Discrete and continuous strategies for working memory

Jorge Yanar[1], Jesse Kaminsky[1], Carlos Brody[1,2]

jyanar@princeton.edu | [1]Princeton Neuroscience Institute, [2]Howard Hughes Medical Institute

Funding
Supported by NIH grant U19NS104648 and the NSF Graduate Research Fellowship (J.Y.)

PRINCETON UNIVERSITY

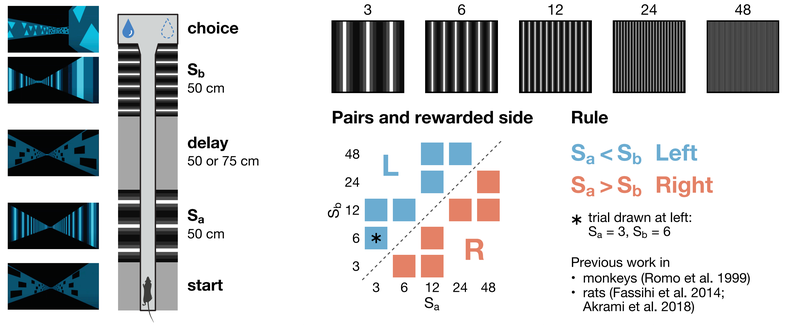
hhmi Howard Hughes Medical Institute

How does the brain hold and compare continuous values in working memory?

- Mice compare two spatial frequencies separated by a delay and report which was higher
- Mice differ in how they use the first stimulus, ranging from graded to step-like (low vs. high) choice strategies
- Delay-period neural encoding of the first stimulus reflects this structure

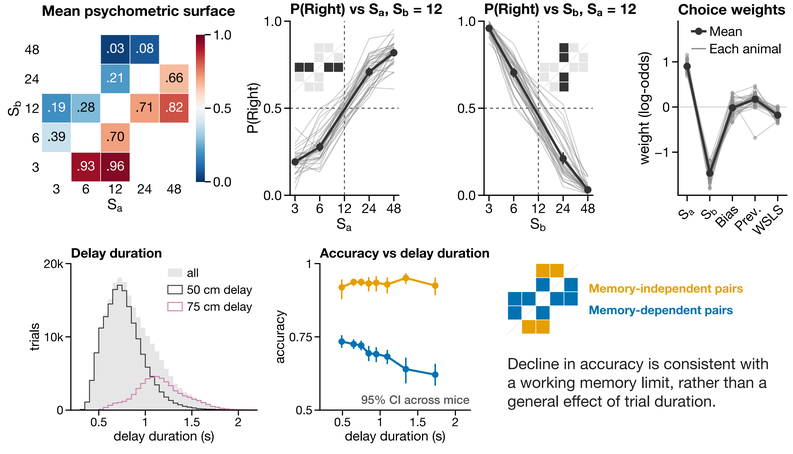
A delayed magnitude comparison task for mice in VR

Mice compare two spatial frequencies, S_a then S_b, and turn left or right for reward. On each trial, the pair (S_a, S_b) is drawn randomly from a set of 12.



Mice use both stimuli, and longer delays only impact memory-dependent pairs

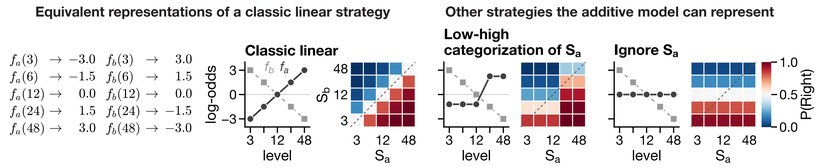
N = 38 mice, 223,247 trials. All sessions on final stage meeting ≥75% accuracy.



An additive choice model captures linear, categorical, and S_b-only strategies

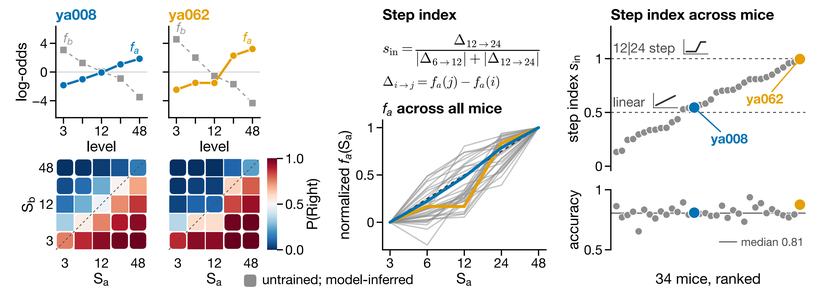
$$\text{logit}\, P(\text{Right}) = w_0 + f_a(S_a) + f_b(S_b)$$

f_a and f_b are scalar look-up tables: each S_a and S_b contributes independently to the log odds. We make no assumption about their shape.



Choice model fits to mice reveal a continuum between linear and category-like treatment of S_a

Model fits to engaged-state trials (GLM-HMM; see next section). Many mice exhibit a large step between S_a 12 and 24. We quantify this with a "step index."

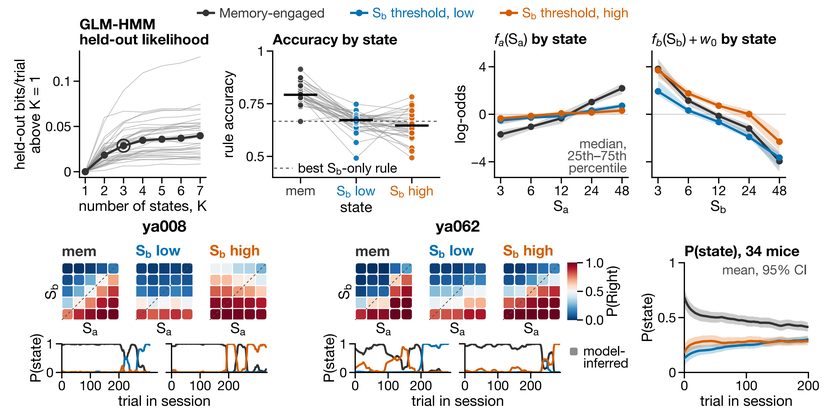


Mice alternate between memory-engaged and memory-disengaged states

We extend the additive choice model via a GLM-HMM approach (Ashwood et al. 2022).

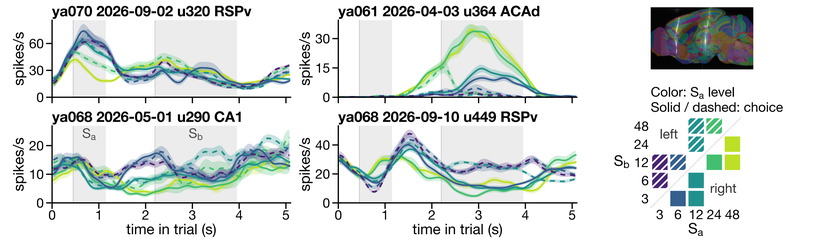
$$\text{For each state } k, \qquad \text{logit}\, P(R \mid k) = w_0^{(k)} + f_a^{(k)}(\mathrm{S_a}) + f_b^{(k)}(\mathrm{S_b})$$

N = 34 mice, 578,522 trials. All trials after each mouse's first day at ≥75% accuracy on the final stage, including low-accuracy days.



Single units across cortex, hippocampus, and thalamus mix S_a, delay, and choice signals

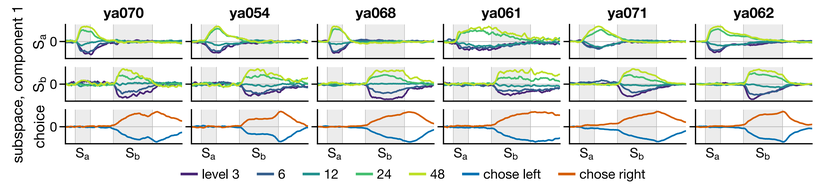
Chronic neuropixels 2.0 probes targeting medial prefrontal cortex (M2, cingulate, prelimbic, and infralimbic cortex) and posterior regions (retrosplenial cortex, hippocampus, and thalamus).



Model-based targeted dimensionality reduction separates S_a, S_b, and choice subspaces

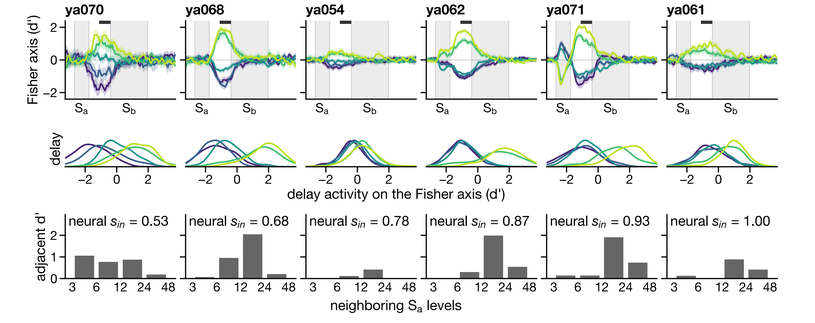
mTDR: Aoi et al. 2020

Held-out trials projected onto first component of each subspace. All recorded units.



In the delay, the neural code for S_a ranges from linear to category-like across mice

We project held-out neural activity onto the full S_a subspace and find the axis that best separates S_a levels in the late delay (Fisher discriminant).



Across mice, the delay-period S_a code tracks the behavioral step index

