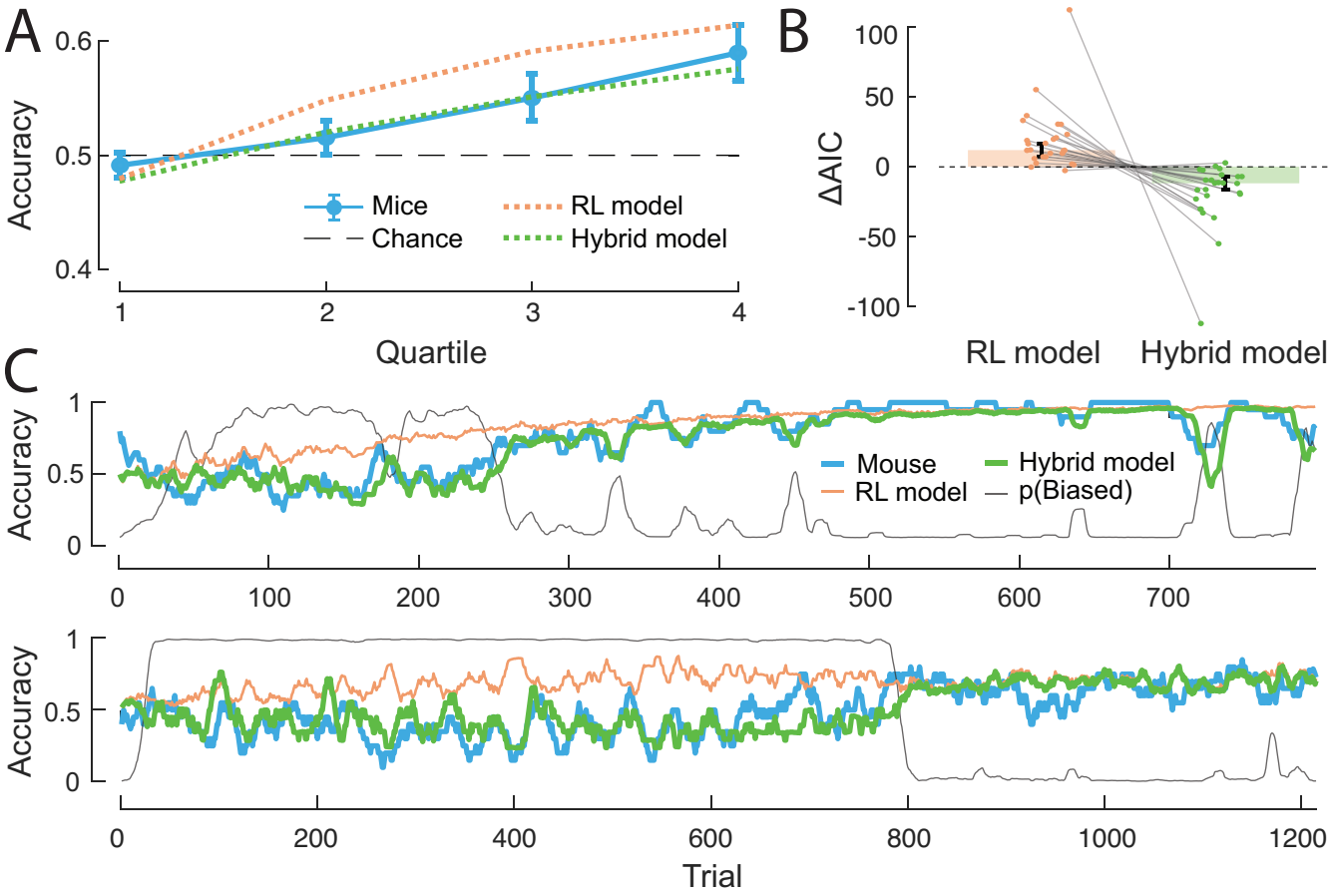

Figure 2: A: Mice and model choice accuracy behavior over quartiles of early learning (single session). B: Model comparison by AIC favors the hybrid model. C: Example individual animal and model learning curves. p(Biased) = 1 - p(RL) is the estimated probability of occupying the "Biased" latent state in the hybrid model.

where α_+ is the learning rate parameter. The biased policy is a heuristic defined by a "bias" parameter between 0 and 1:

$$\pi_t(\text{left}|s_t) = \text{bias}.$$

The hybrid model uses an HMM to construct a mixture policy between the RL and biased policies. Its likelihood is estimated by extending the dynamic noise estimation framework (Li et al., 2024). All models were fitted using maximum likelihood estimation, and we verified that models and parameters are identifiable (Wilson & Collins, 2019).

Results

On average, adolescent animals learned to choose the correct actions associated with odors A & B above chance within the first session (Figure 2A). The hybrid model was a better fit than the RL model according to the Akaike information criterion (AIC; Figure 2B). At the group level, the hybrid model correctly captured decision accuracy across learning (Figure 2A); for individual animals, it successfully explained highly variable trajectories in early learning sessions using the estimated probability of occupying each latent policy state trial-by-trial (e.g., p(Biased); Figures 1B, 2C). As expected, individual animals displayed large differences in estimated latent state occupancy trajectories, perhaps driven by behavioral differences in strategy or bias. Together, these results suggest that the addition of the hybrid mechanism can account for variance in choice behavior that the RL model alone fails to capture.

Discussion

For decades, those using rodent models to study learning have primarily analyzed trained, or stable, behavior, bypassing a moment of learning that could be key to forming and testing novel hypotheses about the neuroscience of how animals learn. Here we present a novel hybrid model applied to the earliest forms of instrumental learning, that allows task acquisition to be studied empirically. In addition to studying C57B/L6 adolescents, we are currently applying our task and model to adult mice and mice with mutations to autism risk genes to examine differences in learning trajectories that may indicate broader neural circuitry changes.

Acknowledgments

This work was supported by the Simons Foundation (Award /613972) and NIMH R01MH119383. We thank Fernanda Castro, Gaby Smith, and Anna Jahng for their help with animal behavior.

References

Ashwood, Z. C., Roy, N. A., Stone, I. R., Laboratory, I. B., Urai, A. E., Churchland, A. K., ... Pillow, J. W. (2022). Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience*, *25*(2), 201–212.

Bolkan, S. S., Stone, I. R., Pinto, L., Ashwood, Z. C., Iravedra Garcia, J. M., Herman, A. L., ... others (2022). Opponent control of behavior by dorsomedial striatal pathways depends on task demands and internal state. *Nature neuroscience*, *25*(3), 345–357.

Li, J.-J., Shi, C., Li, L., & Collins, A. G. (2024). Dynamic noise estimation: A generalized method for modeling noise fluctuations in decision-making. *Journal of Mathematical Psychology*, *119*, 102842.

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, *8*, e49547.