Evidence accumulation in deep RL agents powered by a cognitive model

James Mochizuki-Freeman, Sahaj Singh Maini, Zoran Tiganj
Department of Computer Science, Indiana University Bloomington

MemARI
NeurIPS 2022 | December 2nd
New Orleans, USA

INDIANA UNIVERSITY

Summary

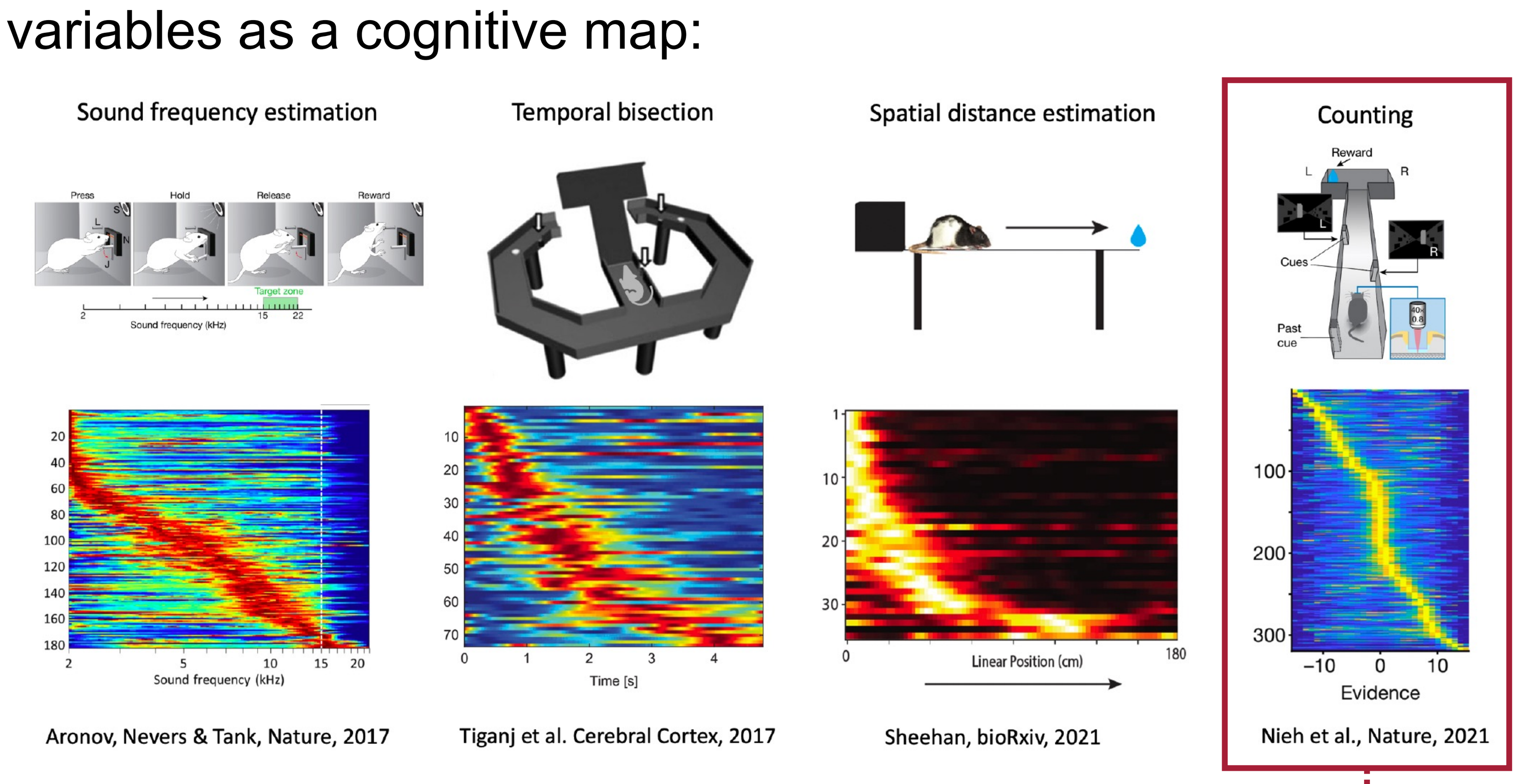
• We integrated a cognitive science model inside an RL agent and trained it to perform a simple evidence accumulation task inspired by behavioral experiments on animals.

• Compared to RNNs and GRUs, our agents were able to **learn faster** and **generalize better**, while also having **significantly fewer parameters**.

• Similar to the animal brain, our agents generated neural activity that reflects a low-dimensional ordered representation of evidence.

Motivation

The brain represents task-relevant physical and abstract variables as a cognitive map:



The **accumulating-towers task** has mice move down a virtual track, observing objects (towers) on the left and right, with the goal being to choose the side that had more towers.

Existing results indicate the existence of cells tuned to the **difference between number of towers**, the population of which tile the entire evidence axis [3].

Laplace Framework

The Laplace transform is implemented via an RNN with analytically-derived weights [1]:

$\alpha = \frac{\Delta n}{\Delta t}$ (change in numerosity)

$s = \begin{bmatrix} 1.6 & 1.366 & 1.167 & \cdots & 0.08 \end{bmatrix}$ (intrinsic decay rate – log-spaced)

$$F_{s;t} = \underbrace{W}_{\mathrm{diag}(e^{-\alpha(t)s\Delta t})} F_{s;t-1} + \overbrace{f_t}^{\delta(0)}$$

$\delta(0)$ [1 0 0 0 0 … 0]

$\mathrm{diag}(e^{-\alpha(t)s\Delta t})$ (recurrent weight matrix)

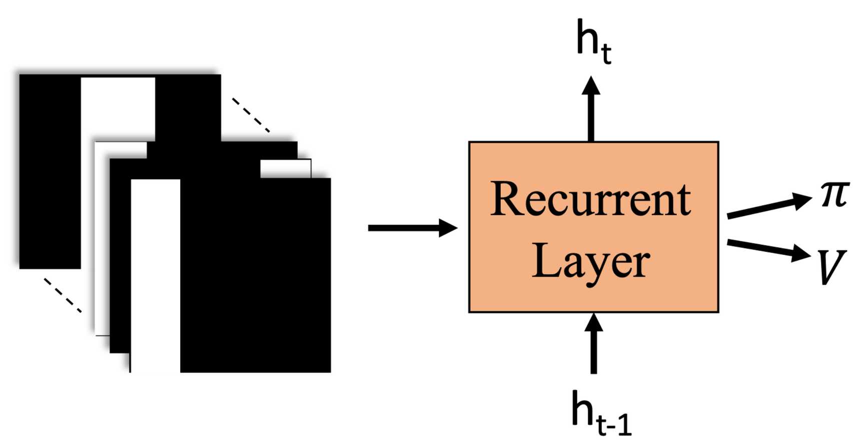
F_{sub} = The point-wise product of two F functions
(see [2] where it is illustrated how a point-wise product of two functions in the Laplace domain results in addition/subtraction in the original domain)

$$\tilde{f}_{n^*;n} = \overbrace{\mathbf{L}_k^{-1}}^{\frac{(-1)^k}{k!} s^{k+1} \frac{d^k}{ds^k}} F_{s;t}$$

$\frac{(-1)^k}{k!} s^{k+1} \frac{d^k}{ds^k}$ (Post's inversion formula)

$\tilde{f}_{sub}$ = The cross-correlation of two $\tilde{f}$ functions

Agent Architecture

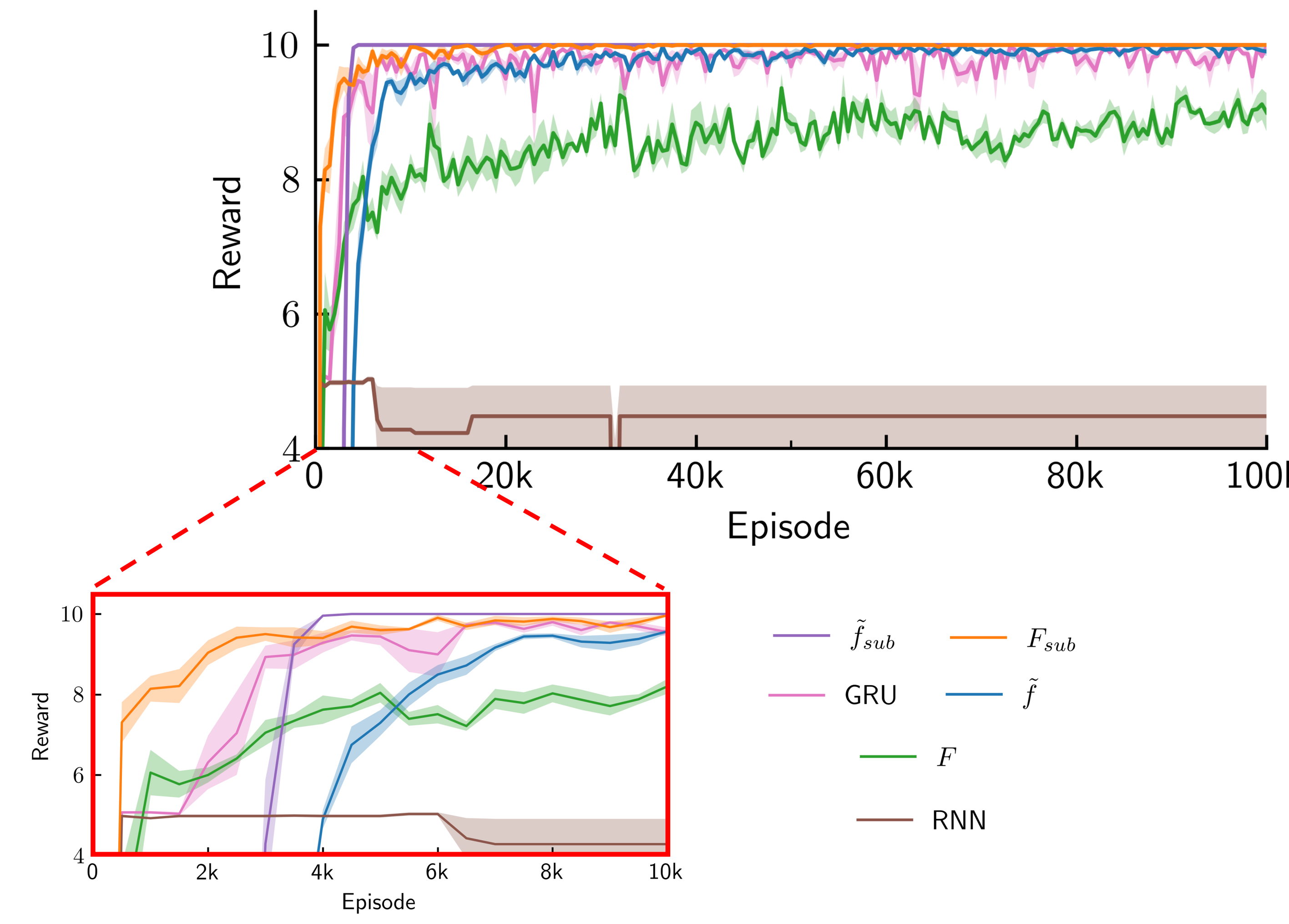


Recurrent Layer Types:

- $\tilde{f}_{sub}$
- $\tilde{f}$ (remove subtraction)
- F_{sub} (remove inverse Laplace transform)
- F (remove both)
- RNN
- GRU
- RNN_{FROZEN}
- GRU_{FROZEN}

Experiments

Our agents were the first to solve the task, and made the fastest initial progress:

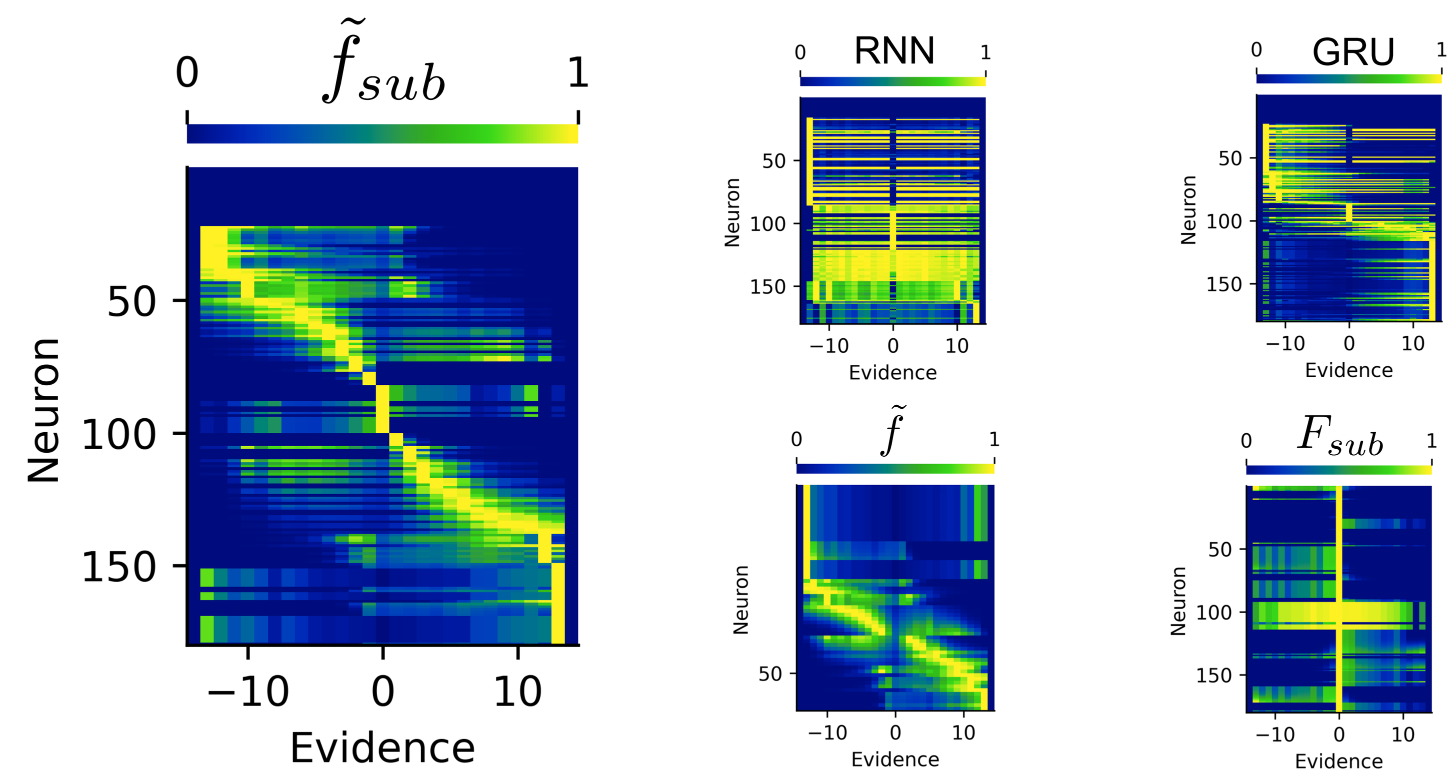


Our agents could generalize better to unseen track lengths:

	d=300	d=3000	d=10000	# Parameters
$\tilde{f}_{sub}$	**10.000 ± 0.000**	**10.000 ± 0.000**	**10.000 ± 0.000**	724
$\tilde{f}$	9.903 ± 0.066	9.925 ± 0.065	9.925 ± 0.041	**244**
F_{sub}	**10.000 ± 0.000**	**10.000 ± 0.000**	**10.000 ± 0.000**	724
F	8.995 ± 0.287	8.675 ± 0.361	8.900 ± 0.272	**244**
RNN	4.475 ± 0.456	4.600 ± 0.281	4.700 ± 0.576	34024
GRU	9.980 ± 0.000	9.600 ± 0.146	8.425 ± 0.504	100624
RNN_{FROZEN}	−21.0 ± 0.000	−201.0 ± 0.000	−601.0 ± 0.000	724
GRU_{FROZEN}	−14.473 ± 5.653	−149.3 ± 44.77	−449.4 ± 131.3	724

Table A1: Mean reward +/- standard error across four runs after 100k episodes of training in d=300 steps long environment. Validation was done in 300, 3000 and 10000 steps long environments.

The neural activity in our agents resembled the cell recordings from the mouse hippocampus:



References

1. M. W. Howard, C. J. MacDonald, Z. Tiganj, K. H. Shankar, Q. Du, M. E. Hasselmo, and H. Eichenbaum. A unified mathematical framework for coding time, space, and sequences in the hippocampal region. *Journal of Neuroscience*, 34(13):4692–707, 2014.
2. M. W. Howard, K. H. Shankar, and Z. Tiganj. Efficient neural computation in the laplace domain. In *Cognitive computations workshop at Advances in neural information processing systems*, 2015.
3. E. H. Nieh, M. Schottdorf, N. W. Freeman, R. J. Low, S. Lewallen, S. A. Koay, L. Pinto, J. L. Gauthier, C. D. Brody, and D. W. Tank. Geometry of abstract learned knowledge in the hippocampus. *Nature*, pages 1–5, 2021.

We gratefully acknowledge support from the Defense Advanced Research Projects Agency (DARPA) under project Time-Aware Machine Intelligence (TAMI) and the National Institutes of Health's National Institute on Aging, grant 5R01AG076198-02. This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute.