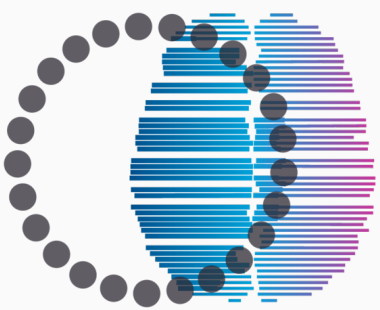


INTERNATIONAL
BRAIN
LABORATORY

Fiete Lab
2020 March Checkpoint

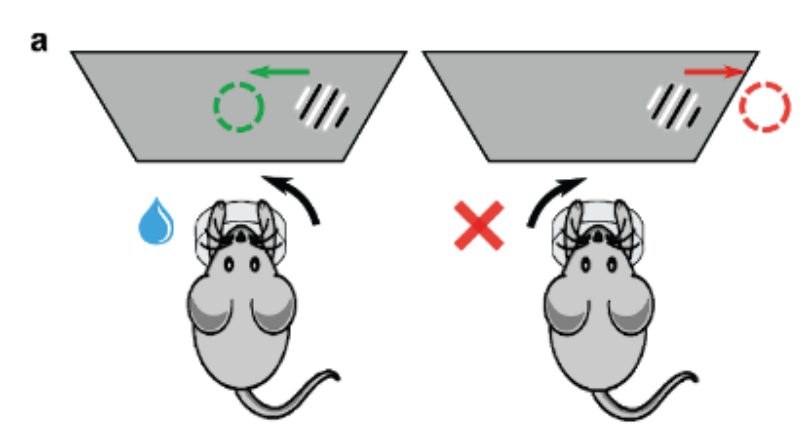
Rylan Schaeffer, Dr. Leenoy Meshulam, Professor Ila Fiete

March 24, 2020

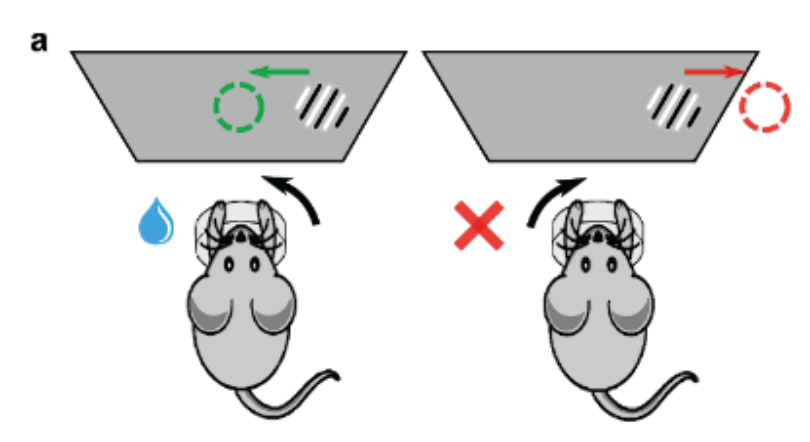
Research Goals

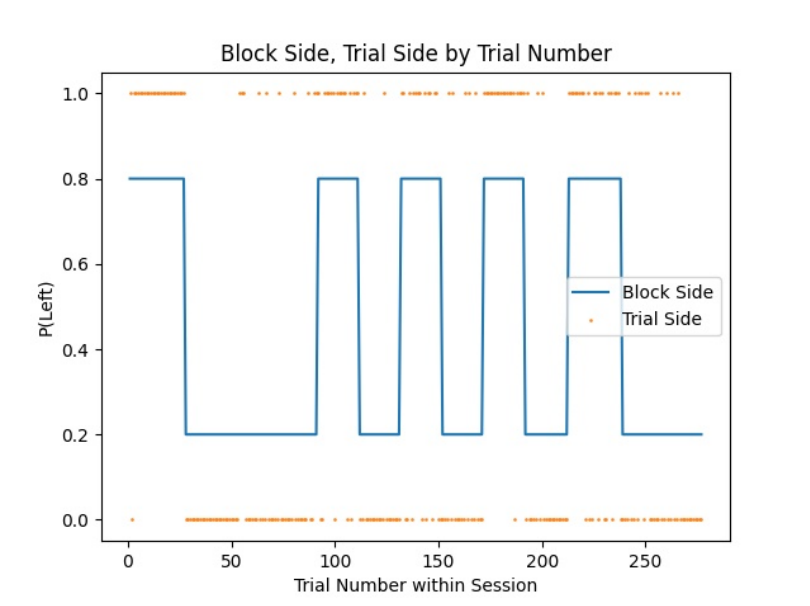
1. Reverse engineer how neurally-plausible mechanistic models solve IBL task
2. Leverage understanding to direct exploration/analysis of biological circuit data

IBL Task

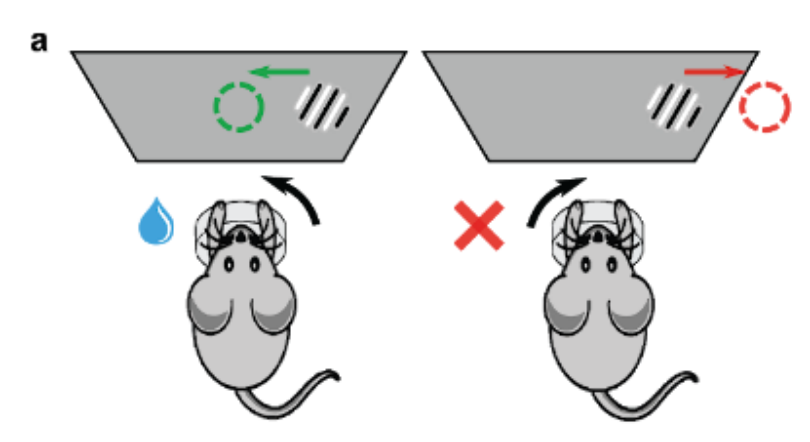


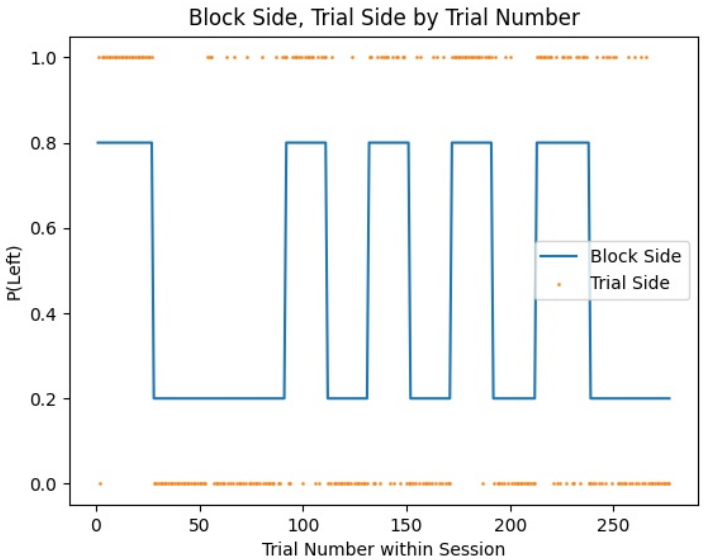
IBL Task

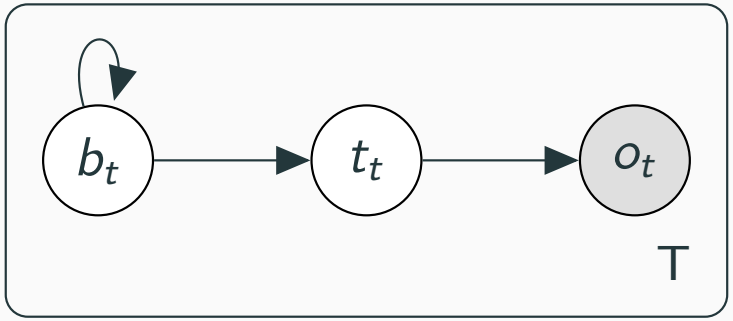

a


Block Side, Trial Side by Trial Number
1.0
0.8
0.6
0.4
0.2
0.0
P(Left)
0
50
100
150
200
250
Trial Number within Session
Block Side
Trial Side

IBL Task







Generative model of IBL task. T trials per session. On trial t: block side b_t, trial side t_t, observation o_t.

$$b_t \sim p(b|b_{t-1})$$

$$t_t \sim p(t|b_t)$$

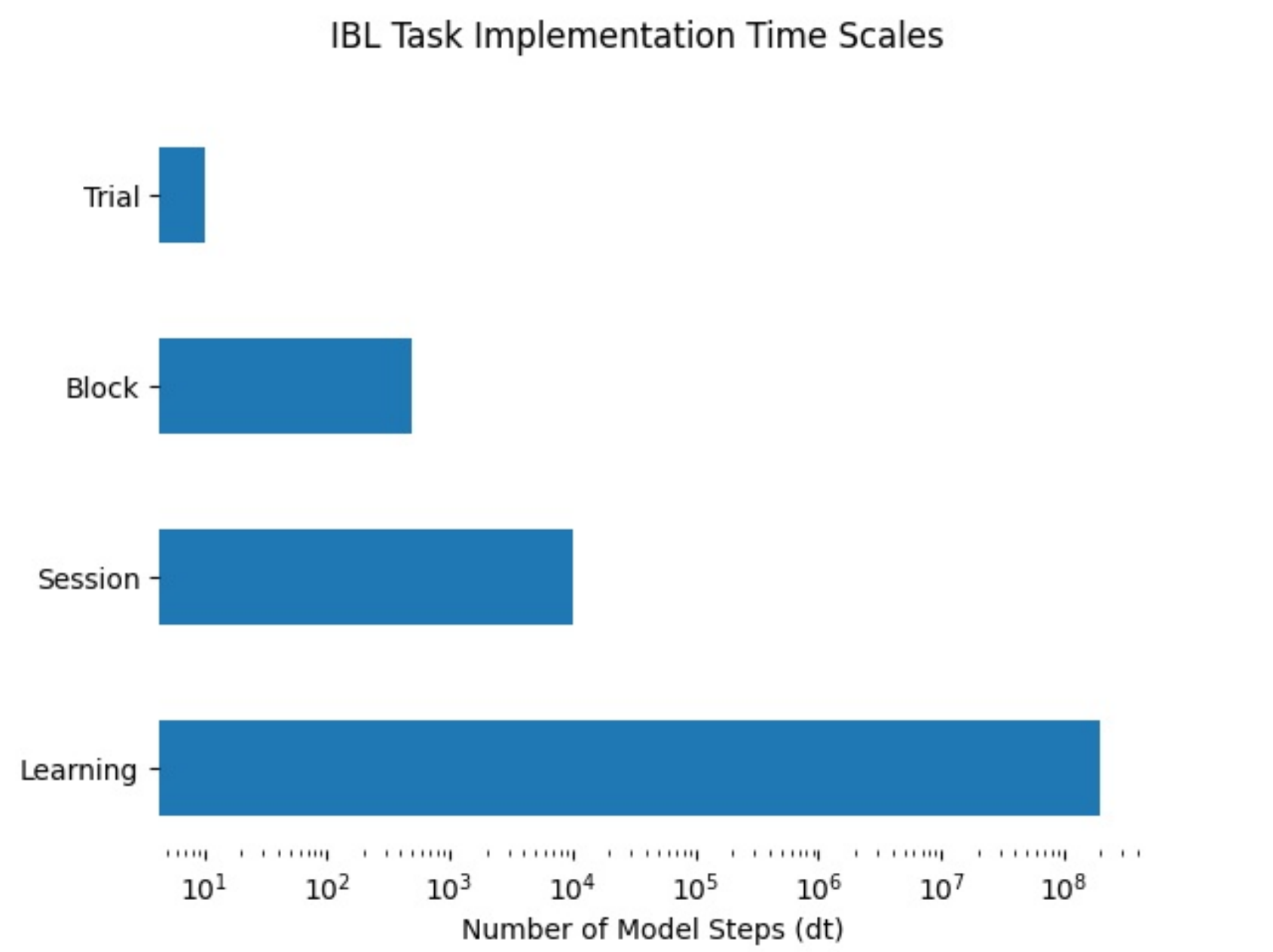
$$o_t \sim p(o|t_t)$$

Research Questions

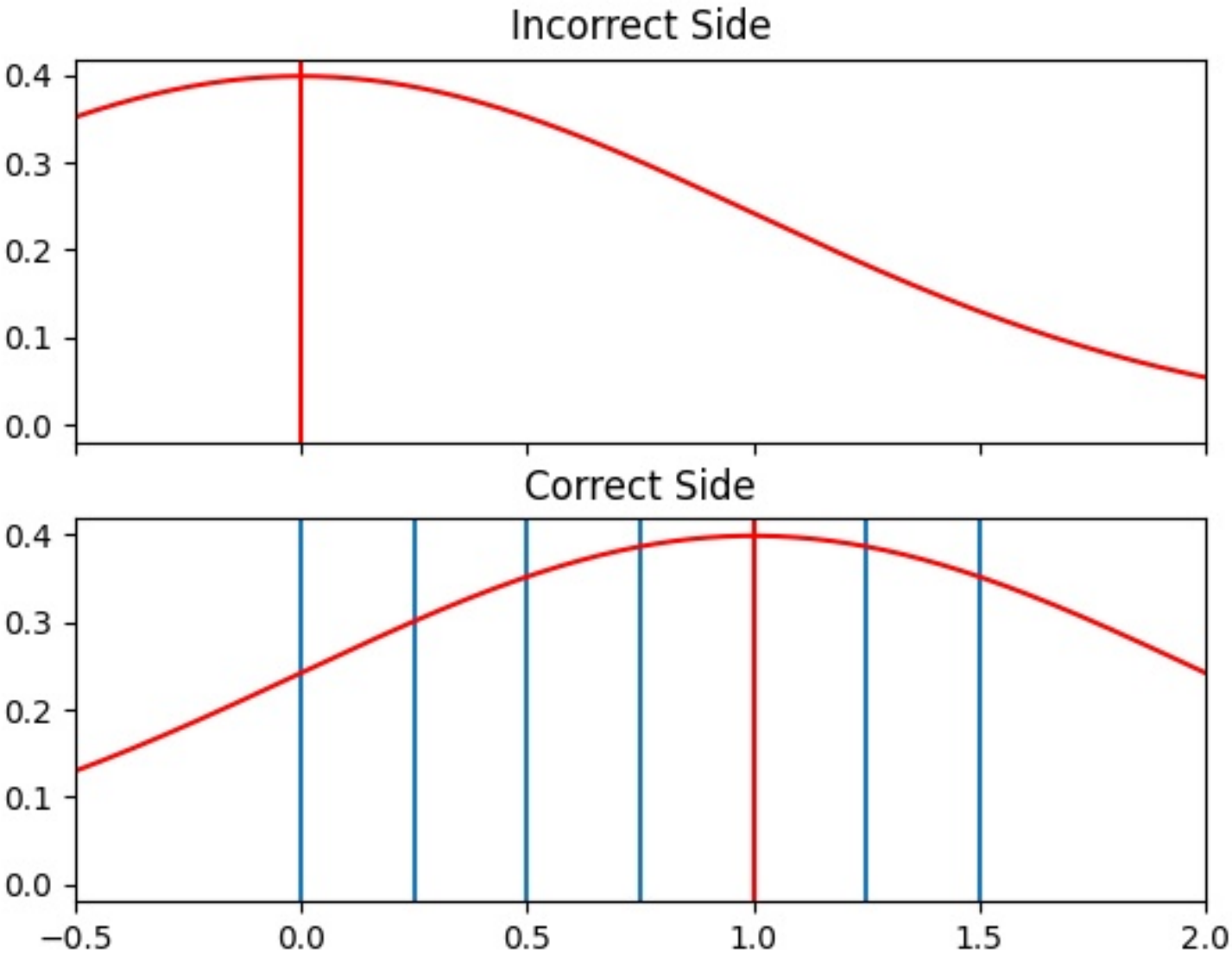
What mechanisms(s) do plausible neural models use to solve the IBL task?

1. Which of these three distributions do networks learn, and over what timescales?
2. How do networks encode current trial side and current block side?
3. How do networks use block side to influence trial side?
4. How do networks use previous stimulus, previous action and the ensuing feedback to update the block side?
5. How do networks use block duration statistics to update block side?
6. How consistent is the learnt mechanism across network design choices?

Implementation Details - Task Timescales



Implementation Details - Within-Trial Stimuli



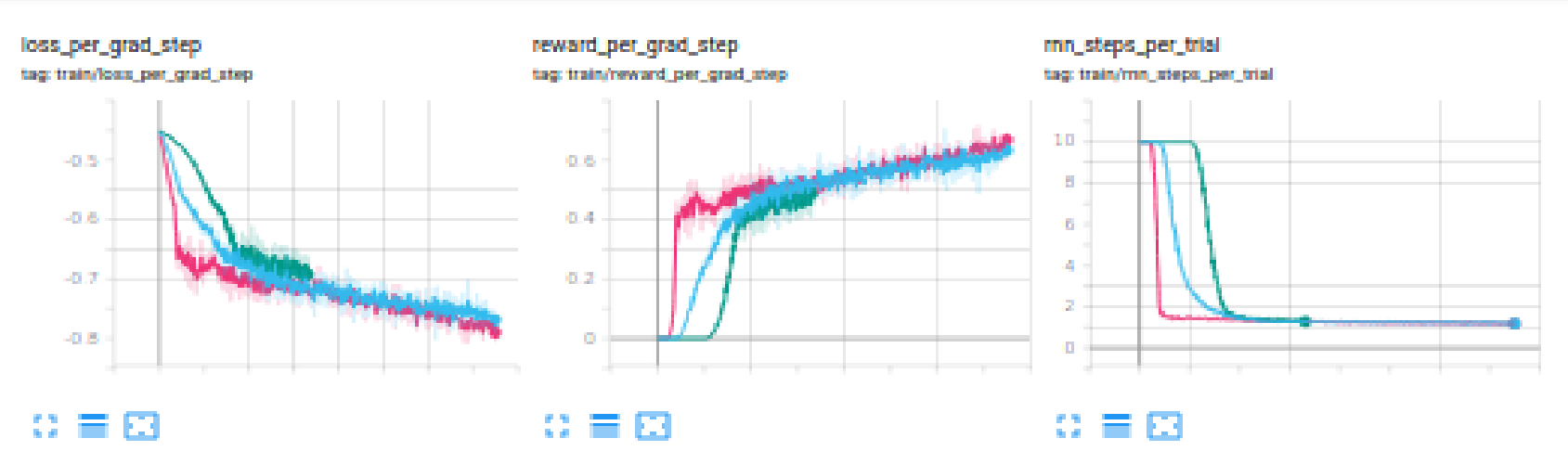
Implementation Details - Model, Loss, Etc.

- Supervised Classification (Binary Cross Entropy)
- Networks: RNN (tanh), LSTM, GRU
- Number of stacked layers: 1
- Hidden dimension: 10, 50, 100, 200
- Optimizer: SGD with LR=0.01, no momentum
- Optional weight initialization and connectivity constraints

Code online at `https://github.com/int-brain-lab/ann-rnns`

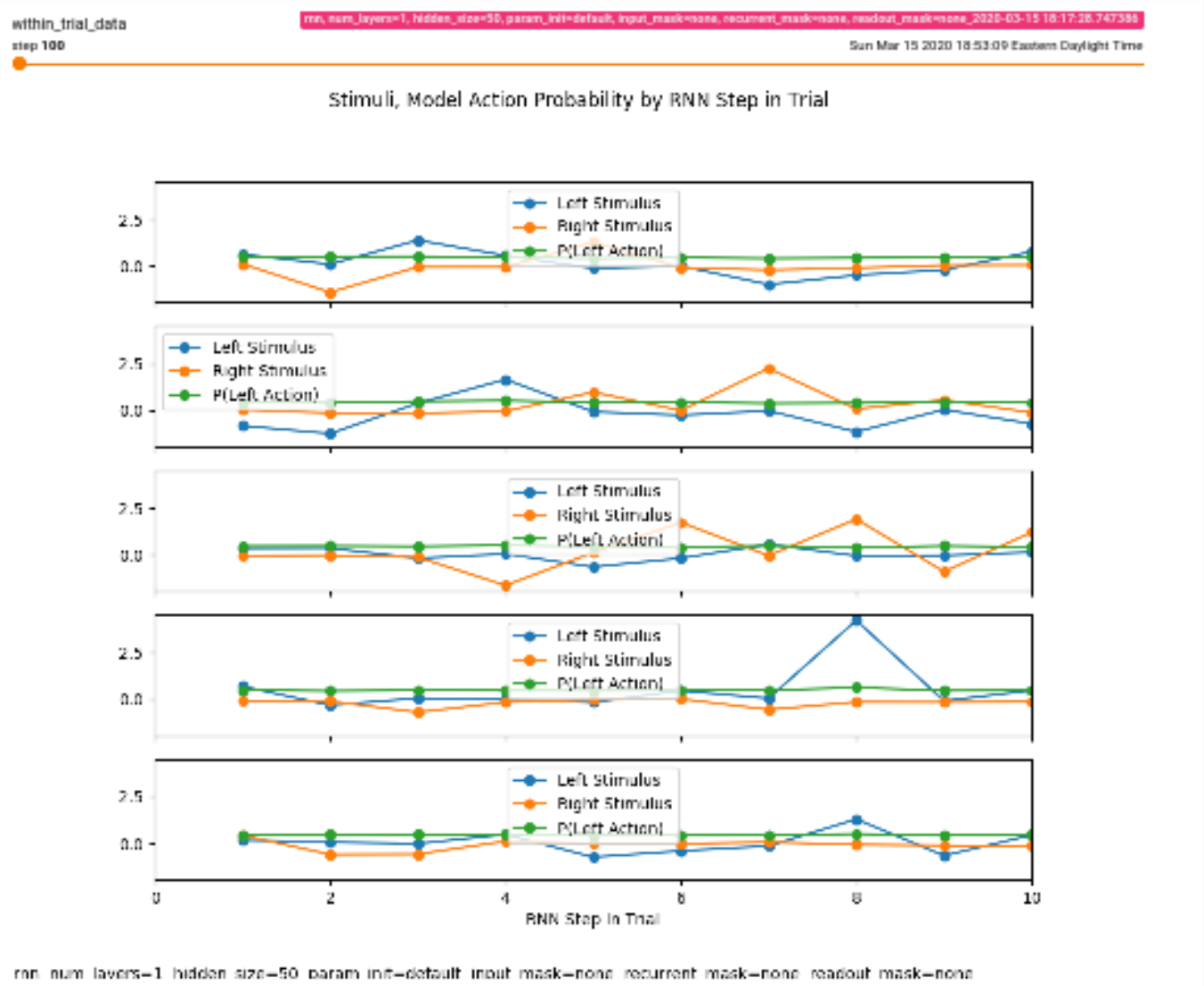
Preliminary Results - Performance

Figure 1: Loss, Avg Reward/Trial, Avg RNN Steps/Trial vs Grad Step



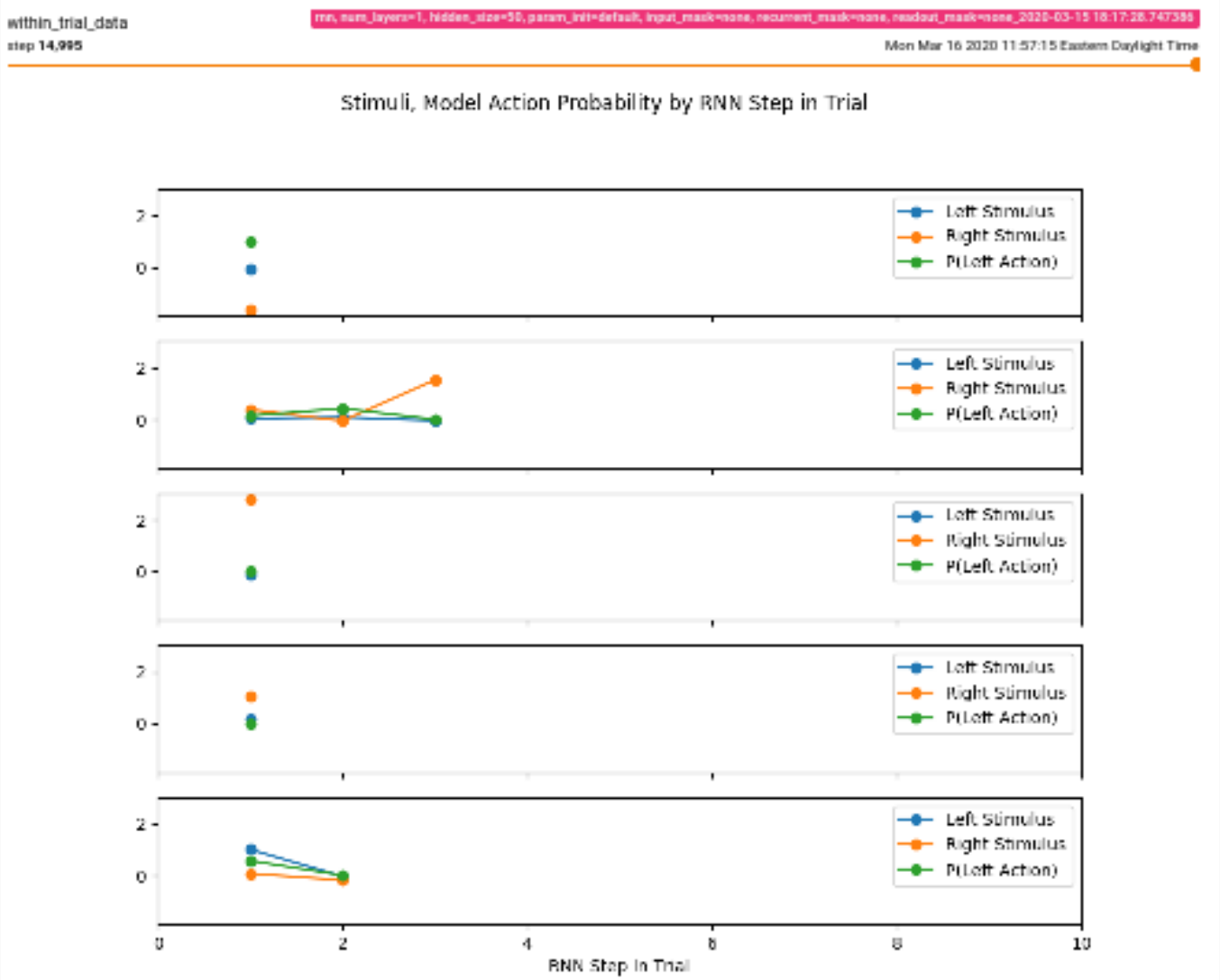
Preliminary Results - Within-Trial Behavior

Figure 2: P(Left), Left Stimulus, Right Stimulus vs *dt* within trial



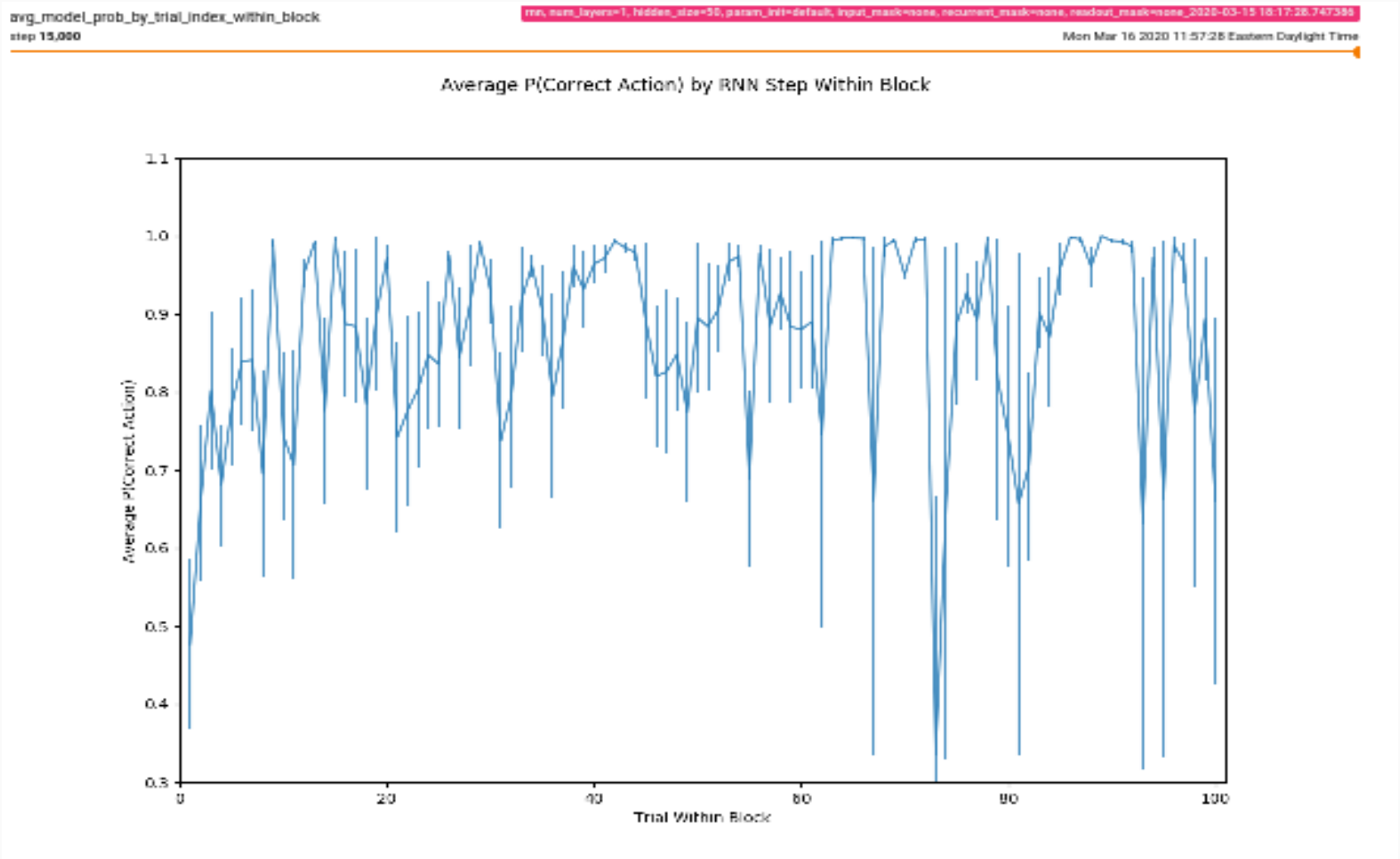
Preliminary Results - Within-Trial Behavior

Figure 3: P(Left), Left Stimulus, Right Stimulus vs *dt* within trial



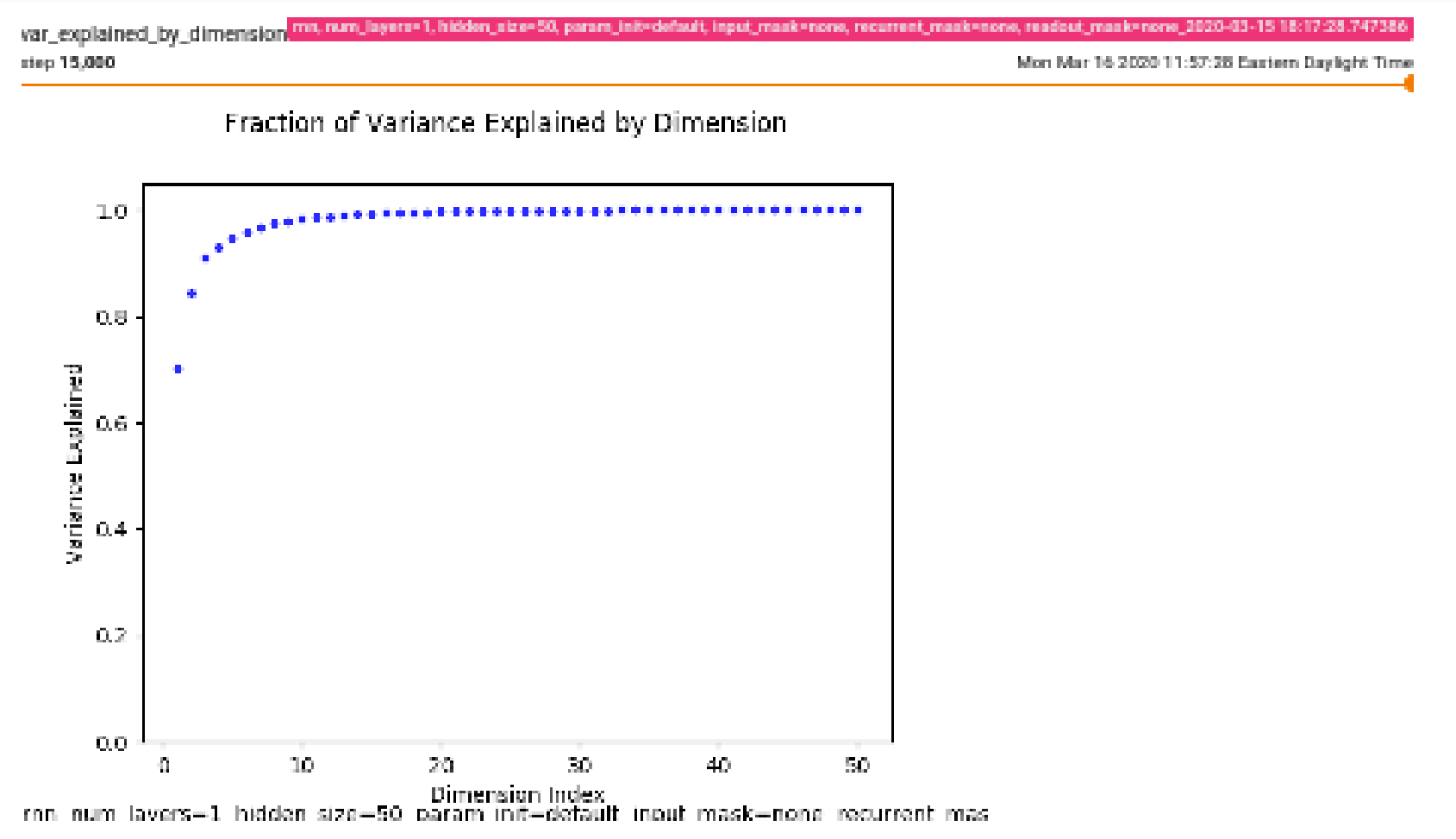
Preliminary Results - Within-Block Behavior

Figure 4: Avg P(Correct Choice) vs Trial within Block (Incorrect Block Duration i.e. mean 60, truncated to [60, 100])



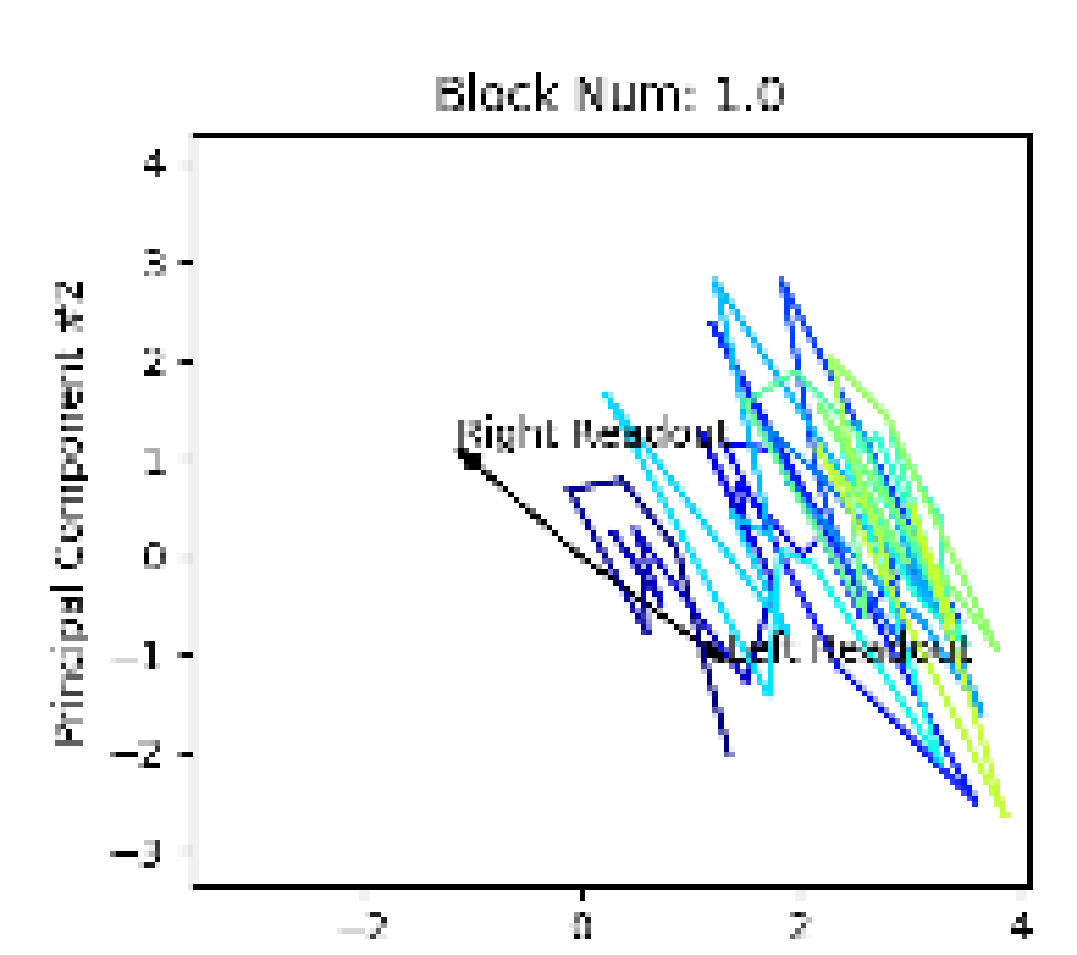
Preliminary Results - Fraction of Variance Explained

Figure 5: Cumulative Fraction of Variance in RNN state vs Principal Component Number



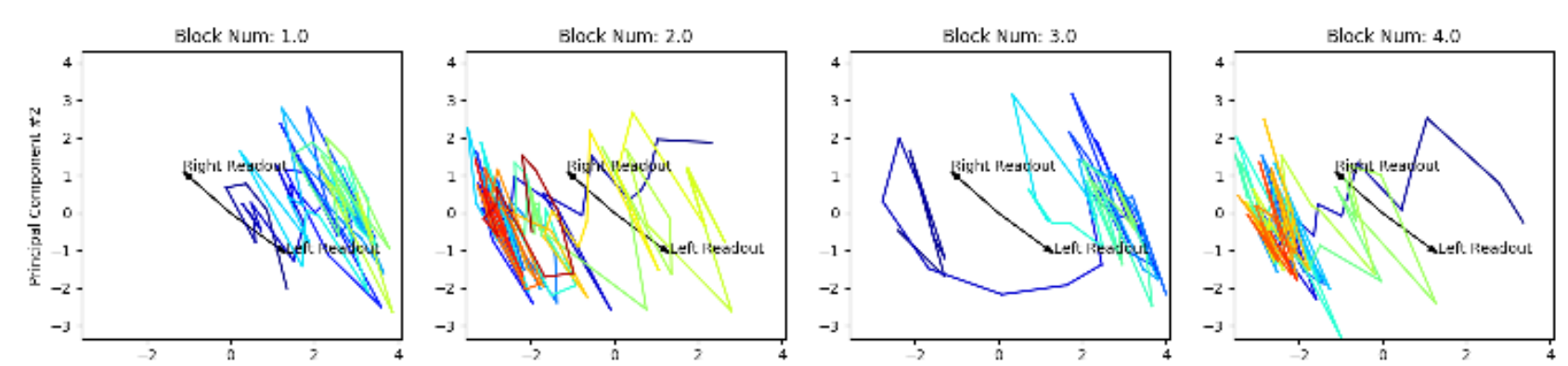
Preliminary Results - PCA-projected RNN State Within-Block

Figure 6: PCA-projected RNN state within block 1 (post-training)



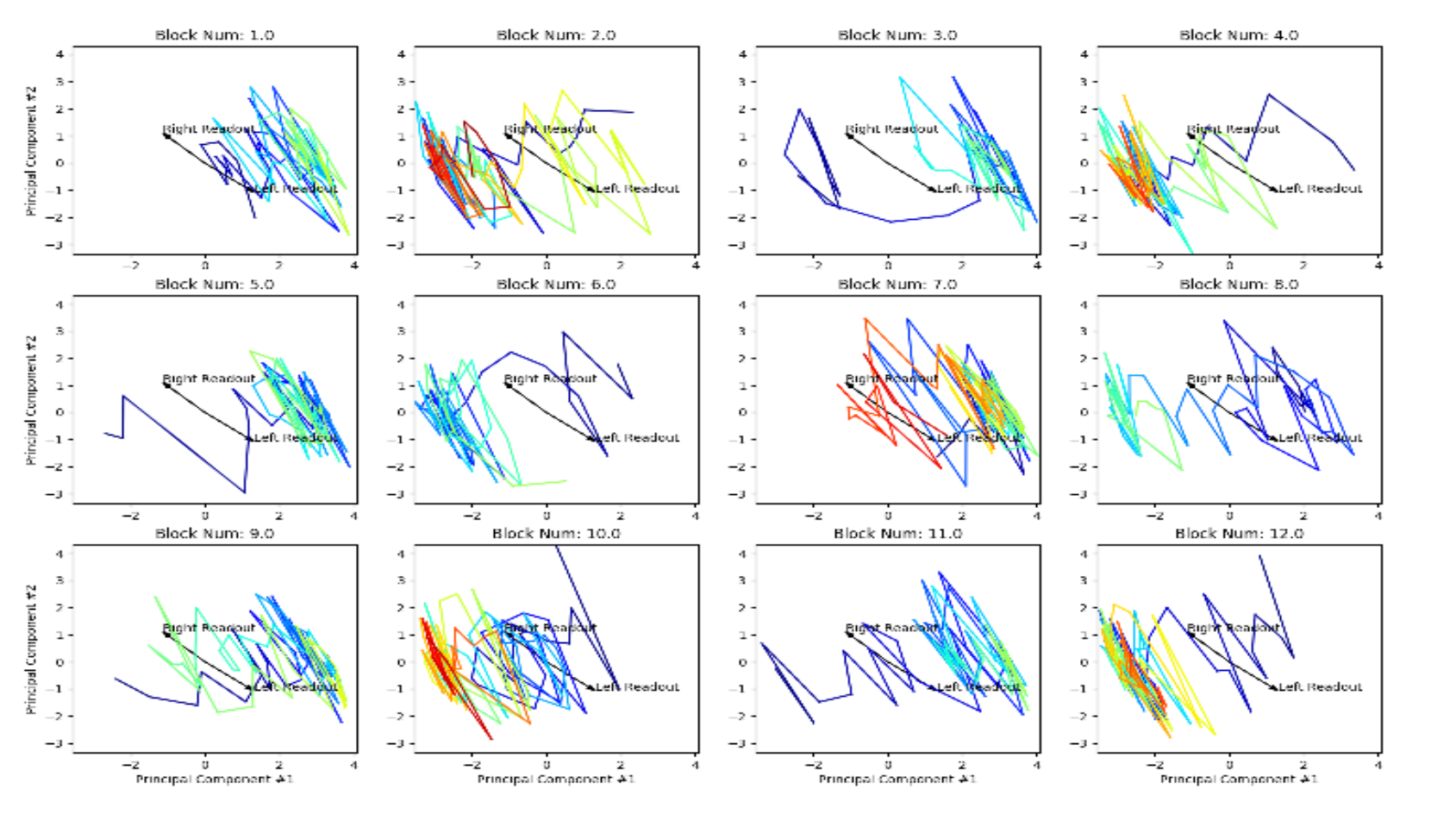
Preliminary Results - PCA-projected RNN State Within-Block

Figure 7: PCA-projected RNN state within blocks 1-4 (post-training)



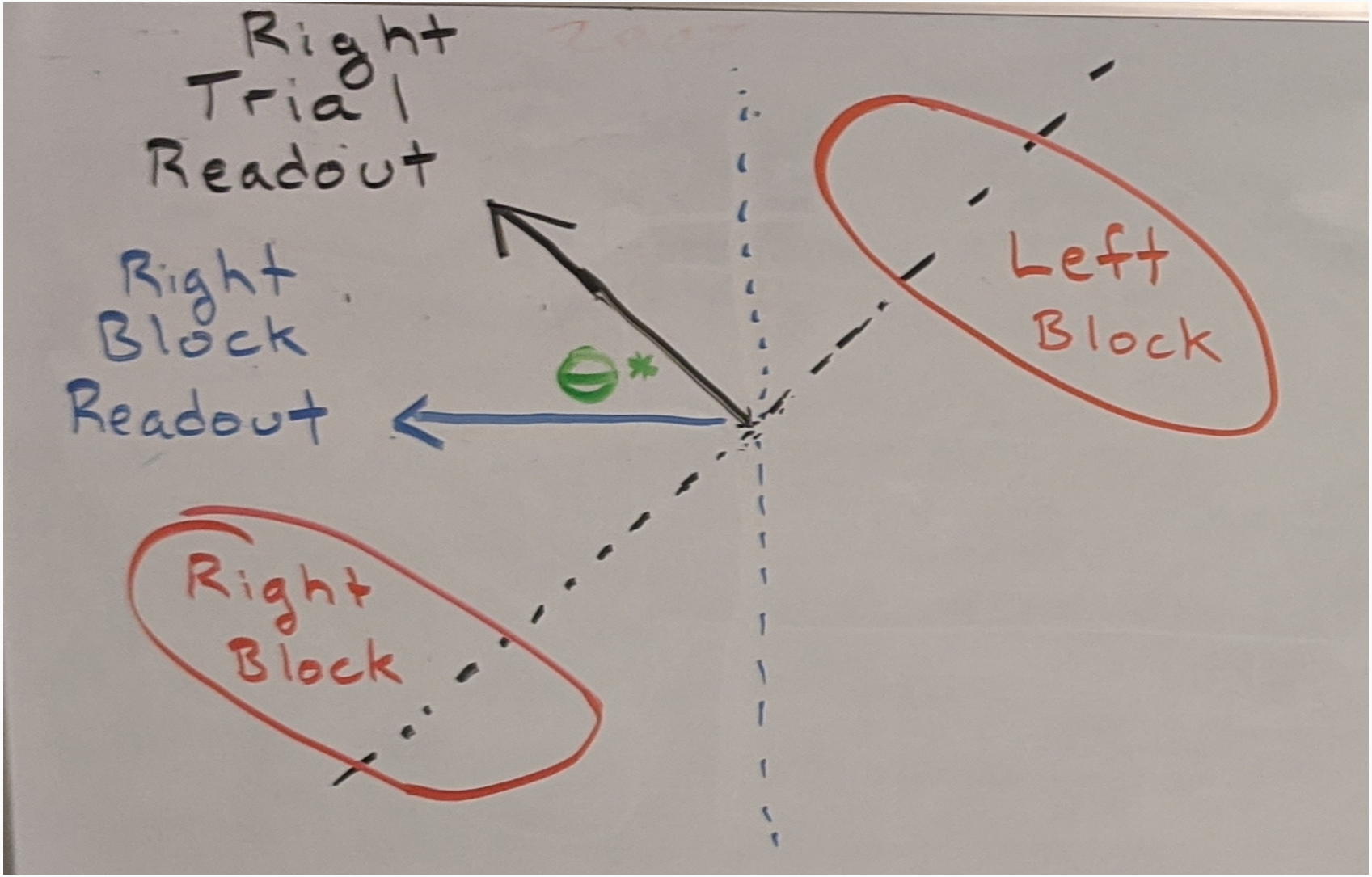
Preliminary Results - PCA-projected RNN State Within-Block

Figure 8: PCA-projected RNN state within blocks 1-12 (post-training)

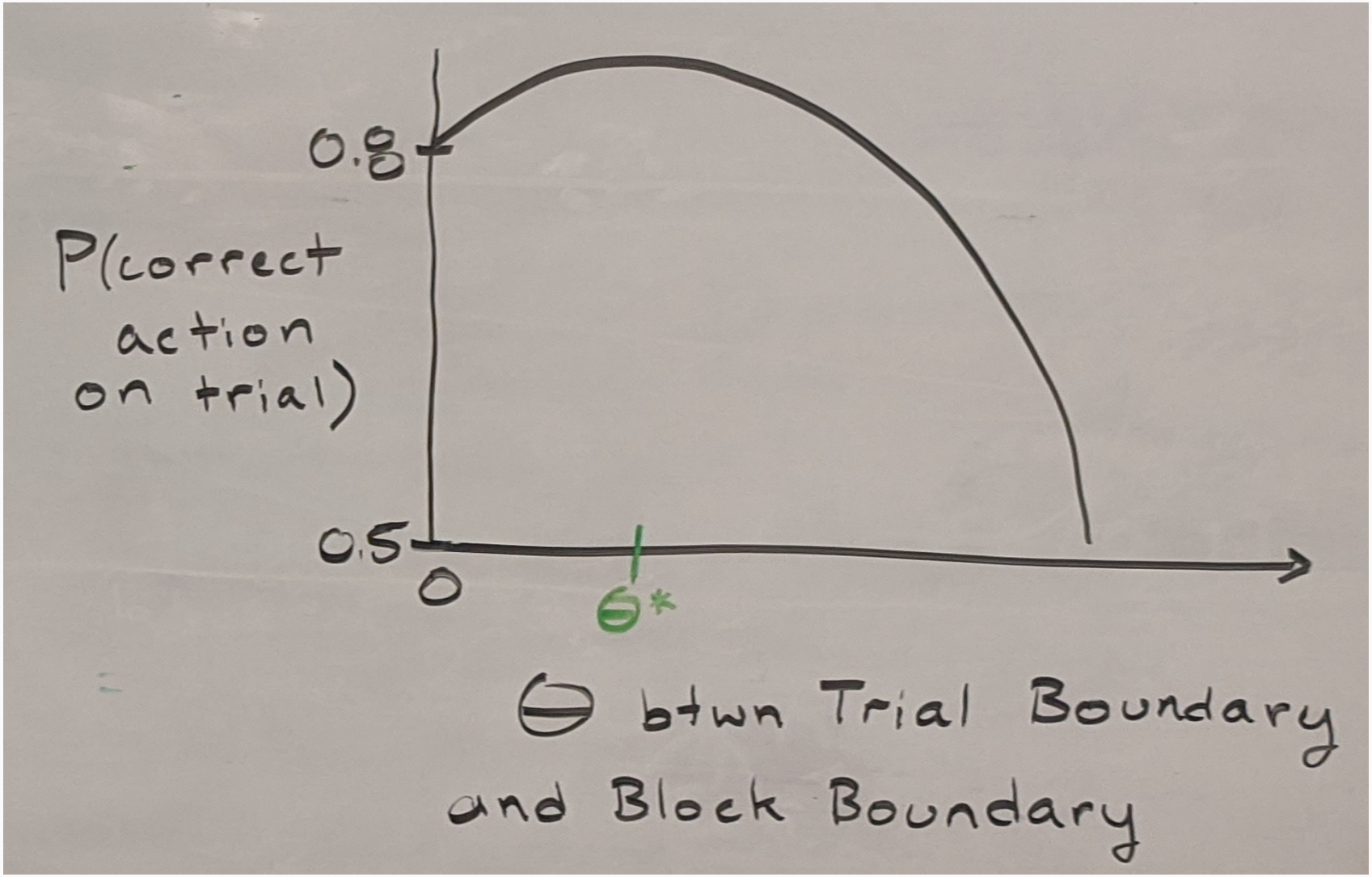


Preliminary Results - Block and Trial Encoding Mechanism

Figure 9: Idealized state space for encoding block side and trial side

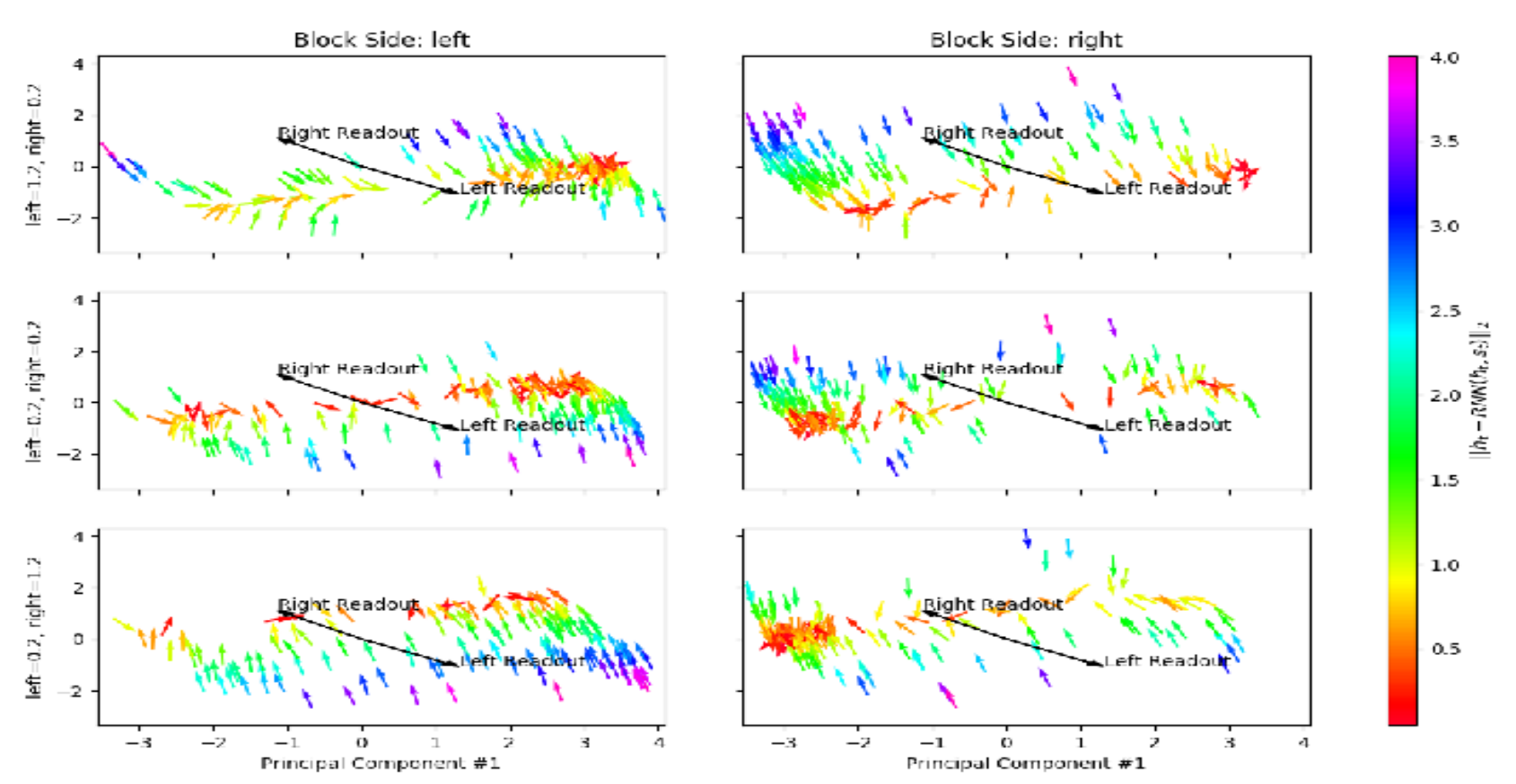


Preliminary Results - Predicted Effect of Rotating Readout



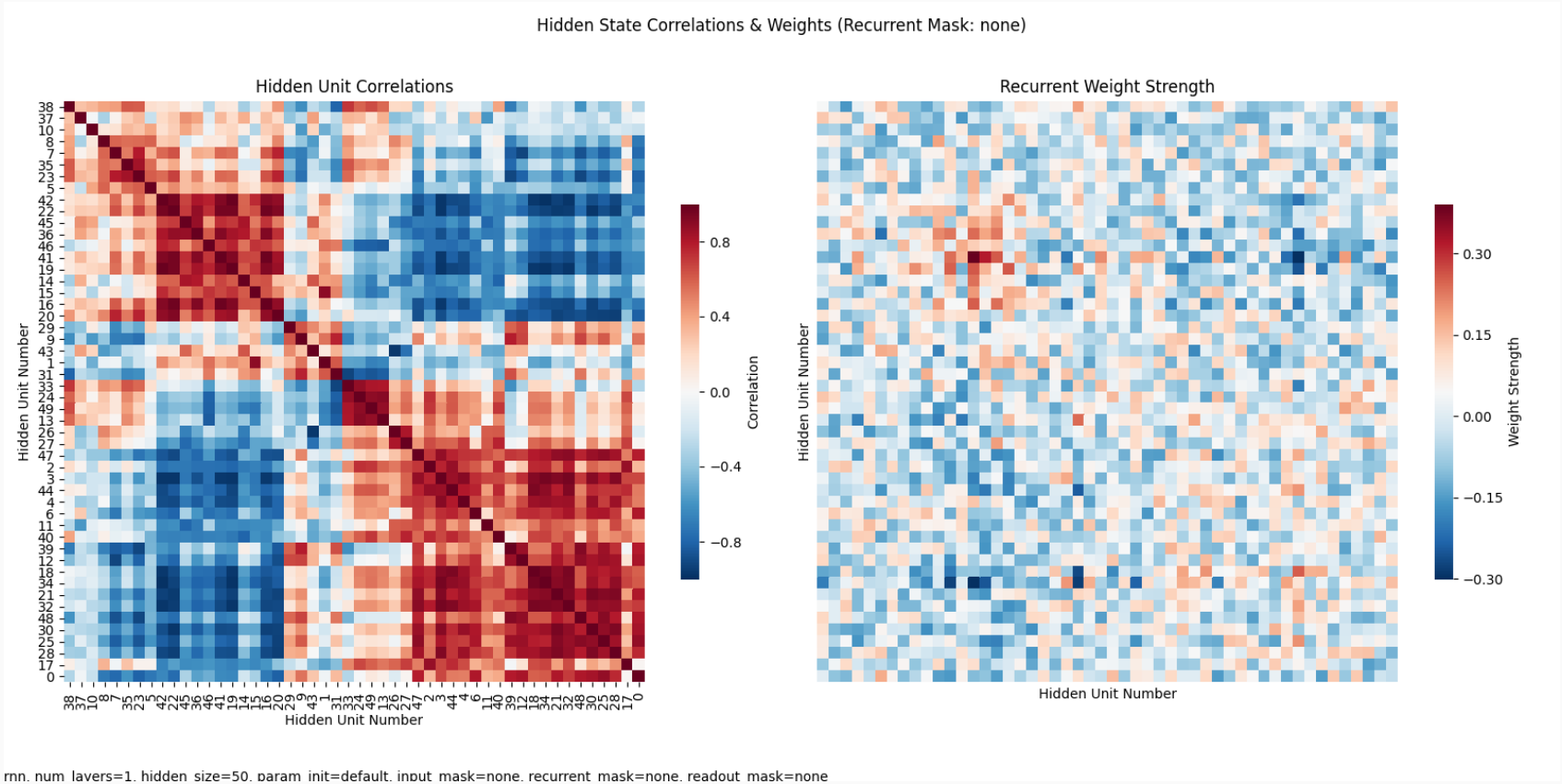
Preliminary Results - RNN State Movement

Figure 10: RNN state movement in response to strong left, ambiguous and strong right stimuli. Color: $||h_t - RNN(h_t, o_t)||_2$



Preliminary Results - Circuit for Bistable Attractor Dynamics

Figure 11: Correlation of RNN state units (left) and Recurrent Weights (right) reveal two self-excitatory, mutually-inhibitory populations



Questions?


DIFFICULT DAY YET I MAINTAIN COMPOSURE
HA THIS SMALL SETBACK AMUSES ME
NATHANWPYLE