Reinforcement learning constrained state space modeling of neural decisions

UNIVERSITY of WASHINGTON COMPUTATIONAL NEUROSCIENCE CENTER

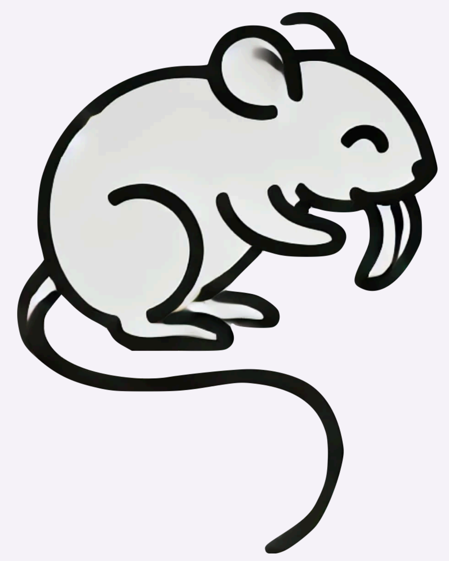
Yusi Chen[1], Shijia Liu [4], Jeremiah Cohen [3], Eric Shea Brown [1,2]

[1]Computational Neuroscience Center, University of Washington; [2] Applied Mathematics, University of Washington; [3] Allen Institute for Neural Dynamics; [4] Harvard Medical School

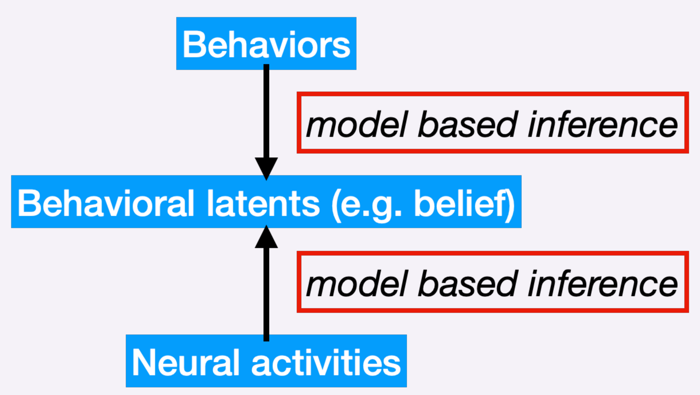
UNIVERSITY of WASHINGTON

Abstract

- State space modeling (SSM) enables latent variable inference and system parameter estimation.
- We proposed **reinforcement learning constrained** SSM [1] to model animal licking.
- Enables the inference of **internal belief** of animals and **data driven discovery of learning rules**.
- Improves upon previous methods by harnessing **within trial details** (e.g. response time, lick number etc.)



Motivation: modeling within trial belief dynamics



Basic assumptions:

- Latent variables decay along time
- Licking as a point process
- Latent variables get updated per lick

Modeling licking as a point process

- Definition:
  - $N(t)$: the number of events up until t;
  - $s(t) = \{s_0, s_1, .., s_{N(t)}\}$: event timing set;
  - $\lambda(t|s(t)) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(N(t+\Delta t) - N(t) = 1|s(t))$: event rate.
- When $N(T) = n$, the probability density function of spike time: $p(s_1, s_2, ..., s_n) = \prod_{k=1}^{n} \lambda(s_k|s_1, .., s_{k-1}) \exp[-\int_0^T \lambda(u|s(u))du]$
- Inhomogenous Markov process (IMI) [2] only considers the influence of most recent event $s_*(t)$
  - $\lambda(t|s(t))$ simplify to $\lambda(t, t - s_*(t))$.
  - Homogeneous Poisson process: $\lambda(t, t - s_*(t)) = \lambda$, $\mathbb{E}[N[0,T)] = T\lambda$, Variance, Fano factor, ITI;
  - Inhomogeneous Poisson process: $\lambda(t, t - s_*(t)) = \lambda(t)$, $\mathbb{E}[N[0,T)] = \int_0^T d\xi \lambda(\xi)$;
  - Inhomogeneous Poisson process with refractory period: $\lambda(t, t - s_*(t)) = \lambda_1(t)\delta(t - s_*(t) > \tau)$

Reinforcement learning constrained state-space modeling

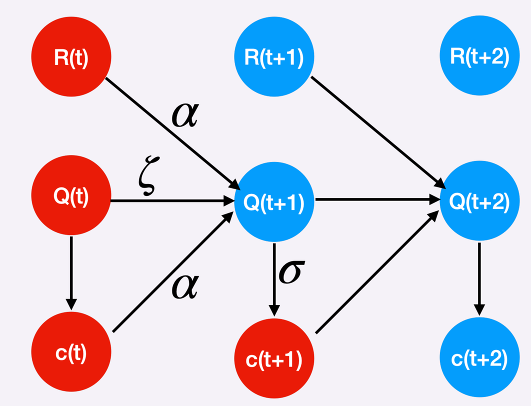
Model assumption	Forward modeling	Constrained state space modeling
Rescorla-Wagner	$Q_{t+1} = \zeta Q_t + \alpha C_t(R_t - Q_t)$ $P(C_t) = \sigma(Q_t)$	$Q_{t+1} \sim \mathcal{N}(\zeta Q_t + \alpha C_t(R_t - Q_t), V)$ $C_t \sim$ Bernoulli$(\sigma(Q_t))$
+ choice kernel	$Q_{t+1} = \zeta Q_t + \alpha C_t(R_t - Q_t)$ $K_{t+1} = \zeta K_t + \alpha C_t(C_t - K_t)$ $P(C_t) = \sigma(\beta_1 Q_t + \beta_2 K_t)$	$L_t = [Q_t, K_t]^T$ $L_{t+1} \sim \mathcal{N}(\begin{bmatrix} \zeta - \alpha C_t & 0 \\ 0 & \zeta - \alpha C_t \end{bmatrix} L_t + \begin{bmatrix} \alpha C_t R_t \\ \alpha C_t C_t \end{bmatrix}, V)$ $C_t \sim$ Bernoulli$(\boldsymbol{\beta}^T L_t)$

Q_t: internal belief; K_t: choice kernel; R_t: reward; $C_t \in \{0,1\}$: decision

Highlight

We applied theory-constrained state space modeling to infer with-in trial belief states from animal behavior and estimate system parameters, advancing the discovery of novel learning theories.

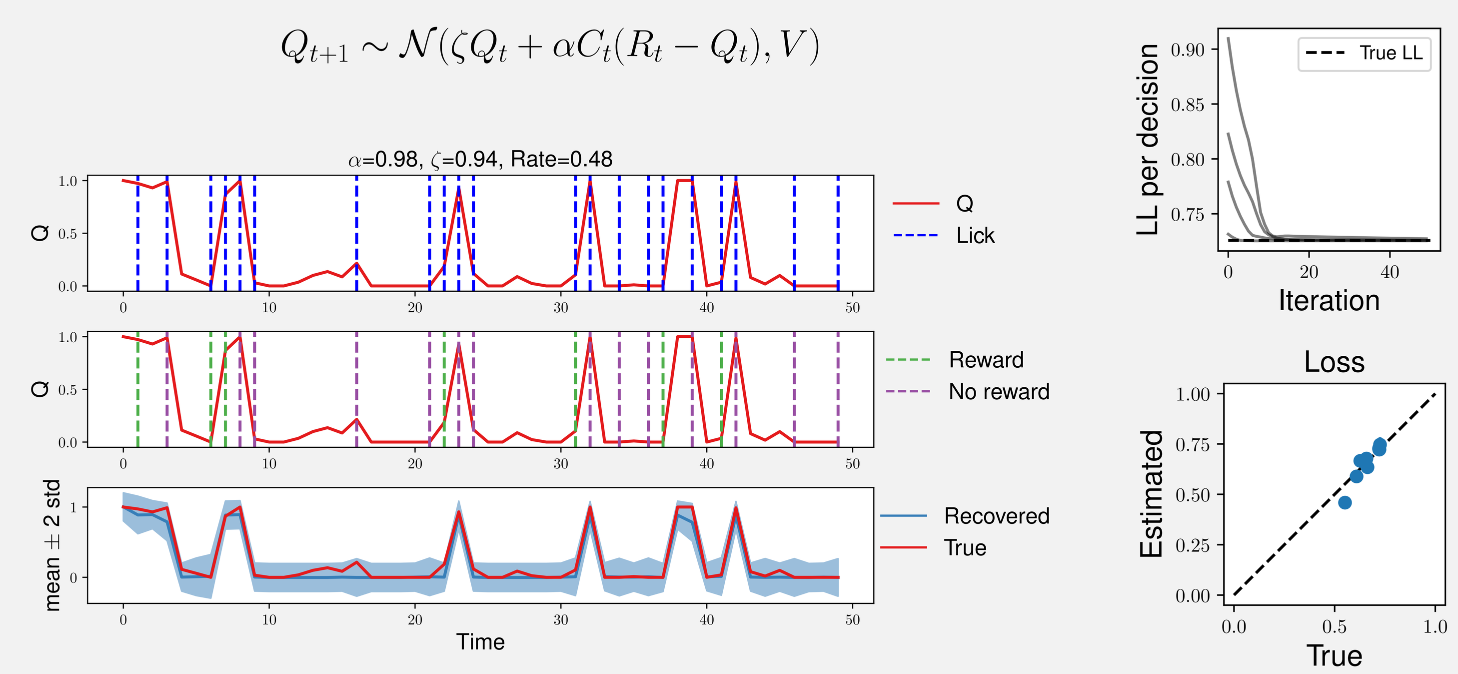
Implementation



- Latent inference: Given $R_{0:t}, C_{0:t+1}$, infer latent Q_{t+1} such that $P(Q_{t+1}|R_{0:t}, C_{0:t+1})$ is maximized.
- Estimated parameter $\alpha^*, \zeta^* =$ argmax $P(C_{t+1}|R_{0:t}, C_{0:t+1}; \alpha, \zeta)$
- Implementation: DYNAMAX, approximate Bernoulli pdf through a mixture of Gaussians.

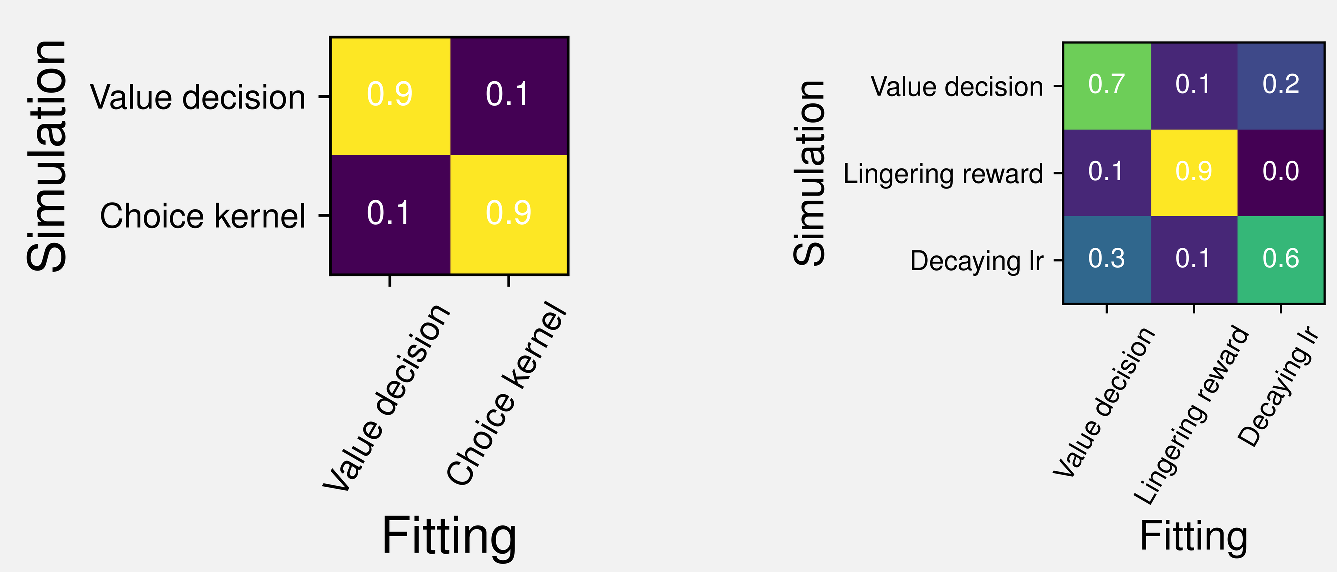
Identify the ground truth parameters

$$Q_{t+1} \sim \mathcal{N}(\zeta Q_t + \alpha C_t(R_t - Q_t), V)$$

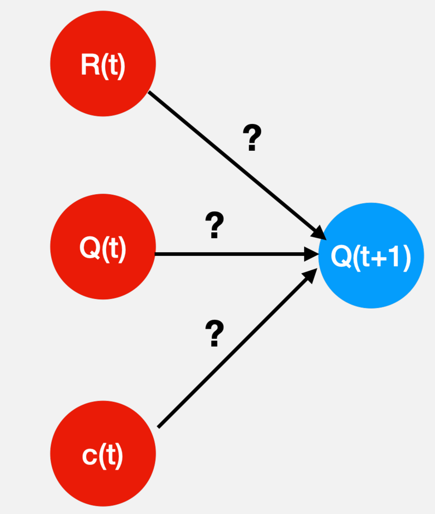


Distinguish between models of latent dynamics

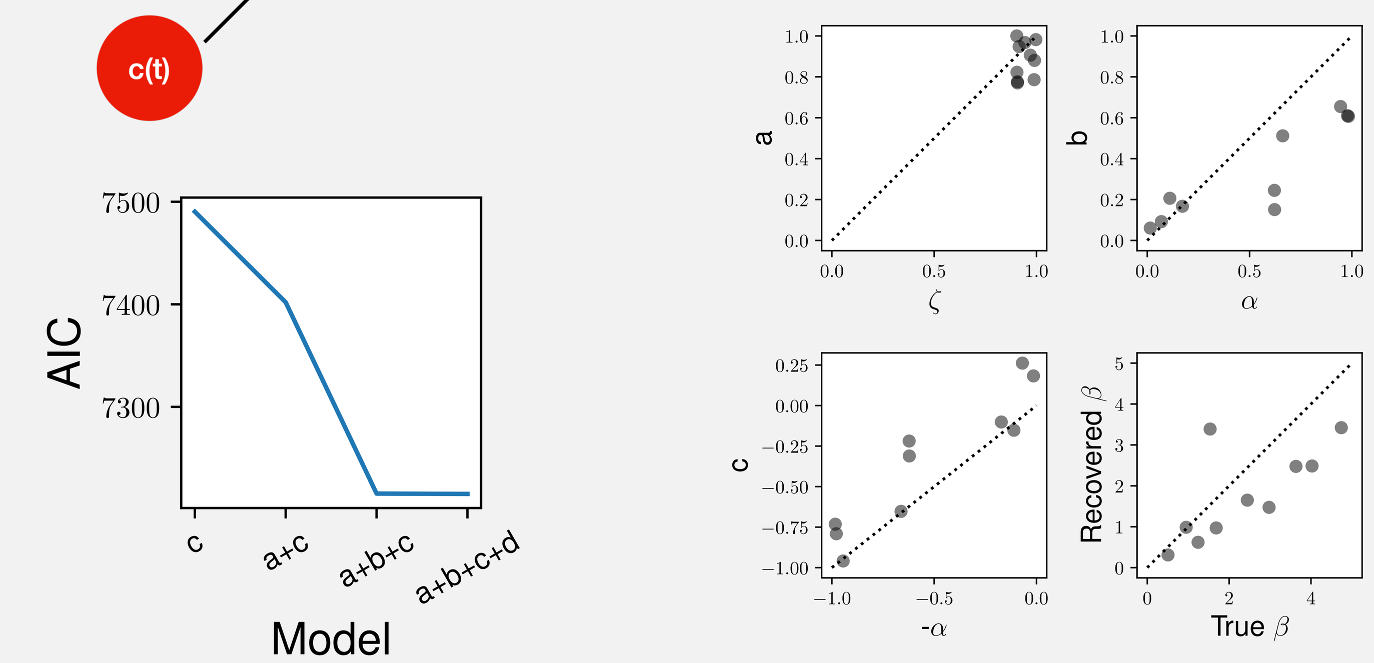
- Value decision: $Q_{t+1} \sim \mathcal{N}(\zeta Q_t + \alpha C_t(R_t - Q_t), V)$
- Choice kernel: $K_{t+1} \sim \mathcal{N}(\zeta K_t + \alpha(C_t - K_t), V)$
- Lingering reward: $Q_{t+1} \sim \mathcal{N}(\zeta Q_t + \alpha C_t(R_t - Q_t), V), R_{t+1} = \eta R_t + r_t$
- Decaying learning rate: $Q_{t+1} \sim \mathcal{N}(\zeta Q_t + \alpha_t C_t(R_t - Q_t), V), \alpha_{t+1} = \eta \alpha_t$



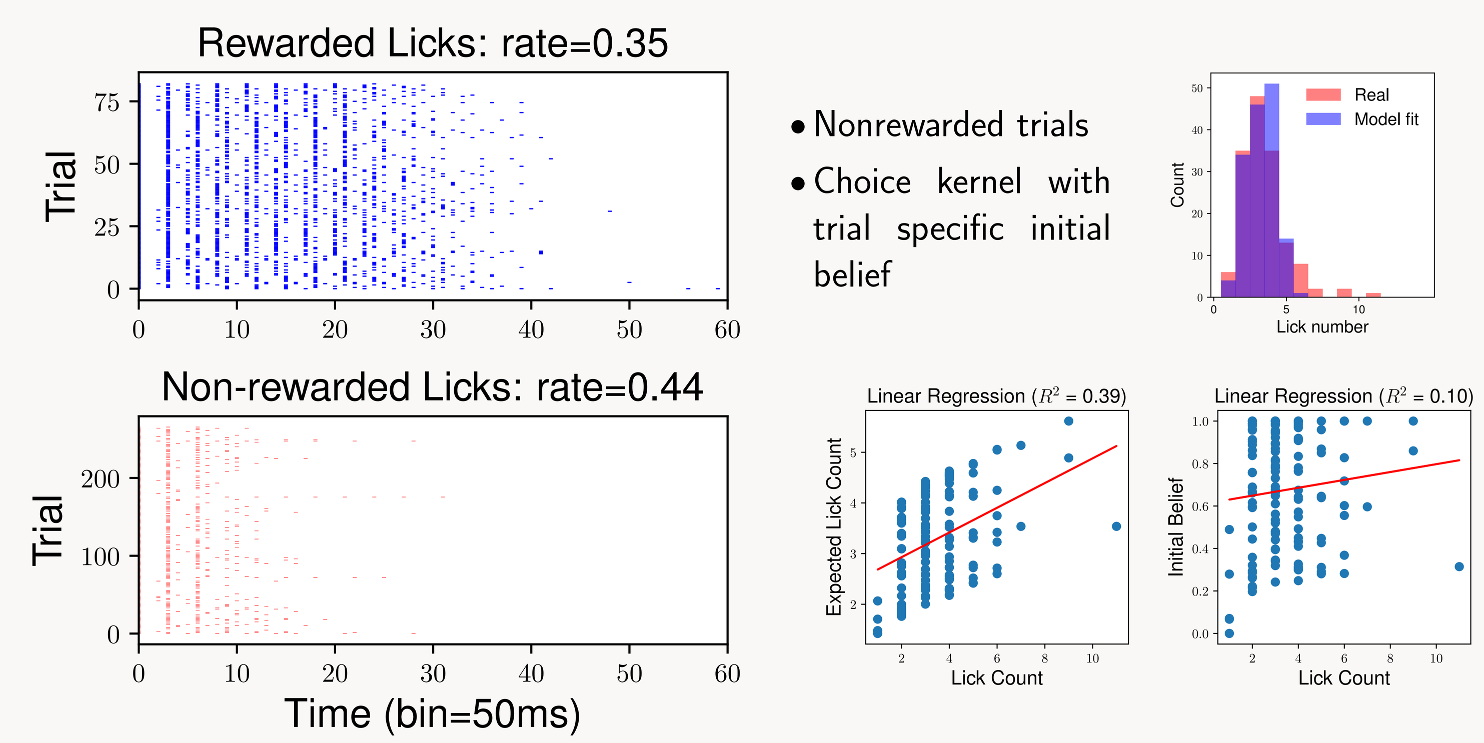
Identify novel learning rules



- Linear relationship: $\mathbb{E}(Q_{t+1}) = aQ_t + bC_t + cR_t$
  - Contains choice kernel model
- Bilinear relationship: $\mathbb{E}(Q_{t+1}) = aQ_t + bC_tR_t + cC_tQ_t + dR_tQ_t$
  - Contains the value decision model
- Generative dynamics: $\mathbb{E}(Q_{t+1}) = \zeta Q_t + \alpha C_t(R_t - Q_t)$
  - Ground truth relationship: $a = \zeta$, $b = \alpha$, $c = -\alpha$



Application to real data and challenges



- Nonrewarded trials
- Choice kernel with trial specific initial belief

Extensions

- Formulate loss function as IMI: $N[s,t) \sim$ Poiss(n, θ)
- Hierarchical modeling of trial-level parameters: $L(\mu, \sigma) = \mathbb{E}_{\theta \sim \mathcal{N}(\mu, \sigma^2)}MAP(\theta)$

References

1. Linderman, S. W. *et al.* Using computational theory to constrain statistical models of neural data. *Current opinion in neurobiology* **46,** 14–24 (2017).
2. Kass, R. E. *et al.* A spike-train probability model. *Neural computation* **13,** 1713–1720 (2001).

Acknowledgements

We'd like to thank Swartz foundation, UW-CNC and Allen Institute for Neural Dynamics.