To attract or to oscillate: Validating dynamics with behavior

Keith T. Murray
Massachusetts Institute of Technology, Cambridge, MA 02139
ktmurray@mit.edu

Recurrent neural networks (RNNs) have driven significant advancements in computational neuroscience, as evidenced by studies showcasing their ability to emulate observed neural dynamics in behaving animals [1]–[4]. These findings substantiate RNNs as robust models for neural systems and reinforce the efficacy of an optimization framework within computational neuroscience. Recently, it has been shown that RNNs trained with large regularizations learn low-dimensional dynamics that appear to explain how the RNNs solve their respective task [2], [5], [6]. We took inspiration from this research direction to investigate the low-dimensional dynamics learned by regularized RNNs trained on a simple pattern recognition task inspired by the card game SET [7].

$$\tau \dot{x}_i(t) = -x_i(t) + \sum_{k=1}^{N} J_{ik} r_k(t) + \sum_{k=1}^{N^{in}} B_{ik} u_k(t) + b_i + \eta_i(t) \quad (1)$$

$$r_i(t) = \tanh(x_i) \quad (2)$$

The regularizations imposed during training were an L2 regularization on the weights defined in (1) and an L2 regularization on the rates defined in (2).

We found that the learned low-dimensional dynamics resembled operations on a finite-state automaton (FSA) [8]. Surprisingly, we also found that the dynamical implementation of the learned FSA changed depending on the time constant, τ, used in (1). RNNs with higher time constants learned FSA as a network of fixed-point attractors in state space [9]. RNNs with lower time constants learned FSA as a network of phase-angle transitions in the space of phase angles of a limit cycle [10]. Theoretically, RNNs with a slow time constant can respond on a quicker timescale [11]; however, the effect of the imposed regularizations and the selected time constant created a large bias on the types of dynamics learned—attractive or oscillatory.

Previous research often selected the time constant, considered to be a hyperparameter, without extensive justification [5], [6]. Our findings challenge the biological realism of the previously discovered dynamics, as they may result from arbitrary hyperparameter selection—a topic further explored in [12]. To validate the choice of the time constant without relying on experimentally gathered neural data, we propose using behavioral data. We observed that different time constants produced varying false positive error rates for ambiguous patterns in our task. We found this relationship to be a bell-curve shaped psychometric function unique to each RNN. By comparing psychometric data from humans to those from RNNs, we could potentially justify the selected time constants and the learned dynamics. This methodology resembles the approach used in [13].

In summary, our study demonstrates that regularized RNNs trained on a simple pattern recognition task can identify patterns through operations on a FSA, but the dynamical implementation of the FSA, either attractive or oscillatory, is dependent on the chosen time constant. To validate the selection of this constant and the learned dynamics, we suggest comparing psychometric data from humans to data from RNNs. RNNs mirroring human psychometric data most closely could be considered more biologically plausible. This approach could potentially provide a means to infer the timescales of cortical areas without single-neuron spike train data [14].

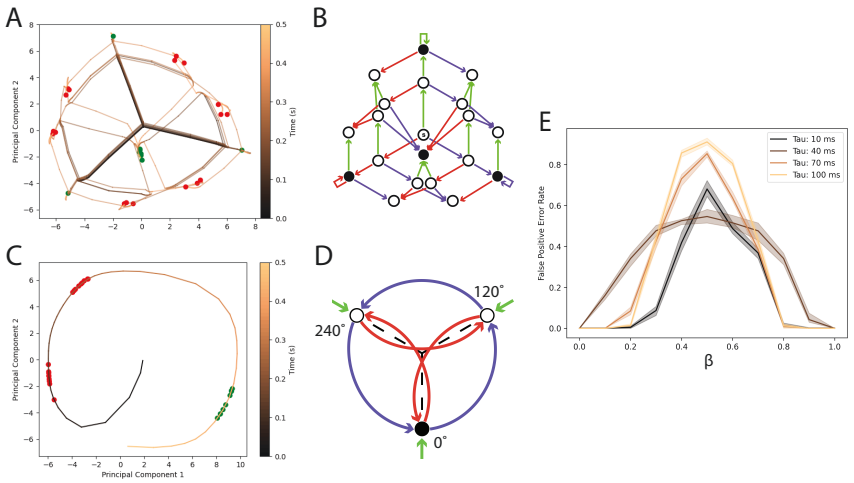

Fig. 1. Effects of τ variation on RNN dynamics and behavior. The task used to train RNNs involved presenting three random vectors, called colors, at randomly selected times and tasking the RNN to determine if the colors were all the same or all different. (A) Dynamics of an RNN (τ = 100 ms) projected onto the first two PCs showing attractive dynamics. Red and green dots indicate invalid and valid patterns, respectively. (B) Corresponding FSA for dynamics from (A). Note the unique encoding directions for each color. Empty and full states represent invalid and valid patterns, respectively. (C) Dynamics for an RNN (τ = 10 ms) showing oscillatory dynamics. (D) Corresponding FSA for dynamics from (C). Note that the FSA is realized in the space of phase angles in the limit cycle identified in (C). Note the unique phase-angle contribution by each color. The starting state is the accepting state. (E) Modulation of τ in a RNN affects false positive error rates for ambiguous patterns, determined by β.

REFERENCES

[1] V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome, "Context-dependent computation by recurrent dynamics in prefrontal cortex," *Nature*, vol. 503, no. 7474, pp. 78–84, 2013.

[2] D. Sussillo, M. M. Churchland, M. T. Kaufman, and K. V. Shenoy, "A neural network that finds a naturalistic solution for the production of muscle activity," *Nature neuroscience*, vol. 18, no. 7, pp. 1025–1033, 2015.

[3] W. Chaisangmongkon, S. K. Swaminathan, D. J. Freedman, and X.-J. Wang, "Computing by robust transience: how the fronto-parietal network performs sequential, category-based decisions," *Neuron*, vol. 93, no. 6, pp. 1504–1517, 2017.

[4] H. Sohn, D. Narain, N. Meirhaeghe, and M. Jazayeri, "Bayesian computation through cortical latent dynamics," *Neuron*, vol. 103, no. 5, pp. 934–947, 2019.

[5] L. Driscoll, K. Shenoy, and D. Sussillo, "Flexible multitask computation in recurrent networks utilizes shared dynamical motifs," *bioRxiv*, 2022.

[6] K. Kay, X. Wei, R. Khajeh, M. Beiran, C. J. Cueva, G. Jensen, V. P. Ferrera, and L. F. Abbott, "Neural dynamics and geometry for transitive inference," *bioRxiv*, 2022.

[7] R. V. E. Gordon, G. Gordon, and H. Gordon, "The joy of SET: The many mathematical dimensions of a seemingly simple card game," Princeton University Press, 2017.

[8] A. Cleeremans, D. Servan-Schreiber, and J. L. McClelland, "Finite state automata and simple recurrent networks," *Neural computation*, vol. 1, no. 3, pp. 372–381, 1989.

[9] M. Khona and I. R. Fiete, "Attractor and integrator networks in the brain," *Nature Reviews Neuroscience*, pp. 1–23, 2022.

[10] S. H. Strogatz, "Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, And Engineering," *CRC press*, 2000.

[11] C. Van Vreeswijk and H. Sompolinsky, "Chaos in neuronal networks with balanced excitatory and inhibitory activity," *Science*, vol. 274, no. 5293, pp. 1724–1726, 1996.

[12] R. Schaeffer, M. Khona, and I. Fiete, "No free lunch from deep learning in neuroscience: A case study through models of the entorhinal-hippocampal circuit," *bioRxiv*, 2022.

[13] S. J. Gershman, "Deconstructing the human algorithms for exploration," *Cognition*, vol. 173, p. 34–42, 2018.

[14] J. D. Murray, A. Bernacchia, D. J. Freedman, R. Romo, J. D. Wallis, X. Cai, C. Padoa-Schioppa, T. Pasternak, H. Seo, D. Lee, et al., "A hierarchy of intrinsic timescales across primate cortex," *Nature neuroscience*, vol. 17, no. 12, pp. 1661–1663, 2014.