# Dynamical phases of short-term memory mechanisms in RNNs

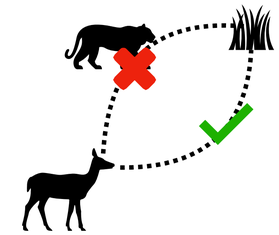
Bariscan Kurtkaya[*1,2], Fatih Dinc[*2,3,4], Mert Yuksekgonul[5], Marta Blanco-Pozo[2,6], Ege Cirakman[2], Mark Schnitzer[2,6,7], Yucel Yemez[1], Hidenori Tanaka[†8,9], Peng Yuan[†10], Nina Miolane[†3]

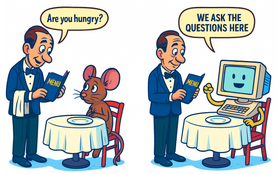
[1]Koc University, [2]CNC Program, Stanford University, [3]Geometric Intelligence Lab, UC Santa Barbara, [4]Kavli Institute for Theoretical Physics, UC Santa Barbara, [5]Computer Science, Stanford University, [6]James H. Clark Center for Biomedical Engineering & Sciences, Stanford University, [7]Howard Hughes Medical Institute, Stanford University, [8]Center for Brain Science, Harvard University, [9]Phi Lab, NTT Research, [10]Institute for Translational Brain Research, Fudan University

Geometric Intelligence Lab, UC Santa Barbara · Stanford University · Fudan University · Koç University · Harvard University

## What is short-term memory?

Short-term memory is a core cognitive mechanism that allows us to temporarily retain information over brief periods—such as remembering how much money we just gave a cashier or tracking a predator's path to avoid becoming prey. Most importantly, understanding this mechanism is crucial, as it may shed light on several neurological and psychiatric conditions, including Alzheimer's disease.






## Why do we conduct simulations instead of wet-lab experiments?

- Wet-lab experiments are often costly, time-consuming, and difficult to scale. Simulations allow us to test hypotheses efficiently before committing to biological studies.
- Due to rapid synaptic reorganization following stimuli or rewards, identifying the underlying mechanisms of memory in vivo is highly challenging.
- Simulations also offer precise control over environmental factors—such as hunger or thirst—that are difficult to regulate in animal experiments.

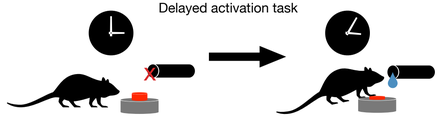
## What tasks do we study?

To examine short-term memory mechanisms, we focused on two classical neuroscience tasks: delayed activation and delayed cue-discrimination. While the delayed activation task isolates the memory component, the cue-discrimination task combines memory with decision-making. Additionally, we explored the impact of a post-reaction period, which is rarely included in RNN training. This extension significantly altered the learned dynamics and reshaped the phase diagram.

**Delayed Activation:**

$$u_t = \varnothing$$

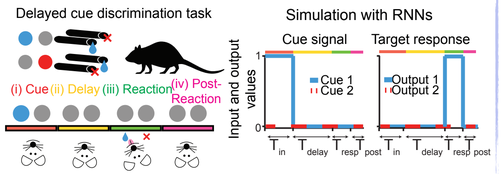
$$o_t^* = \begin{cases} 1, & \text{if } t \in T_{\text{resp}} \\ 0, & \text{otherwise} \end{cases}$$




**Delayed Cue-Discrimination Task:**

$$u_t = \begin{cases} (0,1) \text{ or } (1,0), & \text{if } t \in T_{\text{in}} \\ (0,0), & \text{otherwise} \end{cases}$$

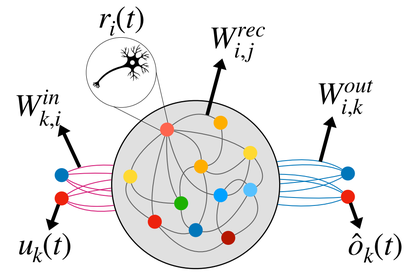
$$o_t^* = \begin{cases} u_{T_{in}}, & \text{if } t \in T_{\text{resp}} \\ (0,0), & \text{otherwise} \end{cases}$$







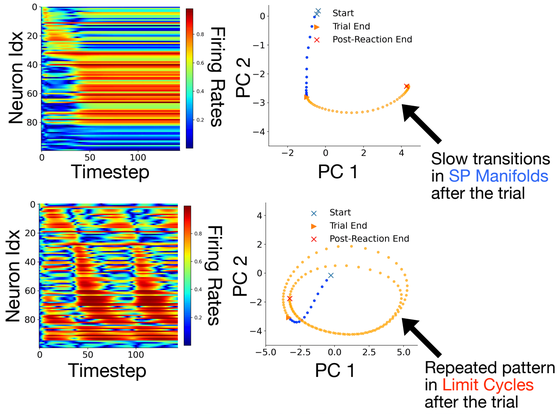
## Leaky firing-rate RNNs




$$\tau \dot{r}(t) = -r(t) + \tanh(W^{rec} r(t) + W^{\text{in}} u(t) + b + \epsilon)$$

$$\hat{o}(t) = f\left(W^{\text{out}} r(t) + b_{\text{out}}\right)$$

## Slow-point manifolds / Limit Cycles




## In a nutshell...

- We identified two distinct mechanisms for maintaining short-term memory: slow-point manifolds that support smooth sequences, and limit cycles that generate temporally localized dynamics.
- We derived theoretical scaling laws linking the learning rate and delay length, predicting the emergence of each mechanism and defining critical thresholds beyond which learning fails.
- We demonstrate that small changes in task structure—such as adding a post-response period—can significantly alter the underlying phase space and learned dynamics.
- We propose testable predictions to guide wet-lab neuroscience experiments in identifying memory-related dynamics in the brain.

## Scaling Law Analysis

Ghost Model [2] for SP Manifold

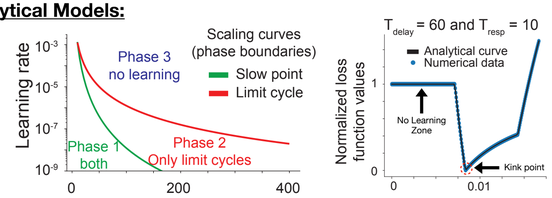
$$\tau \dot{x}(t) = x^2(t) + r \quad \rightarrow \quad \alpha^*_{\text{GM-as}} = \frac{\pi^4 T_{\text{resp}}}{4} T^{-5}_{\text{delay}}$$

Due to the pathological nature of the loss function, there exists a critical learning rate below which the model avoids a no-learning zone [2].

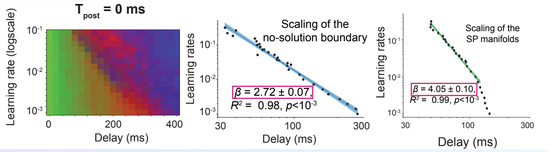
Limit Cycle Model for Limit Cycles

$$\tau \dot{\rho}(t) = (1 - \rho^2(t))\rho(t),\ \tau \dot{\theta}(t) = 2\pi r \quad \rightarrow \quad \alpha^*_{\text{LC-as}} = \frac{T_{\text{resp}}}{4} T^{-3}_{\text{delay}}$$

**Analytical Models:**



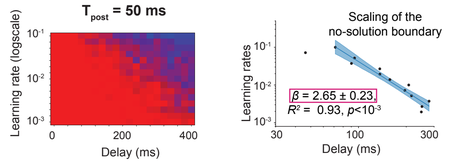

**Full-Rank RNNs:**




Similar to our previous analytical analysis, we examine comparable scaling laws in full-rank RNN experiments. Notably, while analytical models cannot escape the no-learning zone, RNNs can escape from no-learning zone—yet they still exhibit similar scaling behavior.

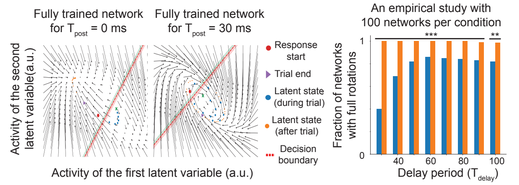
## Post-Reaction Analysis

**Full-Rank RNNs:**




Adding a post-response period to the delayed cue-discrimination task significantly change the phase space—it eliminates slow-point manifolds and restricts the model to converge to limit cycles. However, it does not lead to a significant difference in the scaling of the no-solution boundary.

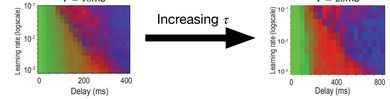
**Low-Rank RNNs:**




To examine the vector field of firing rate dynamics, we conducted experiments on low-rank RNNs using delayed activation tasks. Similar to our previous findings, adding a post-reaction period shifts the phase space, causing convergence to limit cycles instead of slow-point manifolds.

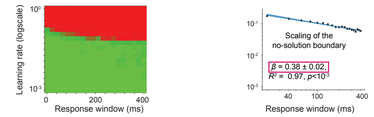
## Additional Analysis

**Longer $\tau$ experiment:**




To examine the effect of the leak coefficient ($\tau$) in our experiments, we increased its value from 10 milliseconds to 20 milliseconds. As we increased the delay length proportionally with $\tau$, the phase diagram remained unchanged.

**No delay experiment:**




To verify the importance of the delay component, we removed the delay period from the task and adjusted the response period accordingly. As a result, the scaling law disappeared—highlighting the critical role of the delay in shaping the learned dynamics.

## References

[1] Rajan, Kanaka, Christopher D. Harvey, and David W. Tank. "Recurrent network models of sequence generation and memory." Neuron 90.1 (2016): 128-142.

[2] Dinc, Fatih, et al. "A ghost mechanism: An analytical model of abrupt learning." arXiv preprint arXiv:2501.02378 (2025).

[3] Sussillo, David, and Omri Barak. "Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks." Neural computation 25.3 (2013): 626-649.

[4] Valente, Adrian, Jonathan W. Pillow, and Srdjan Ostojic. "Extracting computational mechanisms from neural data using low-rank RNNs." Advances in Neural Information Processing Systems 35 (2022): 24072-24086.

[5] Haputhanthri, Udith, et al. "Understanding and controlling the geometry of memory organization in RNNs." arXiv preprint arXiv:2502.07256 (2025).

[6] Dinc, Fatih, et al. "Latent computing by biological neural networks: A dynamical systems framework." arXiv preprint arXiv:2502.14337 (2025).

## Website