Causality and Complexity in Monkey Prefrontal Cortex

Presenting Author : **Luke Chen**

Introduction

Neuroscientists have long been investigating brain function with correlations, i.e. Pearson Correlation, Canonical Correlation Analysis, Representational Similarity Analysis. However, there are concerns and supporting evidence showing correlation does not imply causation. Quoting Judea Pearl: "Statisticians have been immensely confused about what variables should and should not be controlled for, so the default practice has been to control for everything one can measure. It is a convenient, simple procedure to follow, but it is both wasteful and ridden with errors…at the same time, statisticians greatly under rate controlling in the sense that they are loath to talk about causality at all"

Specifically, people have been applying Granger Causality (and transfer entropy) as a tool to claim causality between neural spike trains. However, both these methods have significant flaws. Here we used cutting-edge mathematical tools such as differential geometry and dynamical systems analysis to apply state-of-the-art causality testing, in order to find causation relationships between time-series dynamics in neural spike trains in the PFC.

Methods

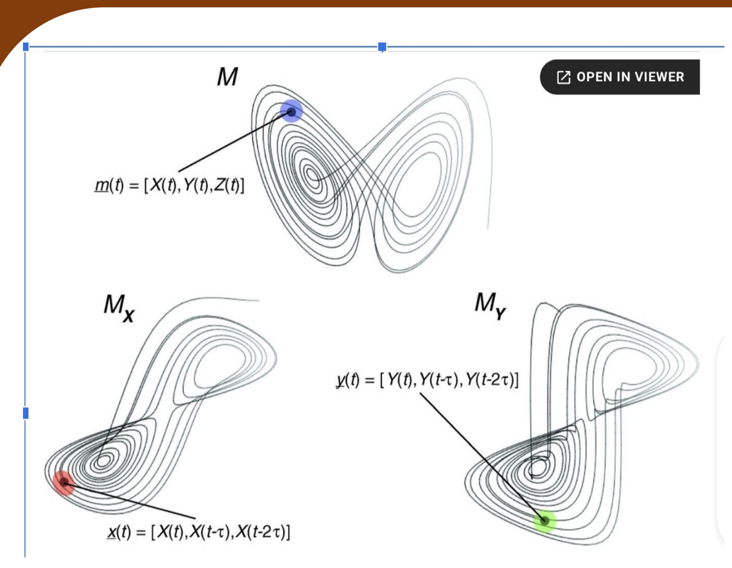

Figure 2: *Convergent cross mapping (CCM)* uses shadow manifold correspondence. This simple Lorenz attractor example shows the attractor manifold of the original system M(X,Y,Z) along with the two shadow manifolds created with time lag = tau. Since X and Y are dynamically coupled, points that are proximal in Mx will be proximal in My (by Takens' Embedding Theorem), allowing us to estimate states in X from Y, (using nearest neighbor embedding), and vice versa. Longer time series cause Mx and My to become denser and neighborhoods in red and green to shrink, giving us more precise cross-map estimates.

Small neural net

Here, we generate synthetic data following the dynamics of a small random neural network, described by the equation

$$x(t+1) = \tanh(aJx(t))$$

where $x(t) \in \mathbb{R}^N$ and J is an $N \times N$ random matrix that is organized into blocks:

$$J = \begin{bmatrix} J^{AA} & J^{AB} \\ J^{BA} & J^{BB} \end{bmatrix},$$

where entries of J^{AA} and J^{BB} are i.i.d. generated as ξz where $\xi \sim \mathcal{N}(0, \sigma)$ and $z \sim Bernoulli(p)$. Similarly, for J^{BA} I use $\xi \sim \mathcal{N}(0, \sigma_M)$ and $z \sim Bernoulli(p_M)$. All entries of J^{AB} are set to 0.

The intuition here is that we have two populations, A and B, with strong intraconnectivity--if a, σ, and N are large enough, dynamics of each of these populations alone is chaotic. Additionally, I add relatively weak random connections from A to B. Thus, A drives B but not the other way around, and the goal is to test whether the CCM can recover this fact.

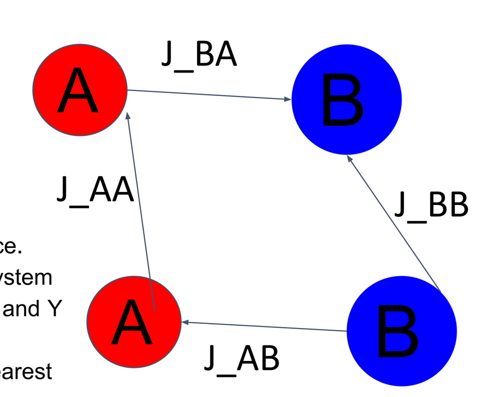

Figure 4: Schematic of Small Neural Net, Populations A and B, Transition Probabilities

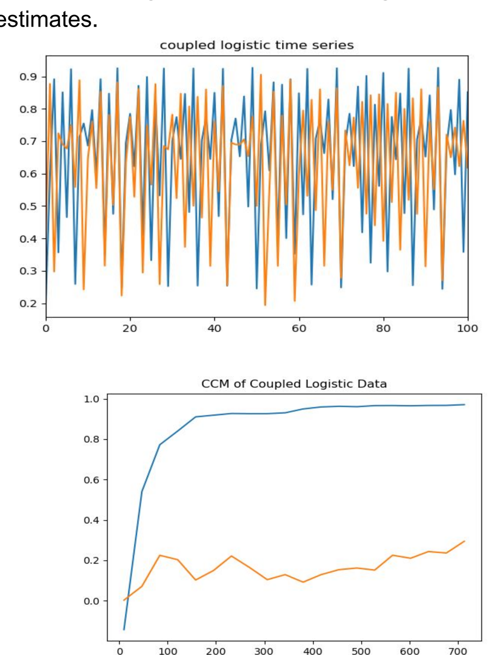

Figure 3
3A: Coupled Logistic Time Series Example with Clear Causality
3B: CCM of Coupled Logistic Time Series Shows a Clear Unidirectional Causality and Strong Convergence as Library Size Increases, As Expected

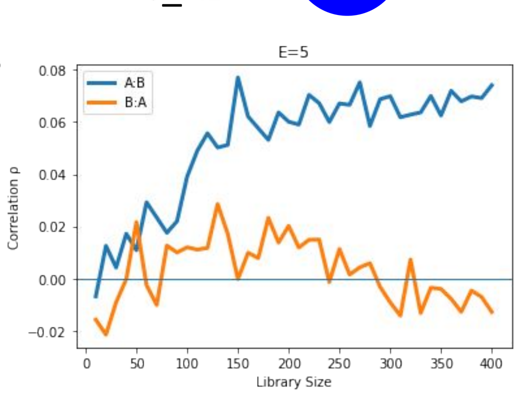

Figure 5: Weak causation from small neural net example demonstrated from A neuron subpopulation to B neuron subpopulation from toy ANN data, but no causation from the B to A direction

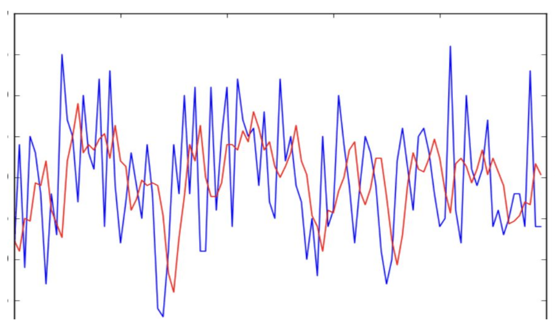

Figure 6: Moving Average Time Series Smoothing Applied to Individual Neural Spike Trains for Noise Reduction Used Before Causality Analysis

Task Structure Description

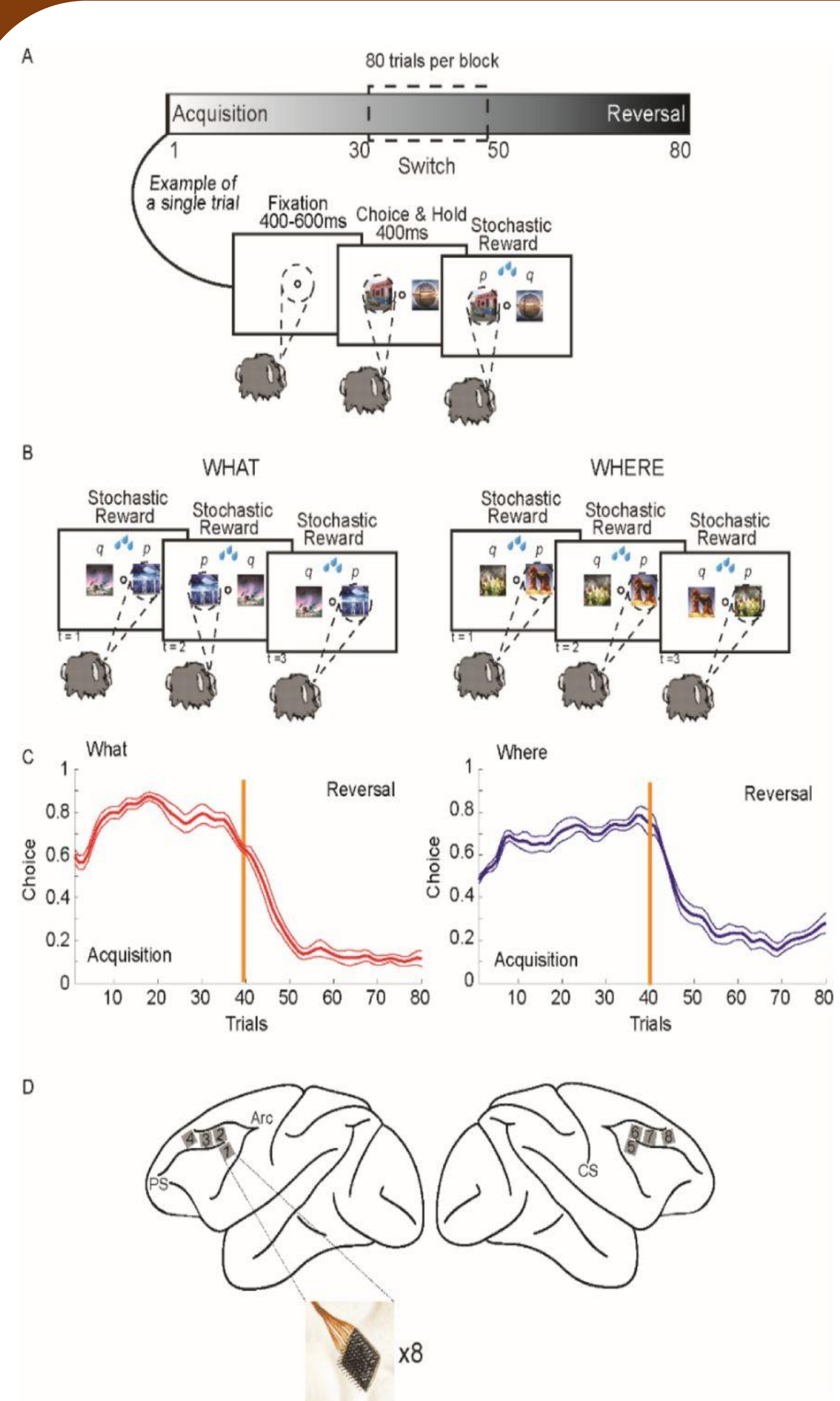

Figure 6: This was a probabilistic reversal task

- **A.** Task was carried out in 80 trial blocks. At the beginning of each block of trials, 2 new images were introduced that the animal hadn't seen before. For each trial the animals fixated, and then two images were presented. Monkeys made saccades to indicate their choice and then were stochastically rewarded
- **B.** 2 Conditions:
  - In "what" condition one of the images was more frequently rewarded (p = 0.7) independently of which side it appeared on, and one of the images was less frequently rewarded (p = 0.3)
  - In "where" condition one of the saccade directions was more frequently rewarded (p = 0.7) and one was less frequently rewarded (p = 0.3) independently of which image was at the chosen location. The condition remained fixed for the entire block. However, on a randomly chosen trial between 30-50, reward mapping was reversed and the less frequently chosen object or location became more frequently rewarded, and vice-versa.
- **C.** Choice behavior across sessions. Monkeys quickly learned the more frequently rewarded image (left panel) or direction (right panel), and reversed their preferences when the choice-outcome mapping reversed. Because the number of trials in the acquisition and reversal phase differed across blocks, the trials were interpolated in each block to make all phases of equal length before averaging. The choice data was also smoothed using Gaussian kernel regression (kernel width sd = 1 trial).
- **D.** Schematic showing locations of recording arrays, 4 in each hemisphere. Array locations were highly similar across animals.

Results

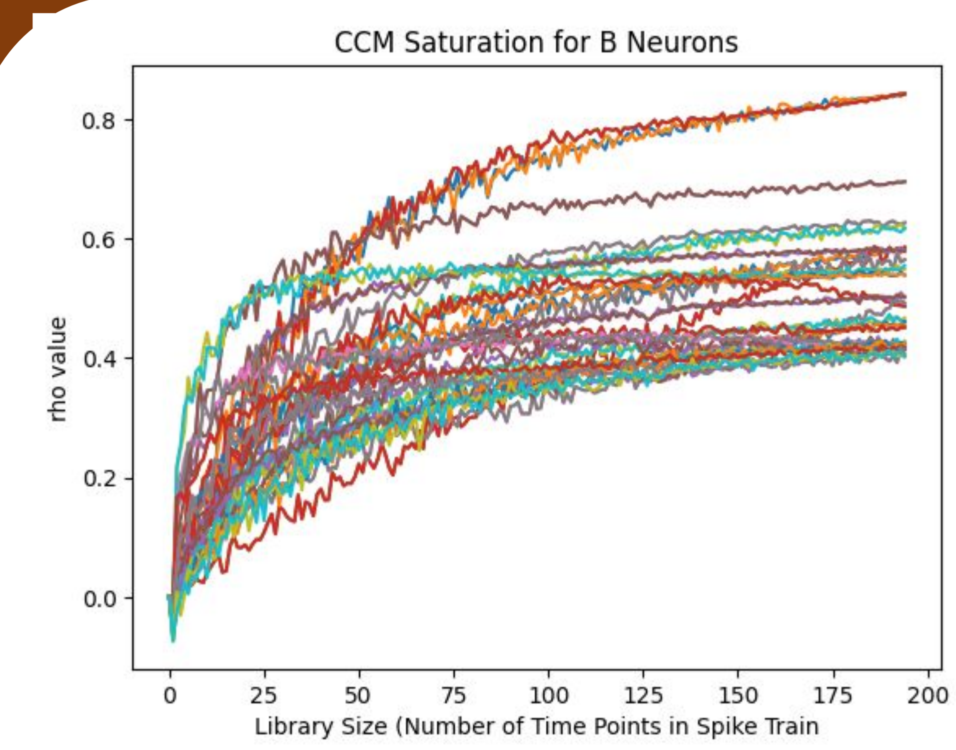

Figure 7: Unidirectional Causality Shown Through Convergence of CCM with Sample Neurons from Area B

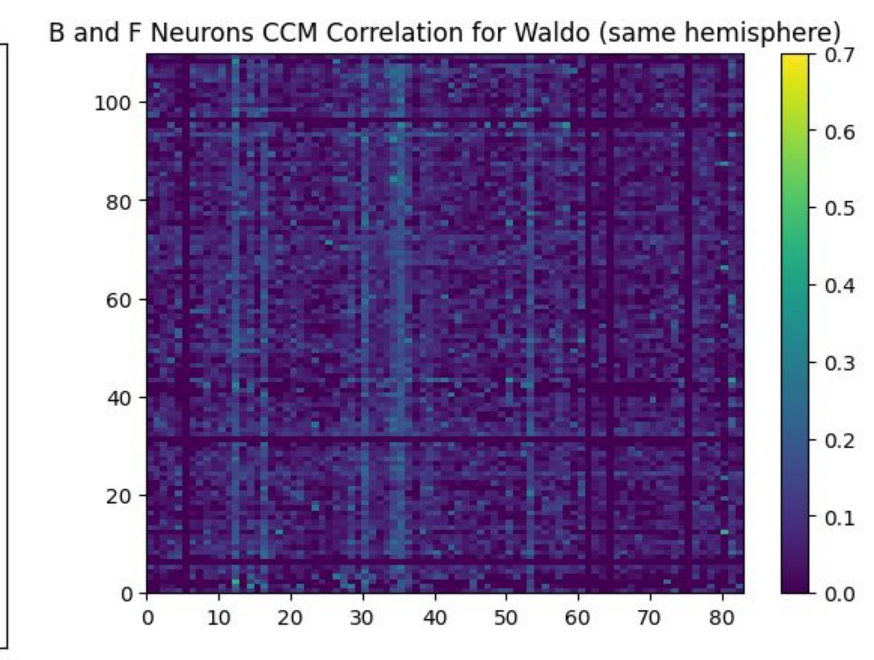

Figure 10: CCM Matrix for Corresponding Regions Across Left and Right Hemispheres of Monkey's PFC

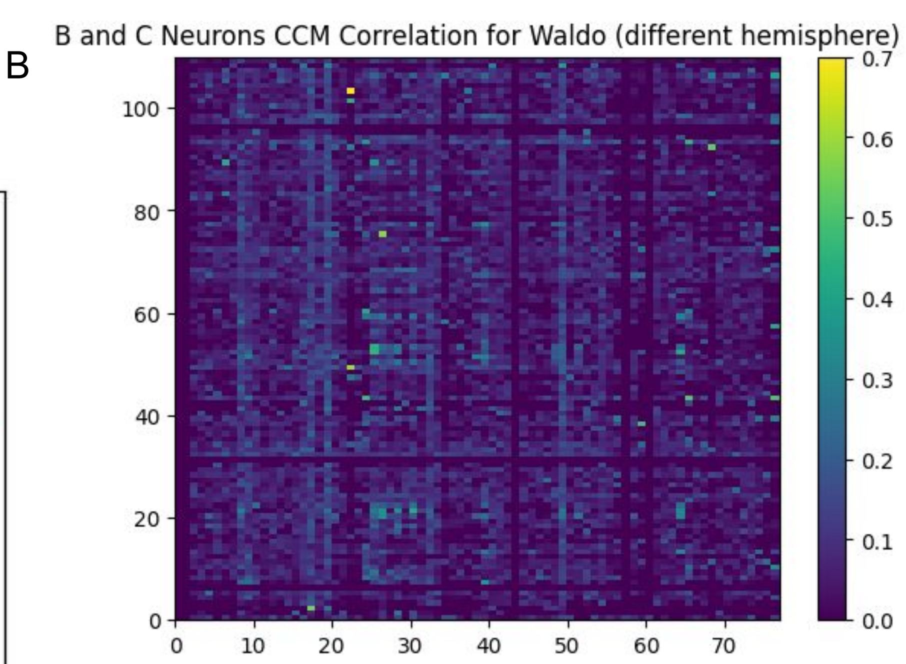

Figure 11: CCM Matrix for Adjacent Regions in Left Hemisphere of Monkey's PFC

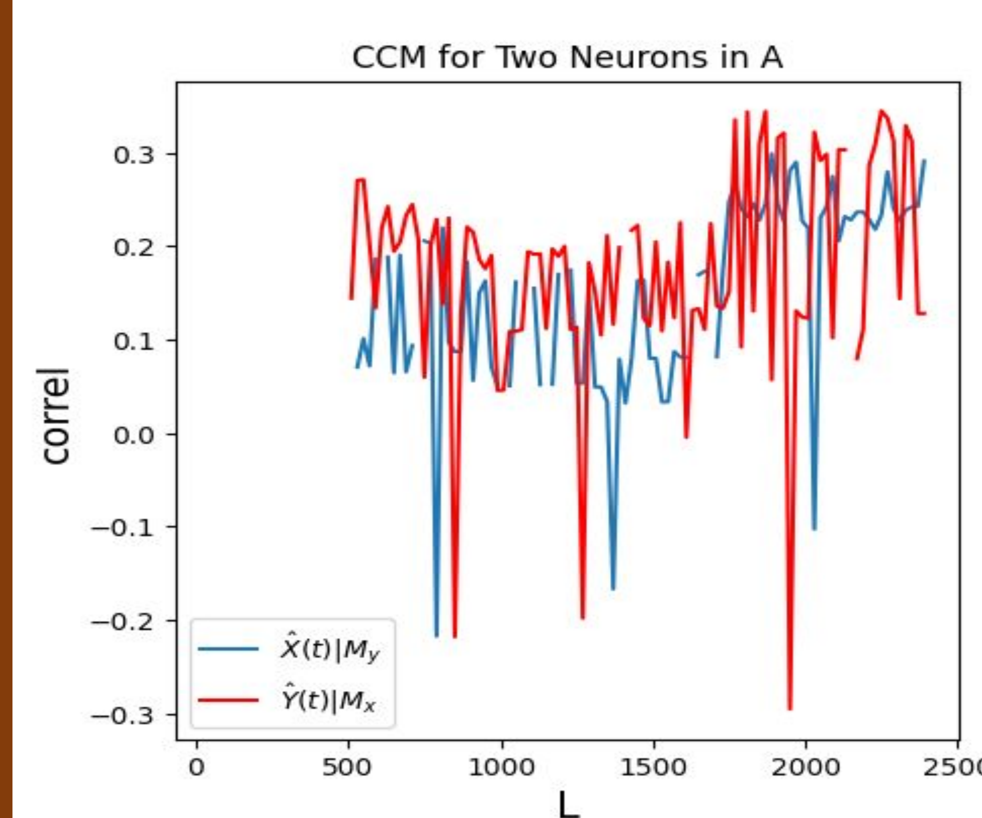

Figure 8: Bidirectional Causality Shown Through Convergence of CCM with Two Sample Neurons in Area A

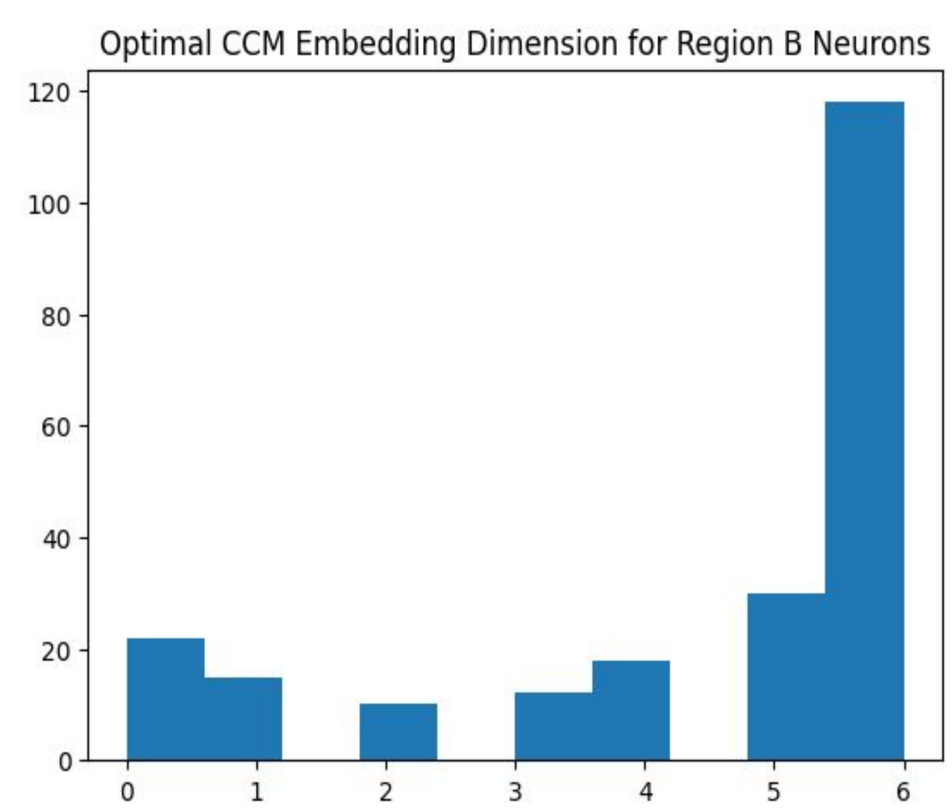

Figure 9: Optimal embedding dimension (E) that gives highest CCM causality for area B neurons shows that the complexity of the underlying dynamical system the neural spike trains seems to trend towards E=6, suggesting high complexity

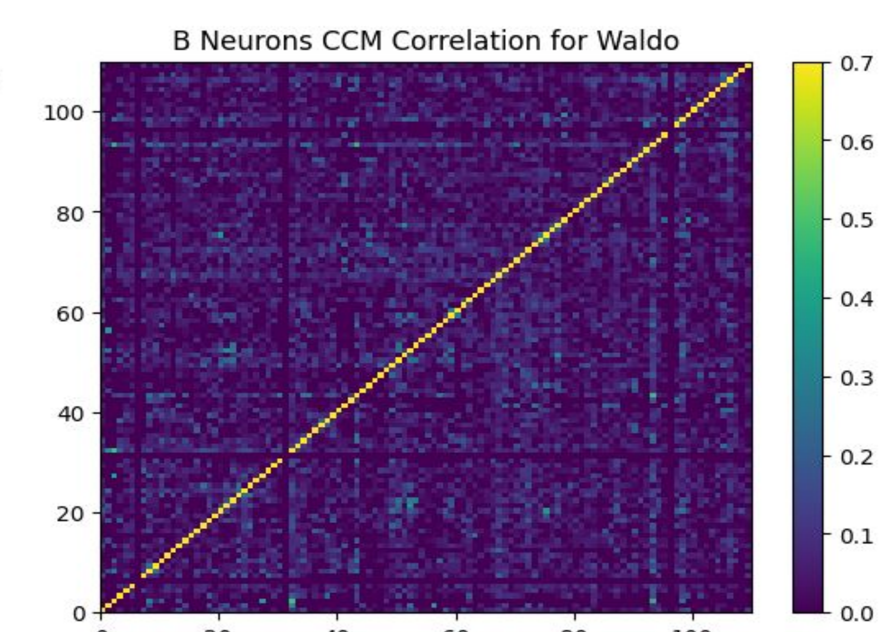

Figure 12: CCM Matrix for Region B shows smaller clusters of adjacent neurons with higher correlation than the others

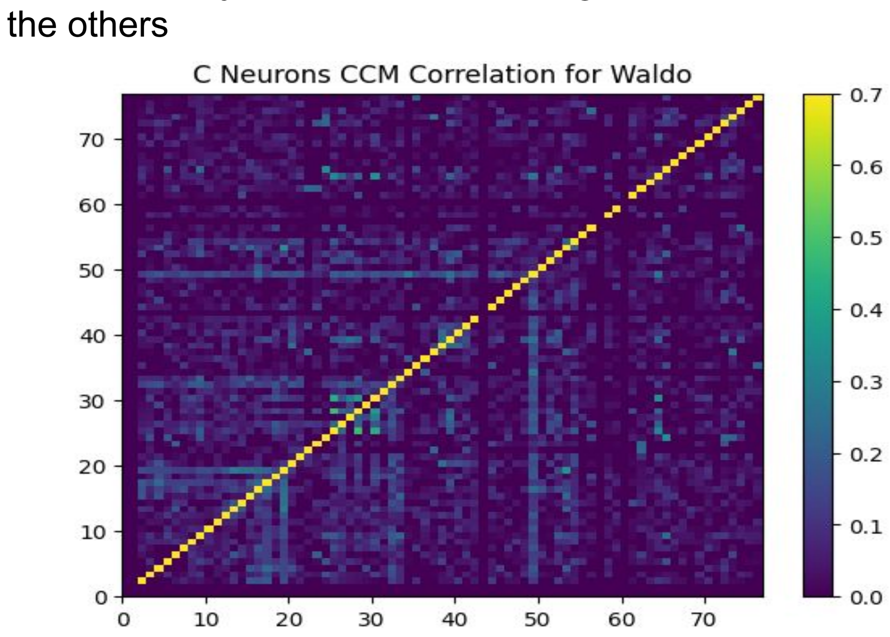

Figure 13: CCM Matrix for Region C shows clusters of adjacent neurons with higher correlation than the others

Conclusion

1. Here, we applied cutting-edge mathematical tools to detect the causality between neural time-series dynamics in a probabilistic reversal learning task.
2. We tested whether our method works properly on a well-designed simulated dataset. We found that the method correctly detects directionality of casuality.
3. Then we tested the method on real data from the vlPFC and cdlPFC/mdlPFC/rdlFPC
4. We detected directionality in specific neuron pairs through convergence of the method **(Figures 7 and 8)**
5. Then we looked at the population causality 1) between adjacent subregions and 2) same subregion across PFC hemispheres. We found some subset of neurons that show a higher causal relationship in adjacent subregions and some subpopulations exhibited strong causal relationship across hemispheres (**Figures 10 and 11)**
6. We showed that optimal embedding dimension peaked around E=6 which implies that the complexity of the underlying dynamical system appears to be most commonly 6 dimensions (**Figure 9)**

Future Directions

1. Is CCM better than traditional methods of testing causality (Granger Causality Test or Transfer Entropy)
2. Can we define any network structure within the population of neurons in one brain area? (using indirect connections, finding graph structure etc.), and use this to further interpret results of CCM?
3. Does the complexity of the underlying nonlinear neural dynamical system change (likely increase) during uncertain stages of the task versus certain stages?