

Discrete boundaries between neural populations in recurrent neural networks

Jacob C. Tanner (jctanner@iu.edu)

Indiana University
Program in Cognitive Science & Department of Informatics
Bloomington, Indiana, 47401

Sina Mansour L. (sina.mansour.lakouraj@gmail.com)

National University of Singapore

Ludovico Coletta (ludovico.coletta@unitn.it)

Fondazione Bruno Kessler, Trento, Trento

Alessandro Gozzi (alessandro.gozzi@iit.it)

Functional Neuroimaging Lab
Istituto Italiano di Tecnologia
Center for Neuroscience and Cognitive Systems
Rovereto, Italy

Richard Betzel (rbetzel@indiana.edu)

Indiana University
Program in Cognitive Science & Department of Informatics
Bloomington, Indiana, 47401

Abstract

Recent theoretical and experimental work in neuroscience has focused on the representational and dynamical character of neural manifolds [e.g. (Ebitz & Hayden, 2021; Saxena & Cunningham, 2019; Mante, Sussillo, Shenoy, & Newsome, 2013; Cunningham & Yu, 2014)]. These neural manifolds are subspaces in neural activity space wherein many neurons coactivate. Importantly, neural populations studied under this “neural manifold hypothesis” are not cleanly divided into separate neural populations. Instead, many neurons contribute to most manifolds in some way or another. Here, we leveraged RNNs as a model system to study the character of *discrete* neural populations. We used a community detection method from network science to produce a partition that separates neurons into distinct populations. These partitions allowed us to ask the following question: *do these discrete boundaries between neural populations matter to the system?* We found evidence that these boundaries do matter to the system. First, we found that these boundaries neatly divide the representational content and role of neurons. Next, we found that these boundaries can be directly inferred from features of the weight matrix and we corroborated this result with structural and functional imaging data from mice and humans. Finally, we found that the dynamics of these RNNs respected the boundaries of neurons into distinct populations.

Keywords: Neural manifolds; Neural populations; recurrent neural networks; representation; dynamical systems

Introduction

Here, we explore the importance of boundaries between discrete neural populations. To do this, we used recurrent neural networks (RNNs) trained on systems neuroscience tasks as a model system to ask the question: *Are boundaries between neural populations important to the system?* (György Buzsáki, 2019). Our investigations broadly fall into three categories: (1) representation and selectivity of neural populations, (2) origin of neural population, and (3) dynamics of neural populations.

We demonstrate many of our main results on an RNN trained to perform a perceptual decision-making task (Molano-Mazon et al., 2022). This task presents the RNN with two stimuli drawn from normal distributions with different means. The task of the RNN is to identify the distribution with the greater mean. This requires the RNN to track previous values of both stimuli and to compare them. The task-relevant variable for this task is the cumulative difference between the means of the two input stimuli distributions (hereafter referred to as *cumulative Δ inputs*). After a fixation period, the RNN makes its decision based on the relative activity of two output neurons. The output neuron with the greatest activity corresponds to the RNNs decision (see Fig. 1a for a schematic of this task).

Representation and Selectivity

In this section, we ask if boundaries between discrete neural populations (Fig. 1b) also neatly divide the selectivity of neurons that represent different task-relevant information. We find that these boundaries not only divide neurons based on their representational profiles (Fig. 1c-e), but lesioning output connections on one side of this boundary only effects task trials requiring the representations of that population (and vice versa; Fig. 1f-h).

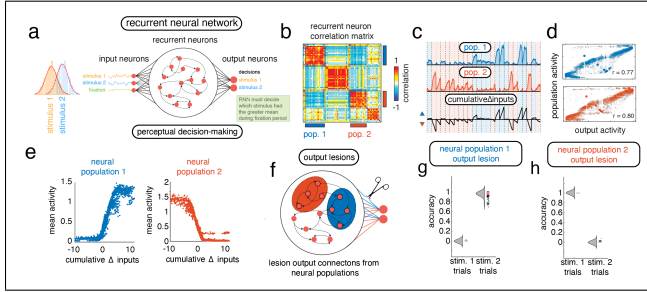


Figure 1: Representations are unique to discrete neural populations.

Origin

In this section, we ask if we can find features of the neuronal connection weights that *create* boundaries between neural populations. We find that neural population boundaries are determined by the sign (+/-) of incoming connection weights. Incoming connection weights can be thought of as living in an N dimensional space, where N is the number of input neurons. For a neuron with two input neurons, its neural population will be determined by the quadrant of the input connection weights in this space (Fig. 2a-d). For a neuron with three input neurons, its neural population will be determined by the octant of the input connections in this space. The number of sectors (quadrants/octants) scales as 2^N . As N increases, we find that the boundary between nearby sectors collapses and the neurons in combined sectors form a larger neural population. We continue to investigate the exact reason for this kind of boundary collapse so that we can predict the nearby boundaries that will collapse.

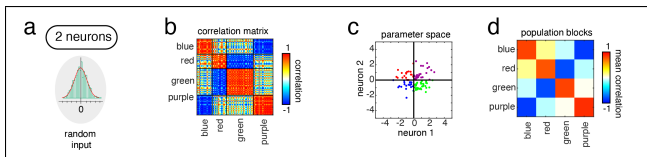


Figure 2: Origins of neural populations when $N = 2$ input neurons.

Dynamics

In this section, we ask if the dynamics of RNNs respect the boundaries between discrete neural populations. In multiple analyses, we find evidence that neural dynamics *do* respect neural population boundaries. These analyses can be subdivided into three subsections: (1) neural dynamics for computation, (2) neural dynamics following neuronal lesions, and (3) neural dynamics defined by higher-order nullclines.

Neural dynamics for computation

Input neuron connection weights define neural population boundaries that are projected onto recurrent neurons. Here, we use those boundaries to determine a lesioning protocol

(Fig. 3e) for recurrent neuron connections. More specifically, we sought to answer the following question: *Do lesions on either side of these boundaries have specific/unique effects on dynamics for computation?*

To answer this question, we created a perturbation paradigm for RNNs trained to perform the perceptual decision-making task, wherein we presented one or the other input neuron with a high-value of input stimulus (Fig. 3a). Trained RNNs respond by maintaining activity far away from a *decision boundary*, effectively performing a dynamic computation for the current value of cumulative Δ inputs (Fig. 3b-c).

After lesioning connections on one side of the boundary, we found that computations for maintaining stimuli 1 information were destroyed, but not computations for maintaining stimuli 2 (and vice versa; Fig. 3e-h).

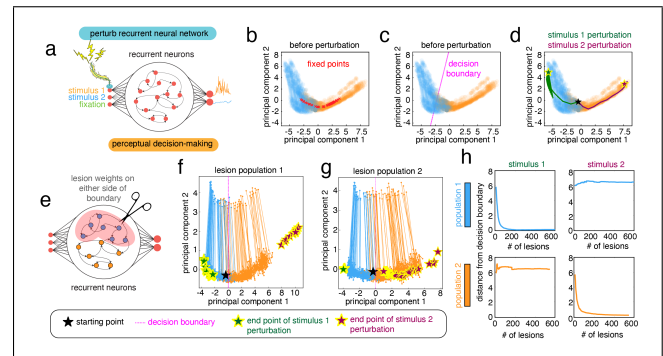


Figure 3: Lesions destroy specific neural computations.

Neural dynamics following neuronal lesions

In this section we investigate the effects of lesions to recurrent neurons on the vector field of the RNN. Vector fields represent the global dynamics of a system by sampling from the directional forces or velocities at each point in space, thus offering a comprehensive view of how the system evolves over time (Fig. 4a). We measured the effects of lesions to these vector fields (flow distance; Fig. 4b,d) and compared these effects across all neurons. Overall, we found that lesions to neurons within the same population boundary have very similar effects on the dynamics of the vector field, whereas neurons in two separate populations have different effects on the vector field (Fig. 4c,d).

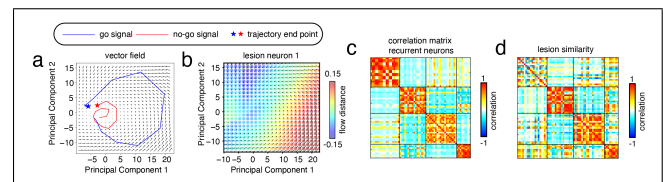


Figure 4: Lesions within a population disrupt dynamics in a similar way.

Neural dynamics defined by higher-order nullclines

In this section, we approximate higher-order nullclines for large RNNs and compare them with neural population structure. Nullclines are lines/surfaces where the derivative of a single state variable is equal to zero, and where all the nullclines intersect, they form fixed points (Fig. 5a). We approximated higher-order nullclines (where all nullclines intersect except for one; Fig. 5b), and we found that the similarity of the higher-order nullclines was reasonably circumscribed by the boundaries between neural populations (Fig. 5d,e).

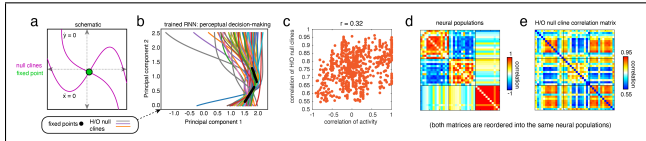


Figure 5: Higher-order nullclines are similar within neural populations.

Conclusion

We conclude that neural population boundaries are important to the system in a variety of ways. We therefore expect that by using methods that approximate these boundaries alongside neural manifold methods that are usually used (like principal component analysis), we will gain additional insights into the dynamics of computation and representation in brains.

References

- Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*, *17*(11), 1500–1509.
- Ebitz, R. B., & Hayden, B. Y. (2021). The population doctrine in cognitive neuroscience. *Neuron*, *109*(19), 3055–3068.
- György Buzsáki, M. (2019). *The brain from inside out*. Oxford University Press.
- Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, *503*(7474), 78–84.
- Molano-Mazon, M., Barbosa, J., Pastor-Ciurana, J., Fradera, M., Zhang, R.-Y., Forest, J., ... others (2022). Neurogym: An open resource for developing and sharing neuroscience tasks.
- Saxena, S., & Cunningham, J. P. (2019). Towards the neural population doctrine. *Current opinion in neurobiology*, *55*, 103–111.