Deciphering value learning rules in fruit flies using a model-driven approach

Rishika Mohanta[1,2], Adithya Rajagopalan[1,3], Glenn Turner[1]

[1]HHMI Janelia Research Campus, Ashburn, VA, USA
[2] Indian Institute of Science Education and Research, Pune, Maharashtra, India
[3] The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Find me here!

Me

Or come find me!

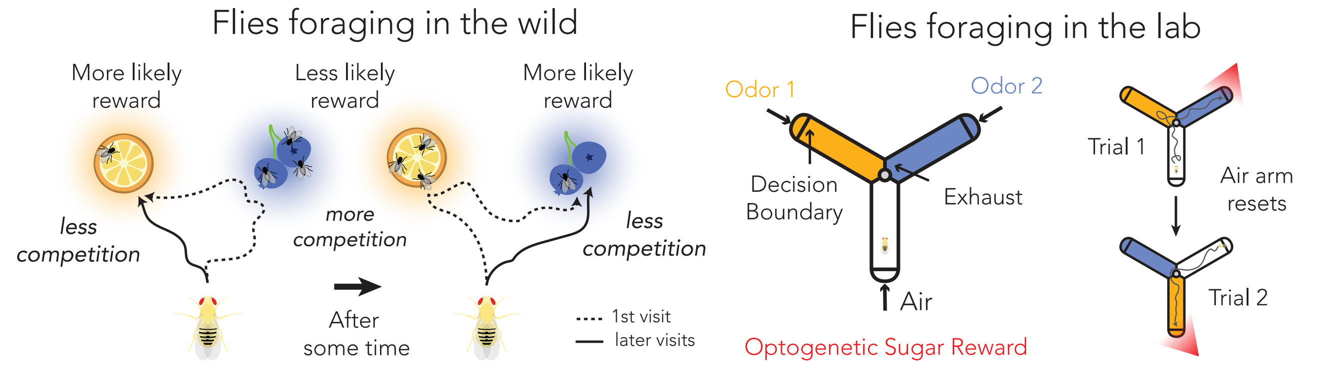
1 | Introduction

Humans and flies both face foraging challenges in daily life.

Humans foraging in the wild

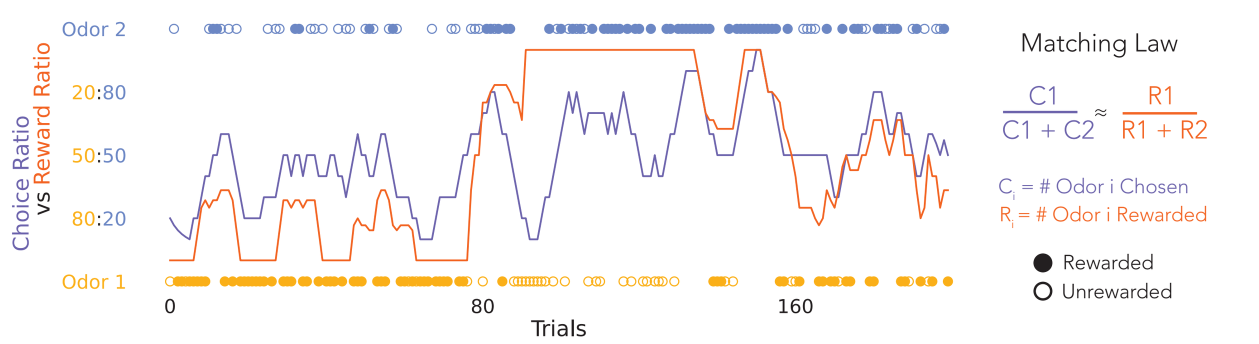
Flies foraging in the wild

Flies foraging in the lab



Decision-making process underlying foraging can be replicated in an Y-maze with two odors.

Flies show operant matching strategy (Rajagopalan et al. 2022)



Fundamentally different strategies can show identical operant matching

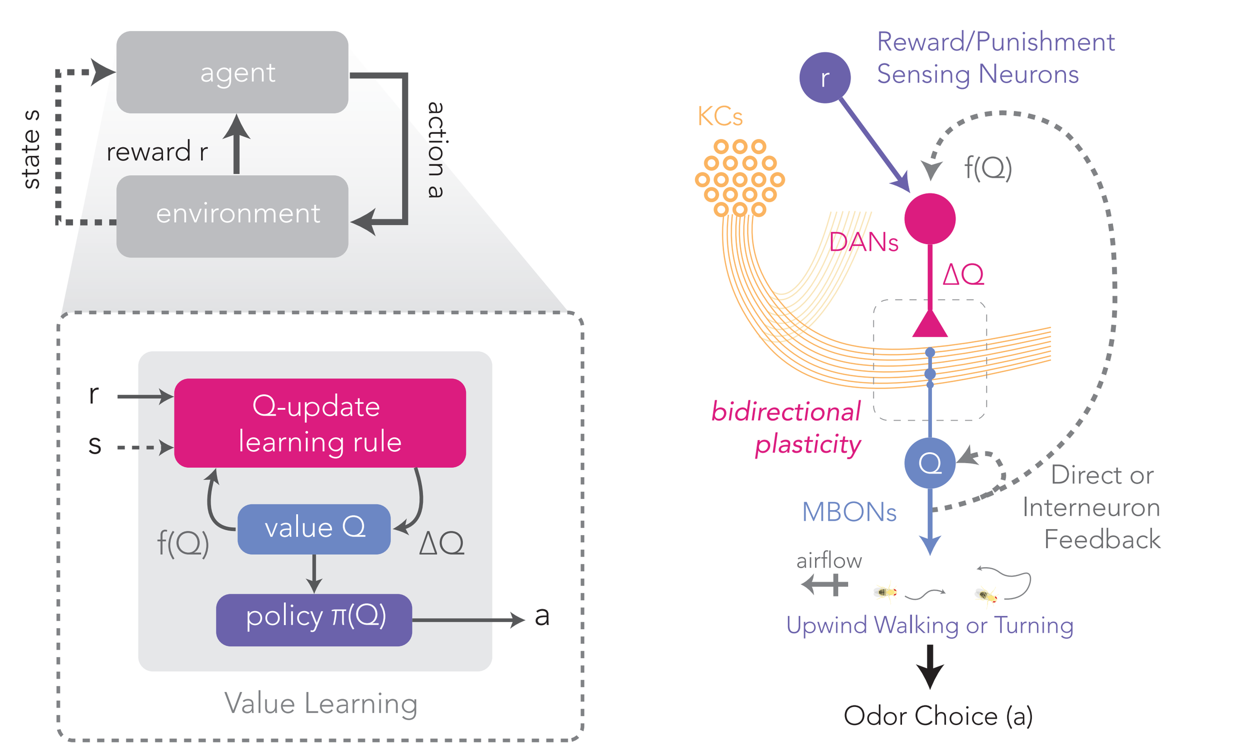
Strategy	Reward: ●	●	●	○	●	●	○	●	●	○	Choice Ratio		Reward Ratio
Persevere on 1st Reward Omission	1	1	1	1	1	1	1	2	2	2	7/10	≈	5/7
Switch on every Reward Omission	1	1	1	1	2	2	2	1	1	1	7/10	≈	5/7

Matching oversimplifies the behavior and throws away important information.

What cognitive processes contribute to the learning rule that governs foraging behavior?

Reinforcement Learning to the Rescue

Q-Learning, a variant of Reinforcement Learning can easily be mapped to the learning and memory circuit of the fruit fly brain to create a library of learning rules.



Value learning framework

$$Q_{t+1} = f(Q_t, C_{1...t}, R_{1...t})$$
$$C_{t+1} \sim \pi(Q_t)$$

where $Q_{t+1}(odor)$ = odor value at trial t+1
C_{t+1} = chosen odor at trial t+1
R_{t+1} = reward recieved at trial t+1
$\pi(x)$ = policy that maps value to p(action)

'Cognitive' Value Learning Algorithms

Classical 'Q Learning' Variants

$$Q_{t+1}(odor) = \begin{cases} (1-\alpha')Q_t + \alpha(R_t + \gamma\tilde{Q}) & \text{if chosen and rewarded} \\ (1-\tau)Q_t + \alpha(\Theta + \gamma\tilde{Q}) & \text{if chosen and unrewarded} \\ (1-\kappa)Q_t & \text{if not chosen} \end{cases}$$

Habit-Value Arbiter 'Q Learning' Variants

$$V_{t+1}(odor) = \begin{cases} (1-\alpha_v)V_t + \alpha_v(R_t + \gamma\tilde{V}) & \text{if chosen} \\ (1-\kappa_v)H_t & \text{if not chosen} \end{cases}$$

$$H_{t+1}(odor) = \begin{cases} (1-\kappa_h)H_t + \alpha_v & \text{if chosen} \\ (1-\kappa_h)H_t & \text{if not chosen} \end{cases}$$

$$Q_t(odor) = w_t\theta_V V_t + (1-w_t)\theta_H H_t$$

$$w_t = \sigma(w_V MAD(V_{1...t}) + w_H MAD(H_{1...t}) + w_b)$$

'Behavioral' Policy

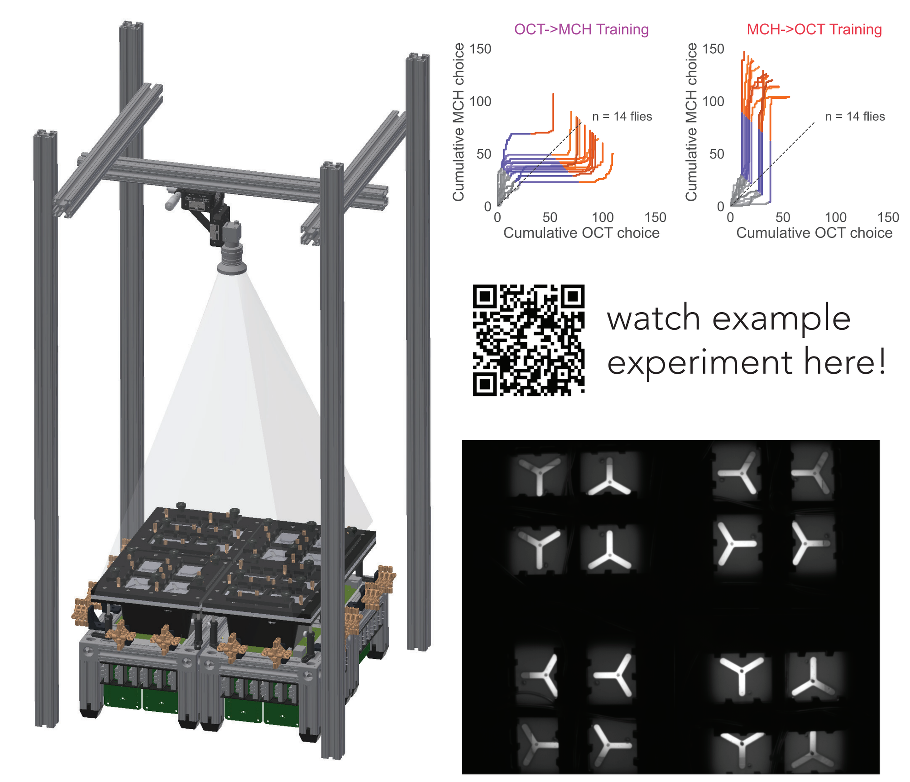
Accept-Reject Policy

$$q_i = \sigma(mQ_i + c), \ i = 1,2$$

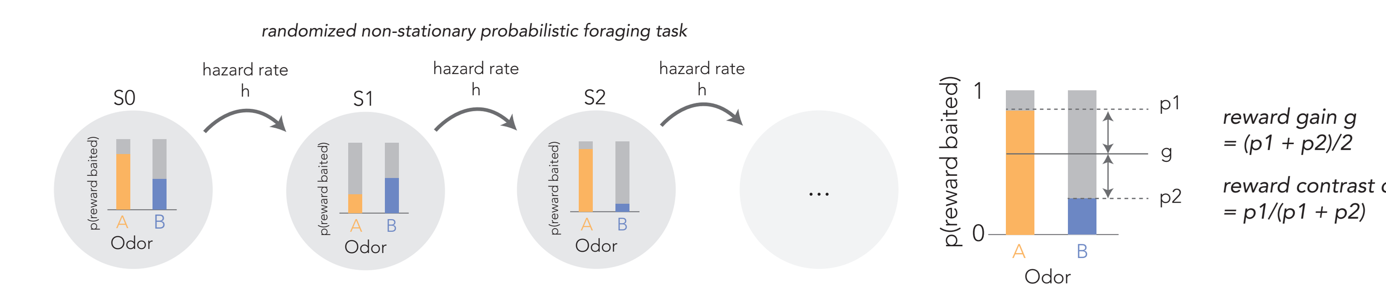
$$[\pi(Q)]_i = \frac{q_i(3-q_{i'})}{3q_i + 3q_{i'} - 2q_iq_{i'}}$$

$$i, i' = 1,2 \text{ or } 2,1$$

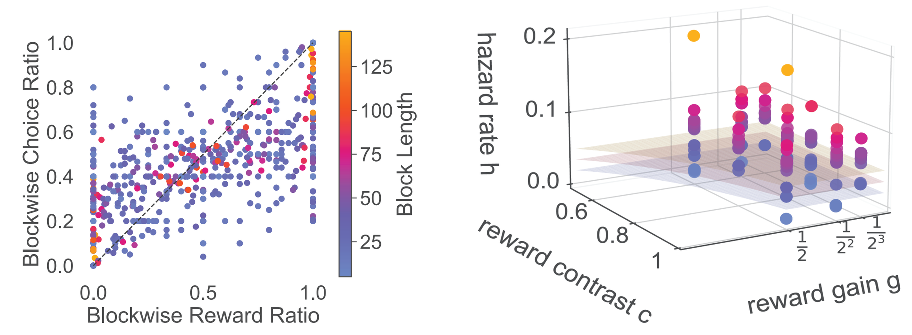
2 | High-throughput Y-Maze Assay

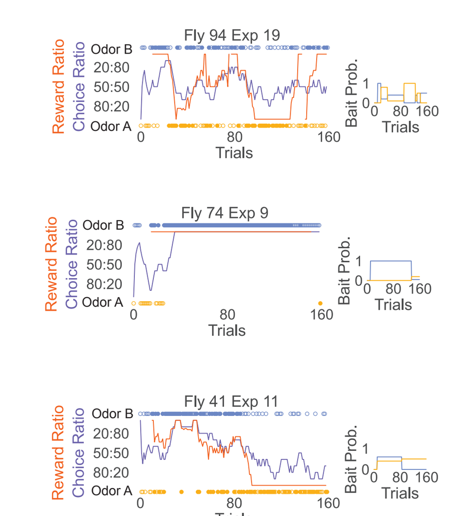


3 | Designing Foraging Experiments



Variable-Block experiments designed to span the space of 2AFC tasks show operant matching behavior.

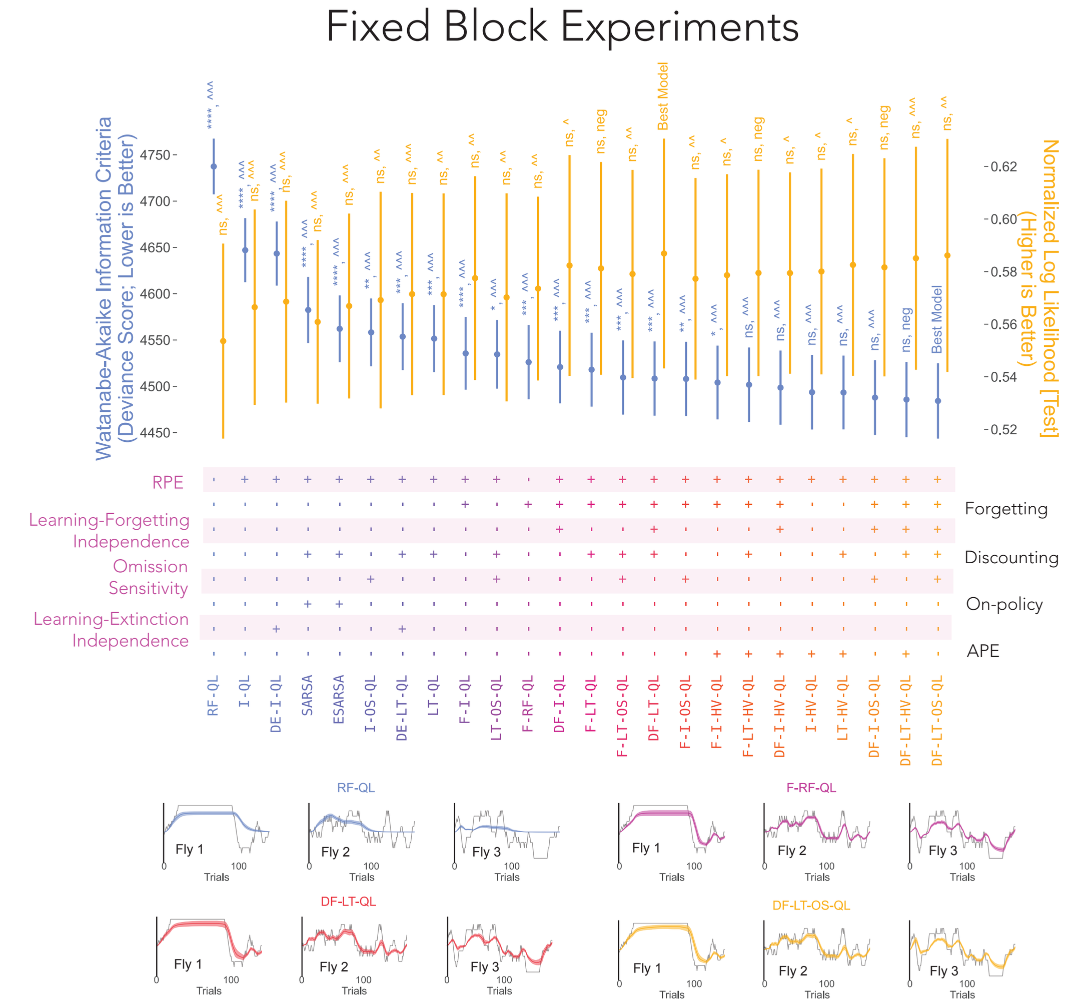




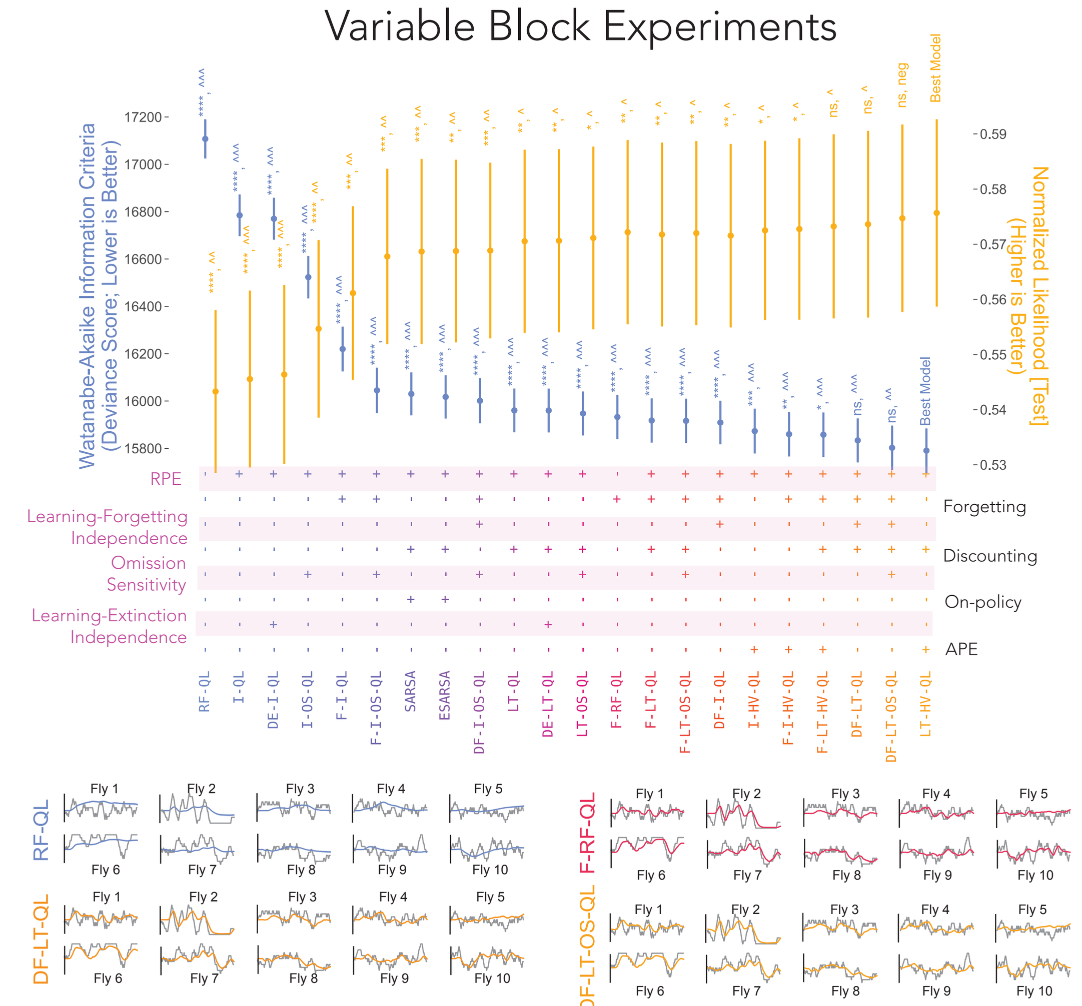
4 | Value Learning Rule Library

Q-Learning models of Rajagopalan et al. (2022) Fixed Block experiments and Variable Block experiments both reveals that including learning-independent forgetting, perseverance, and temporal discounting in the value update improves the model's explanatory power.

Fixed Block Experiments

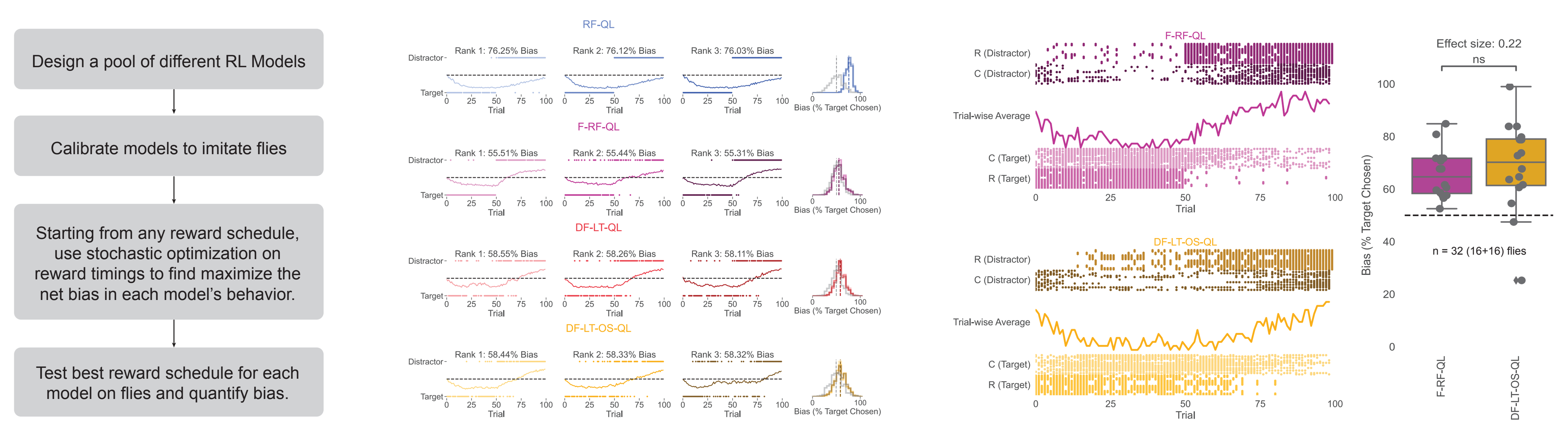


Variable Block Experiments



5 | Evaluating Learning Rules by Biasing flies with Choice Engineering

We utilize **"choice engineering"** to maximally bias preference between two alternatives (Dan & Loewenstein, 2019) in order to test between different learning rules

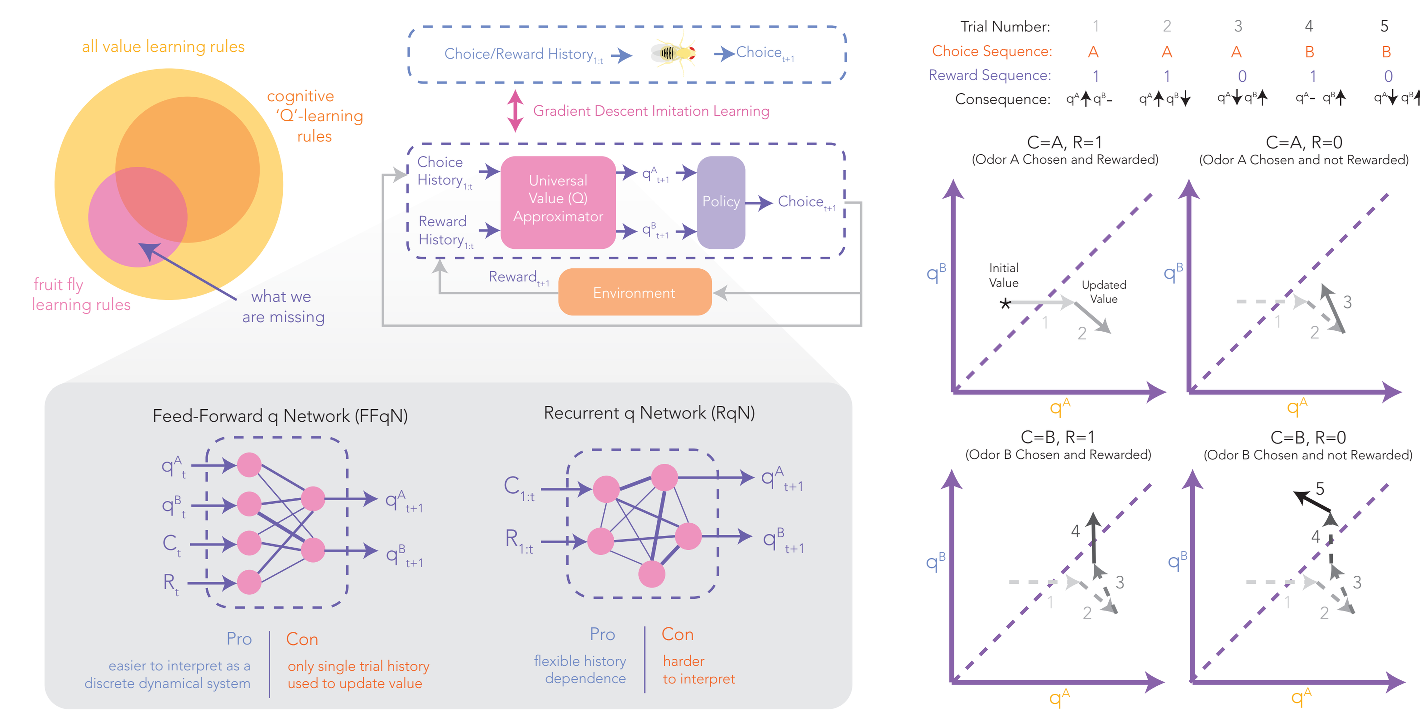


Different learning rules show maximum bias with different reward schedules and predict weak bias.

On comparing two models (excluding and including reward prediction error), we find slightly higher bias for the latter but the results are not statistically significant.

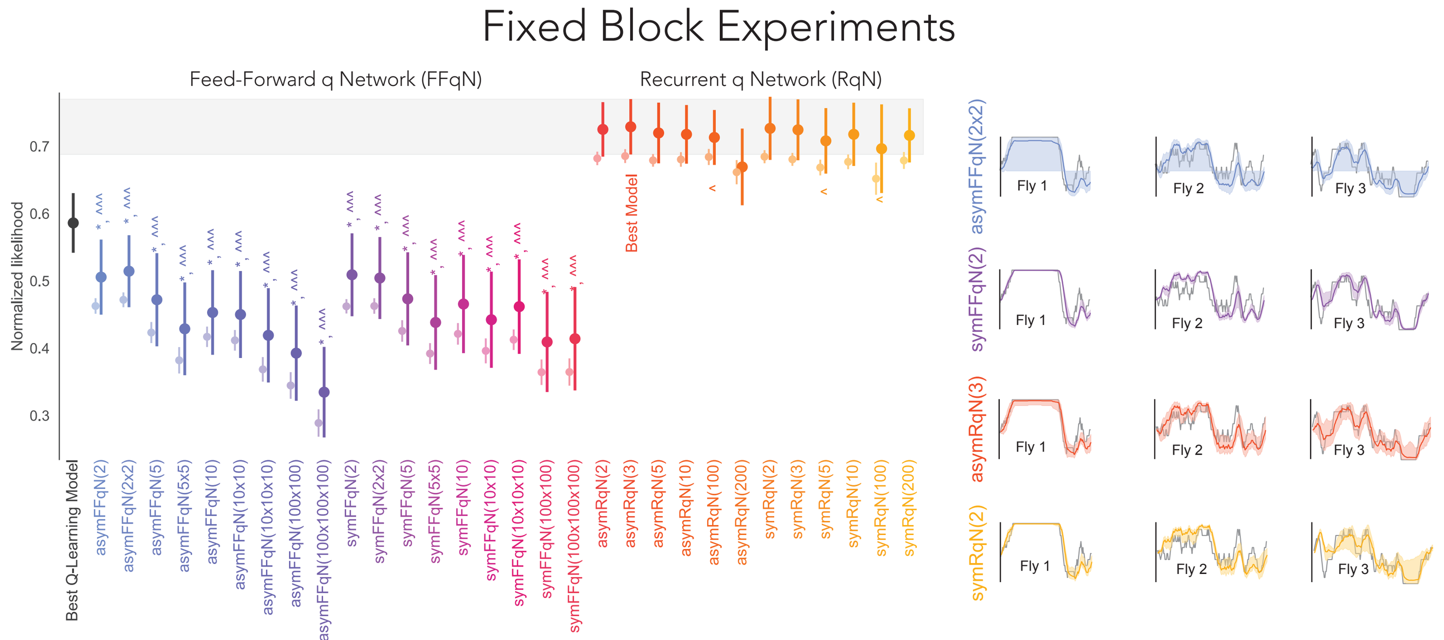
6 | De-novo Learning Rule Discovery

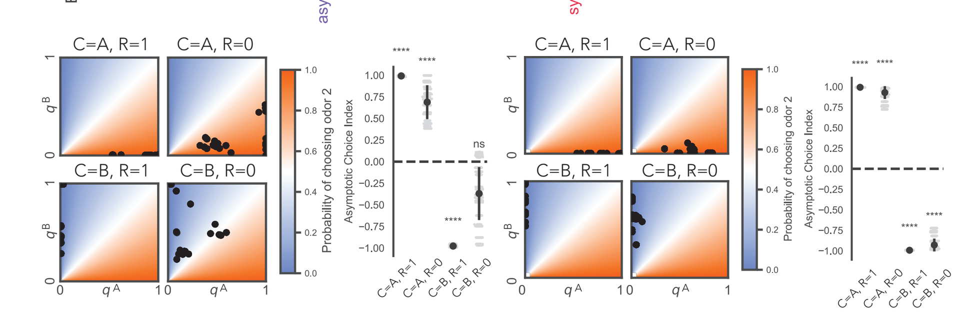
Value learning rule library samples only a small fraction of possibilities.



We use artificial neural networks to create a "universal value approximator". After training them to imitate fly behavior, we analyze them as dynamical systems to understand the learning rules that govern value updates.

Fixed Block Experiments

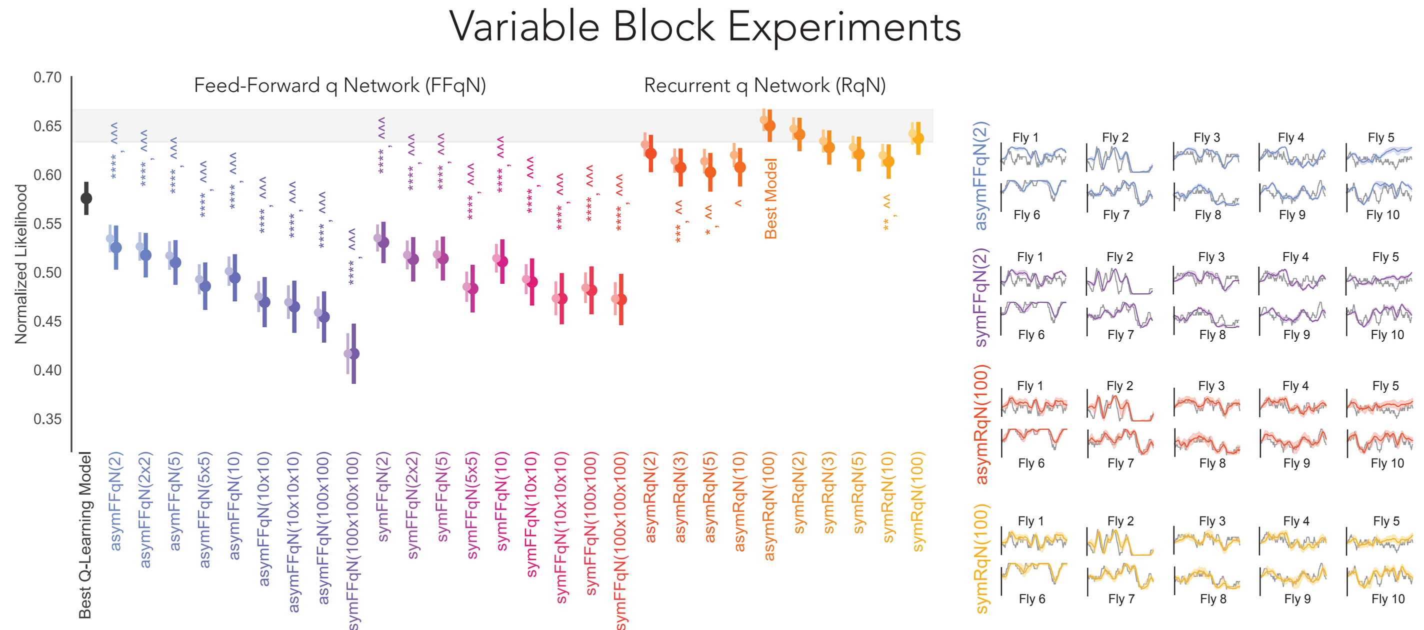


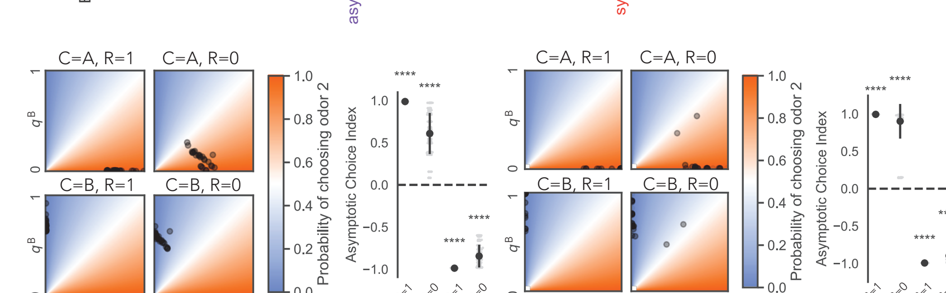


Small FFqNs and RqNs can explain and predict the observed choice behavior

Analysing the dynamics of the FFqNs reveals perservative attractors in the absence of rewards.

Variable Block Experiments





Also, RqNs perform better than the best learning rules in the library.

This suggests that there is more to fly behavior than trial-wise recursive value updates.

References

1. Rajagopalan, Adithya, et al. "Expectation-based learning rules underlie dynamic foraging in Drosophila." bioRxiv (2022).
2. Dan, Ohad, and Yonatan Loewenstein. "From choice architecture to choice engineering." Nature communications (2019).

Acknowledgements

We thank Adithya Rajagopalan and the members of the Turner and Aso Labs at HHMI Janelia for their valuable feedback throughout the duration of the project. We thank the Janelia Experimental Technology Team and FlyCore at Janelia for their support with the rig design/fabrication and fly husbandry. We also thank Dr Kevin Miller and Dr Maria Eckstien from Google Deepmind, Dr. Jan Funke, Dr. James Fitzgerald, Dr. Srini Turaga, Dr. Tzushuan Ma, Maanasa Natarajan and Yash Mehta from Computation & Theory at Janelia, and Dr. Collins Assisi from IISER Pune for their mentorship and support. Finally, We thank HHMI and KVPY Fellowship for funding and IISER Pune for the studentship that allowed me to conduct this project.