Title **One-shot learning of paired associations by a reservoir computing model with Hebbian plasticity**

300 word summary One-shot learning can be achieved by animals and algorithms, but how animals do it is poorly understood as most of the algorithms are not biologically plausible. Experiments studying one-shot learning in rodents have shown that after initial gradual learning of associations between cues and locations, new associations can be learned with just a single exposure to each new cue-location pair. Foster, Morris and Dayan (2000) developed a neuro-symbolic actor-critic and coordinate learning agent that exhibited one-shot learning to displaced single locations in an open field maze using dead reckoning. While the temporal difference rule for learning the agent's coordinates was biologically plausible, the agent's memory mechanism for learning target coordinates was not, nor did they address one-shot learning of multiple cue-location pairs that rodents are also capable of (Tse et al., 2007). Here we extend the biological plausibility of that agent by replacing the symbolic memory mechanism with a reservoir of recurrently connected neurons resembling cortical microcircuitry. Biologically plausible learning of goal coordinates was achieved by subjecting the reservoir's output weights to synaptic plasticity governed by a novel 4-factor variant of the exploratory Hebbian (EH) rule gated by reward. The agent's current coordinates and goal coordinates were passed to a pretrained neural network that performed vector subtraction and selected the direction of movement towards the target. Our fully neural agent trained by Hebbian plasticity combines functions thought to involve the hippocampus and prefrontal cortex such that the memory system can store in one shot goal coordinates that can be recalled when a relevant cue is presented, while the coordinate system acts as a cognitive map encoding relational information for goal directed dead reckoning. As with rodents, the biologically plausible agent exhibited one-shot learning in the multiple cue-location paired associations task of Tse and colleagues.

Central Question Animals can flexibly adapt previous learning to solve new tasks within a few trials. How they do so is poorly understood, as most machine learning algorithms for few-shot learning are not biologically plausible. Tse and colleauges (2007) showed that rodents gradually learned multiple cue-location paired associations, after which they learned each of two novel cue-location pairs after just a single trial[1]. This one-shot learning depended on the hippocampus and correlated with activity changes in the prefrontal cortex. Foster and colleagues (2000) developed a neuro-symbolic actor-critic and coordinate learning agent that was able to perform one-shot navigation to single displaced locations using dead reckoning[2]. However, the neuro-symbolic agent was only able to navigate to single targets. Here, we develop a fully neural agent that solves the Tse and colleagues' multiple paired association task with one-shot learning of novel paired associations.

Approach The task was modelled with 2 parts, learning of 6 cue-location original paired associates (OPA) over 17 sessions and learning of 2 novel paired associates (NPA) or 12 random cue-location pairs for a single session with a single trial per cue, followed by a non-rewarded probe session. Agents randomly started from the north, south, east or west walls and had to navigate in a 1.6 m square maze to the correct target spanning 3 cm in radius. Trials ended after 600 s with each simulation timestep representing 100 ms.

The **first** agent was an actor-critic (Fig. 1A) with a single nonlinear hidden layer with weights trained by backpropagation. The **second** agent is a modified coordinate learning system[2] (Fig. 1B) with episodic memory, learning of self-position, and a symbolic vector subtraction motor controller. The agent learned its position $p(t)$ as a 2D coordinate vector using dead reckoning derived temporal difference error (Eq. 1)

$$\delta(t) = -\Delta a(t) + p(t) - p(t - \Delta t) \quad (1)$$

Cue-coordinate associations were handcrafted such that the symbolic memory module stored cues and current coordinates $p(t)$ as goal coordinates $g(t)$ when the agent was rewarded. When queried with a cue in the next trial, the associated goal coordinates were recalled using a distance-based metric. Thereafter, vector subtraction was performed between $g(t)$ and $p(t)$ to choose a movement direction $q(t)$ to reach the goal by direct heading. The **third** and **fourth** agents use a reservoir (Fig. 1C) with two reservoir readout units $g(t)$, representing X and Y coordinates, in place of the memory module and a pretrained neural network that performs vector subtraction between current $p(t)$ and goal $g(t)$ coordinates to select direction of movement. The readout weights were either trained by reward modulated perceptron rule or the exploratory Hebbian (EH) rule[3] (Eq. 2) to store the agent's coordinates $p(t)$ as goal coordinates $g(t)$ only when the reward was disbursed

$$\Delta W^{g}(t) = r(t) \cdot (g^{noisy}(t) - \bar{g}(t)) \cdot M(t) \cdot R(t) \quad (2)$$

We noticed that when the reservoir was queried with the same cue in the subsequent trial, the readout units were able to recall the correct goal coordinates, demonstrating one-shot learning of cue-coordinate pairs.

Results All three agents gradually learned to navigate to the correct location given the cue – exhibiting decreased latency (Fig. 1E) and increased amounts of time spent at the correct location (Fig. 1F). When agents were exposed to two, six (Fig. 1G) or 12 NPAs (Fig. 1H) for a single trial per cue, only the symbolic and reservoir agents spent higher amount of time at the correct location during the probe trials (chance = 16.7%).

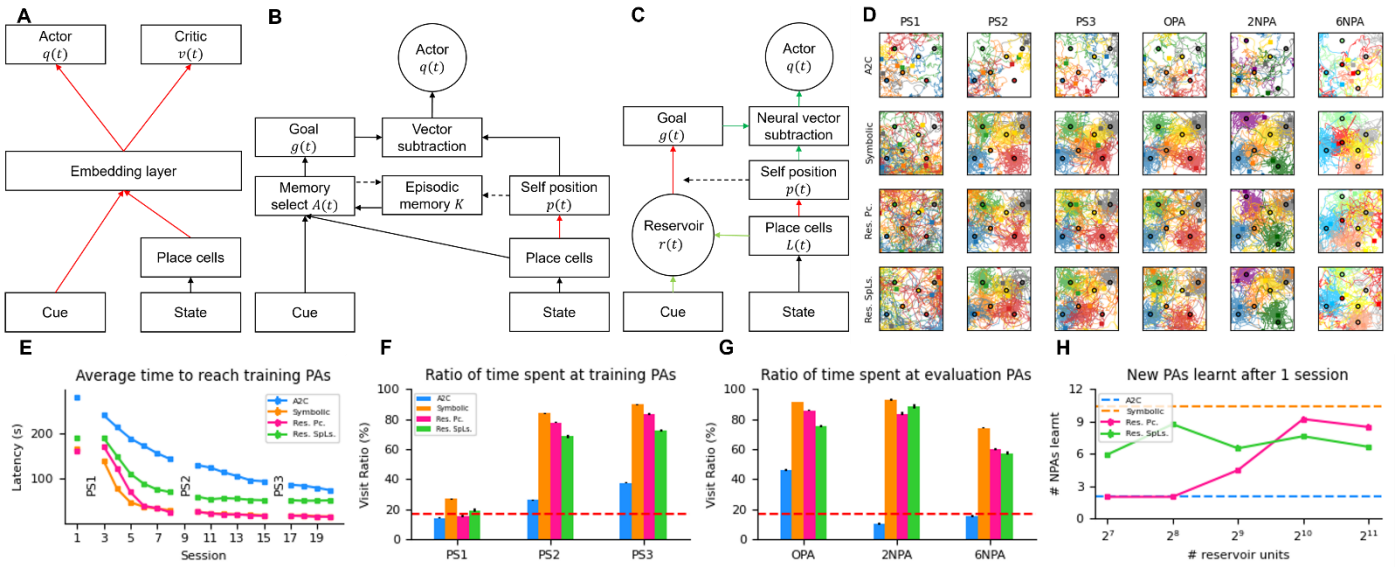

Figure 1. Biologically plausible agent & one-shot learning of multiple paired associates. A) Actor-critic (A2C) agent trained by backpropagation. B) Symbolic agent based on Foster et al (2000). **C)** Agent with reservoir trained by reward modulated Hebbian plasticity and a neural vector subtraction module. **D)** Example trajectory for all agents during probe session PS1, PS2, PS3 and after one OPA, 2NPA, 6NPA training session. **E)** Decreasing latency to reach OPA cue-location pair. **F)** Increasing amount of time spent at correct cue-location during probe sessions. Chance performance = 16.7%. **G)** Only symbolic and reservoir agents show one-shot learning performance for two (2NPA) and six (6NPA) novel pairs. **H)** Symbolic (orange) and A2C (blue) agents learned 11 and 2 PAs respectively after one session. Reservoir agents' (pink and green) learned more PAs in one-shot (2 to 9) when number of reservoir units were increased. EH rule (green) allowed reservoir with lesser units to learn more PAs in one-shot.

Conclusion

Canonical reinforcement learning agents, such as the actor-critic, update their policy incrementally, restricting their one-shot learning abilities. The one-shot learning demonstrated by Tse and colleagues can be replicated by having two systems trained by Hebbian plasticity, an episodic memory system to store and recall goal coordinates after a single trial, and the gradual learning of self-coordinates during the trial. Additional neural systems such as a pretrained vector subtraction module facilitates direct heading to goals. However, the rule or schema that different cues are associated to specific locations is handcrafted in the current model. Learning such a schema using biologically plausible learning rules remains to be demonstrated. Nevertheless, our fully neural agent replicates the major one-shot learning phenomenon described by Tse and colleagues (2007).

References

1. Tse, D. *et al.* Schemas and Memory Consolidation. *Science.* **316**, 76–82 (2007).

2. Foster, D. J., Morris, R. G. & Dayan, P. A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* **10**, 1–16 (2000).

3. Hoerzer, G. M., Legenstein, R. & Maass, W. Emergence of complex computational structures from chaotic neural networks through reward-modulated hebbian learning. *Cereb. Cortex* **24**, 677–690 (2012).