The Emergence of Abstract and Episodic Neurons in Episodic Meta-RL

Badr AlKhamissi [1] Muhammad ElNokrashy [2] Michael Spranger [1]

[1]Sony CSL [2]Microsoft EGDC

Abstract

We analyze the reinstatement mechanism introduced by Ritter et al. (2018) to reveal two classes of neurons that emerge in the agent's working memory (an epLSTM cell) when trained using episodic meta-RL on an episodic variant of the Harlow visual fixation task. Specifically, *Abstract* neurons encode knowledge shared across tasks, while *Episodic* neurons carry information relevant for a specific episode's task.

Task Summary

We develop a simplified symbolic version with exact parallels to the task structure of the Harlow visual fixation task found in the PsychLab environment (Leibo et al., 2018) but which factors out the visual and spatial modeling of the environment.

Figure 1. Illustration of the 1D Symbolic Harlow task. **(a)** Fixation cross at the center of agent's receptive field. **(b)** Objects placed in agent's receptive field upon fixation. **(c)** Top-down view of agent in the environment.

The task consists of a one-dimensional circular state-space with 16 discrete cells. The agent can observe 8 cells at any time step (see Figure 1 for an illustration of the task). The agent can then select one of two actions $|\mathcal{A}| = 2$: move one cell to the *left*, or to the *right*. The goal is for the agent to learn to choose the rewarding object by orienting it towards the center of its receptive field, after fixating on the cross.

The Episodic LSTM Cell

We use the reinstatement mechanism from Ritter et al. (2018)—defined as:

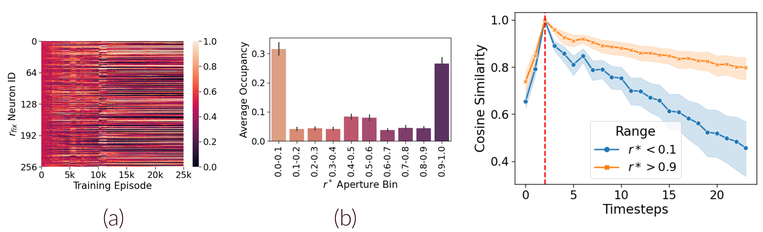
$$\mathbf{r}_t = \sigma\left(\mathbf{W}_{rx}\mathbf{x}_t + \mathbf{W}_{rh}\mathbf{h}_{t-1} + \mathbf{b}_r\right) \tag{1}$$

where $\mathbf{r}_t$ controls the flow of information from the retrieved memory $\mathbf{m}_t$ into the epLSTM cell state:

$$\mathbf{c}_t = \mathbf{i}_t \odot \tilde{\mathbf{c}}_t + \mathbf{f}_t \odot \mathbf{c}_{t-1} + \underline{\mathbf{r}_t \odot \tanh\left(\mathbf{m}_t\right)} \tag{2}$$

Recurrent Neurons and the Reinstatement Gate

We note that since $\mathbf{r}_t[j] \in (0, 1)$ (by the σ function) modulates the reinstatement of individual neurons from $\mathbf{m}_t$, we may interpret $\mathbf{r}_t[j]$ as the importance of neuron j for recurring episodic information.



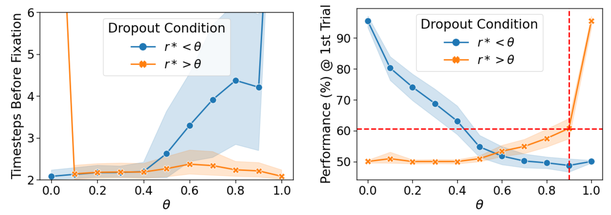
(a) (b) (c)

Figure 2. **(a)** The values of $\mathbf{r}_{\text{fix}}[j]$ for each neuron j across training episodes. **(b)** Average percentage of neurons in $\mathbf{r}^*$ that appear within a bin of "openness" during testing across 30 different seeds. **(c)** Average vector similarity between $\mathbf{c}_{\text{fix}}$ at the first fixation in an episode (vertical line), and $\mathbf{c}_t$ at every other step.

About 25% of the neurons have become biased to be open ($\mathbf{r}^*[j] \geq 0.9$), while $\sim 30\%$ are biased to be closed ($\mathbf{r}^*[j] < 0.1$).

What Happens When You Squeeze Out Either Type

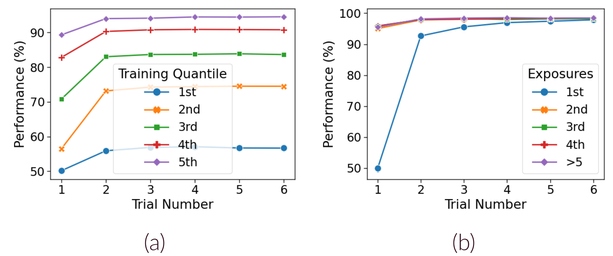
To clarify the roles of individual neurons in $\mathbf{c}_t$, we test while gradually masking out the neurons $\mathbf{c}_t[j]$ based on $\{j : \mathbf{r}^*[j] \lesseqgtr \theta\}$ then analyze the behavioral change in two signals: **(a)** number of steps before fixation, and **(b)** first trial performance.



(a) Time steps (b) Performance

Figure 3. When dropping *episodic* ($\mathbf{r}^* \geq \theta$) or *abstract* ($\mathbf{r}^* < \theta$) neurons at θ: **(a)** Average time-steps before fixating. **(b)** Average first trial performance. Note the 30% regression when dropping *episodic* neurons at $\theta = 0.9$ (dashed lines).

Task Performance



(a) (b)

Figure 4. Training and testing Performance **(a)** Average training performance at each trial number, per training quantile. **(b)** Testing performance at each trial number, per number of exposures to a specific task.

The performance on the first trial improves and is not stuck at random as in the classical Harlow experiment because the agent is able to reinstate relevant information when it re-encounters a specific task. While at testing time it can be seen that when a task reoccurs the agent immediately identifies it and is able to solve it from the first trial.

What's Next?

The findings in this paper implies that one does not need to store the whole cell-state when committing it to the long-term memory module since only a fraction of the activations are actually going to be reinstated. In the case of the experiments conducted in this paper, this method can save up to 75% of the storage cost for the memory module while maintaining close to optimal performance after deployment once $\mathbf{r}^*$ is computed.

Wang et al. (2018) had shown that the meta-RL framework has direct connections with structures and functions in the brain. Inline with this theory and the work presented in this paper, previous work have shown that the PFC contain single neurons that encodes abstract rules (Wallis et al., 2001). Future work may extend the analysis to different episodic tasks, and utilize the findings for incorporating stronger inductive biases. We hope this work meaningfully furthers the sharing of insights between the neuroscience and machine learning fields.

References

Leibo, J. Z., de Masson d'Autume, C., Zoran, D., Amos, D., Beattie, C., Anderson, K., Castañeda, A. G., Sanchez, M., Green, S., Gruslys, A., Legg, S., Hassabis, D., and Botvinick, M. (2018). Psychlab: A psychology laboratory for deep reinforcement learning agents. *CoRR*, abs/1801.08116.

Ritter, S., Wang, J. X., Kurth-Nelson, Z., Jayakumar, S. M., Blundell, C., Pascanu, R., and Botvinick, M. M. (2018). Been there, done that: Meta-learning with episodic recall. In *ICML*.

Wallis, J., Anderson, K., and Miller, E. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature*, 411:953–6.

Wang, J. X., Kurth-Nelson, Z., Kumaran, D., Tirumala, D., Soyer, H., Leibo, J. Z., Hassabis, D., and Botvinick, M. (2018). Prefrontal cortex as a meta-reinforcement learning system. *bioRxiv*.