A new multi-level modeling framework provides evidence for the simulation of object dynamics in the dorsomedial frontal cortex.

Daniel Calbick[1], Jason Z. Kim[2], Hansem Sohn[3], Mehrdad Jazayeri[4], Ilker Yildirim[1]

[1] Interdepartmental Neuroscience Program, Yale University; [2] Department of Physics, Cornell University; [3] Center for Neuroscience Imaging Research & Department of Biomedical Engineering, Sungkyunkwan University; [4] Department of Brain & Cognitive Sciences, MIT

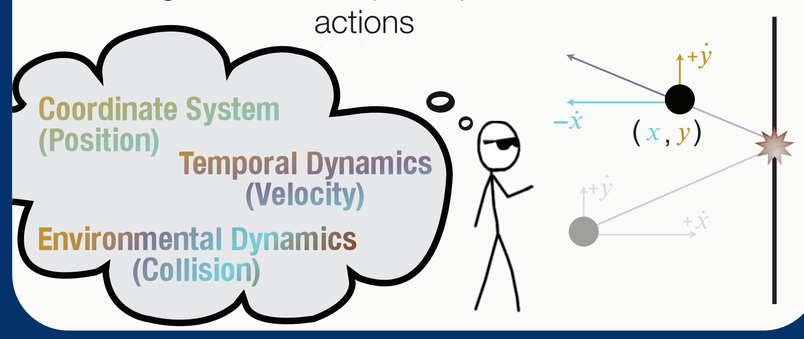
Email: daniel.calbick@yale.edu

Mental Simulation & World Models

Runnable mental models of our environments enable flexible planning.

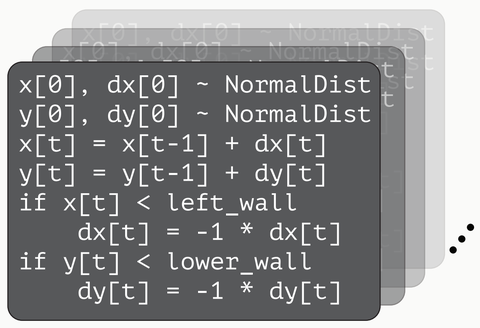
Two aspects:

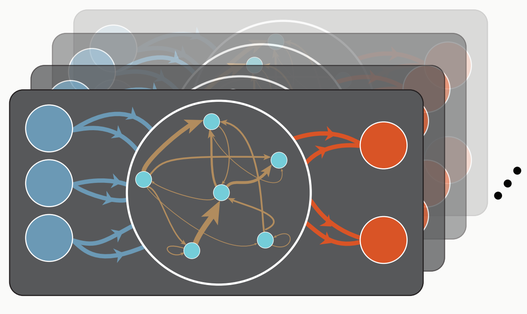
- Representation of objects and how they move and interact with each other
- Planning with these object representations to allow actions

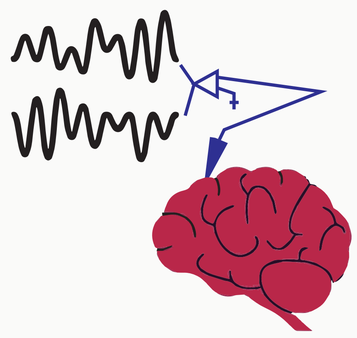


Hypothesis Driven Recovery of World Models in the Brain

Candidate Symbolic World Model ↔ Programmable Neural Networks (Mapable to neural implementation) → Empirical Testing in Neural Data





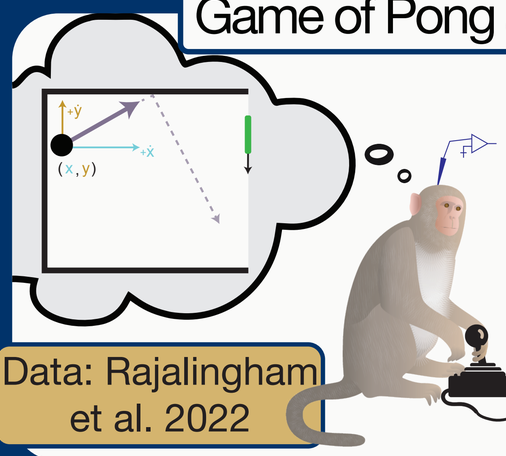


Key Issue: Existing frameworks do not allow for hypothesis-driven exploration of structured world models in neurobiology.

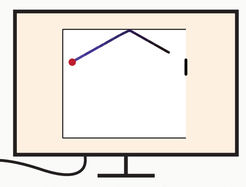
Probabilistic models of cognition: Allow for hypothesis-driven exploration but not relatable to neural data.

Task-optimized DNNs: Yield black-box hypothesis as a function of architectures, training sets, and objective functions

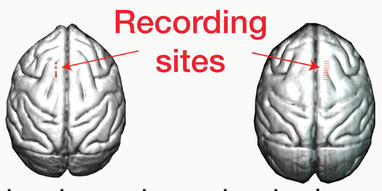
Game of Pong as Example World Model



79 Pong trials, played by 2 trained primates



Neurons recorded from dorsomedial frontal cortex (DMFC)

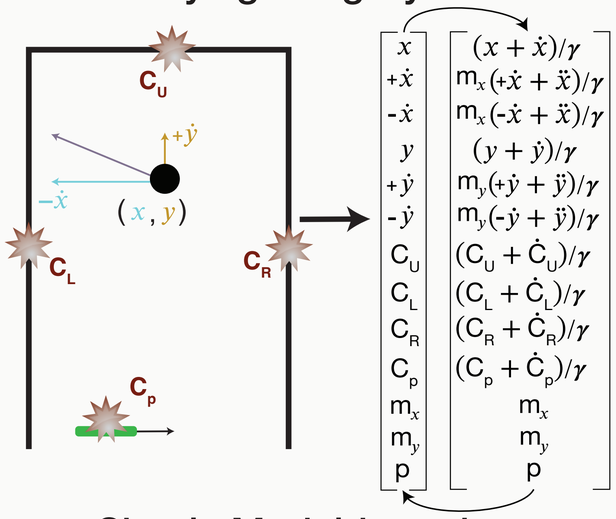


Data: Rajalingham et al. 2022

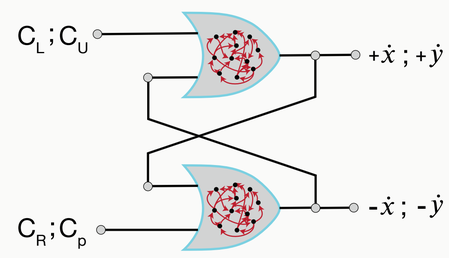
Hypothesis: Planning based on physical simulation of approximate object dynamics.

Programmed RNN World Model of Pong

Hypothesized World Model Codifying Pong Dynamics

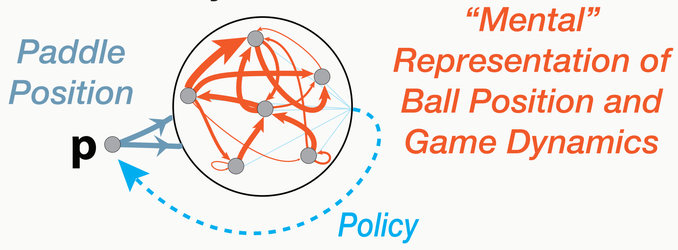


Collision Detection & Resolution via Programmed Neural S-R Latch Circuit

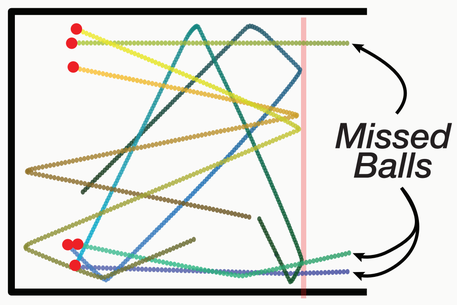


Programmed RNN OR Logic Gate

Simple Model-based Policy with Linear

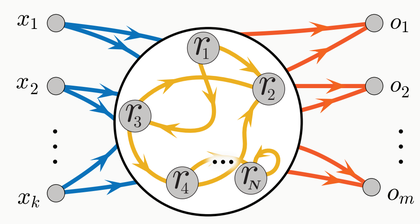


Example Gameplay



Symbolically Programming (not training) RNNs to Implement Hypothesis

1. Coarse Network Architecture & Equations

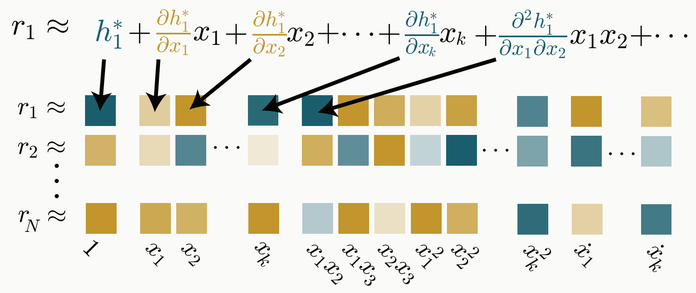


$$\frac{1}{\gamma}\dot{\boldsymbol{r}} = -\boldsymbol{r} + \boldsymbol{g}\,(A\boldsymbol{r} + B\boldsymbol{x} + \boldsymbol{d})$$

$$\boldsymbol{o} = W\boldsymbol{r}\ ;\ W = \underset{W}{argmin} \left\| W\mathrm{r} - o \right\|$$

2. Decomposed Network State (r)

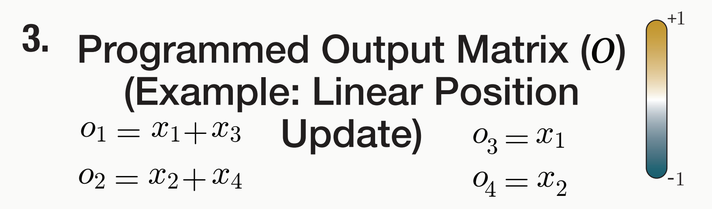
Taylor Series Decomposition



3. Programmed Output Matrix (O) (Example: Linear Position Update)

$o_1 = x_1 + x_3$ $\quad o_3 = x_1$

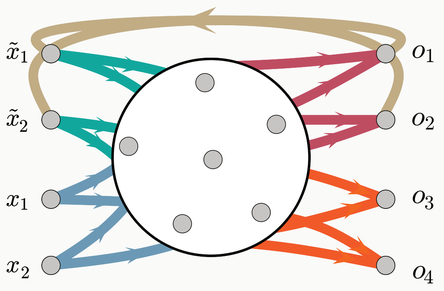
$o_2 = x_2 + x_4$ $\quad o_4 = x_2$



4. Scripting & Feedback

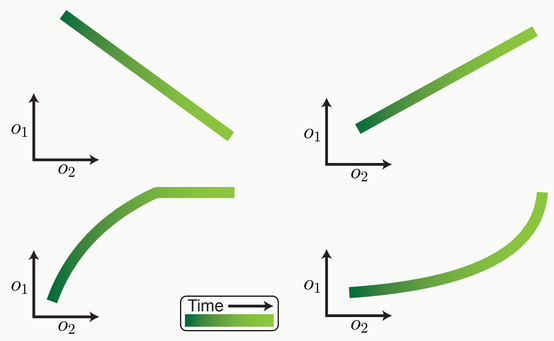
$$\frac{1}{\gamma}\dot{\boldsymbol{r}} = -\boldsymbol{r} + \boldsymbol{g}\,(\tilde{B}\,\tilde{W}\boldsymbol{r} + B\boldsymbol{x} + \boldsymbol{d})$$

$$A = \tilde{B}\tilde{W}$$



Feedback allows for the linking of simple programmed functions for more complex dynamics between symbolic variables.

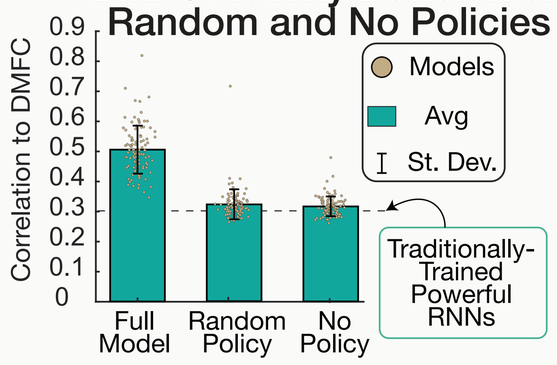
5. Simulated Trajectories from RNN with Time-varying Velocity Inputs



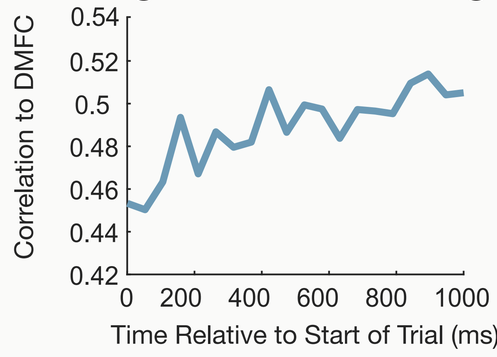
Method: Kim & Bassett 2023

Results

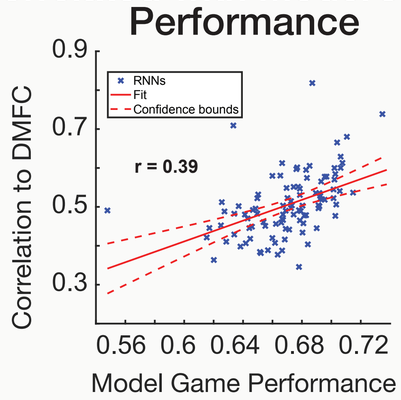
Programmed RNN explains DMFC Activity Better Than Random and No Policies



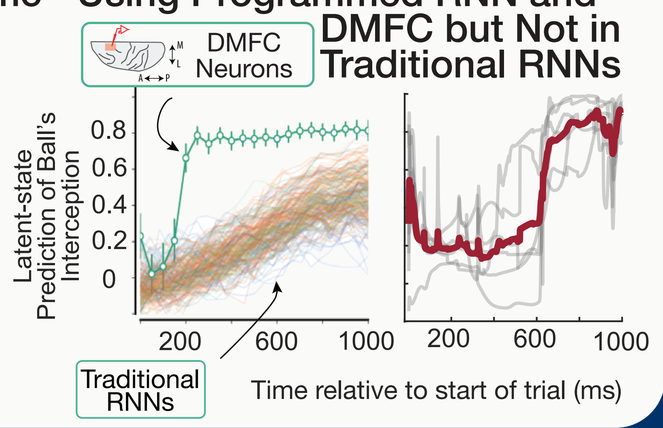
Similarity to DMFC Strengthens as Trial Progresses



Model-Brain Similarity is Correlated with Model's Game Performance



Non-linearity Seen in Prediction Using Programmed RNN and DMFC but Not in Traditional RNNs



Conclusions

- Evidence DMFC neurons may be simulating object dynamics
- Our findings highlight a new multi-level modeling framework and its potential for exploring world models in the brain.

References & Acknowledgements

Kim, J. Z., & Bassett, D. S. (2023). A neural machine code and programming language for the reservoir computer. Nature Machine Intelligence.

Rajalingham, R., Sohn, H., & Jazayeri, M. (2022b). Dynamic tracking of objects in the macaque dorsomedial frontal cortex. bioRxiv, 2022–06.

We thank the Yale Center for Research Computing for their support in managing the Milgram computing cluster.