Title: A Hippocampus-Inspired Episodic Memory for Reinforcement Learning

Introduction: Episodic memories—memories of real, personal experiences—are fundamental to human intelligence. For instance, an attentive fire drill participant can recall *perceptually rich* data (i.e., what the escape path looks like) after just one or two exposures (an ability called *few-shot learning*) and can prevent new memories from overwriting this perceptually rich data (an ability called *continual learning*). Existing neuroscience models of episodic memory have primarily been evaluated in situations that omit at least one of these three factors (few-shot learning, continual learning, and perceptual richness)[1]. I propose the evaluation of one such model in a novel environment that stresses all three. Furthermore, I hypothesize that this model will be more memory efficient than existing AI techniques for episodic-like memory.

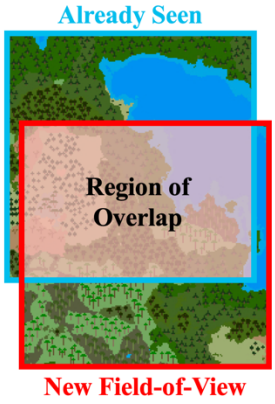
Background: Much evidence from neuroscience points towards the *Cornu Ammonis* 3 (CA3) region of the hippocampus as the locus of the episodic memory system[2]. Unlike other areas of the brain, which contain dense, localized neural connections (i.e., neurons primarily connect to other nearby neurons), CA3 contains sparse, global connections that allow it to effectively act as a single network. One leading theory suggests that CA3 behaves as an *associative* network, a type of neural network that performs *pattern completion* by recovering previously saved patterns (episodic memories in this case) from noisy or incomplete inputs during memory recall[2]. The same theory posits that the *sparsity* of CA3 might enable *pattern separation*, the act of magnifying the differences between individual episodic memories to prevent them from overwriting one another (i.e., to enable continual learning)[2].

Researchers within reinforcement learning (RL), an AI subfield that studies how agents learn from rewards, are also studying the episodic memory system. "Episodic control"—the use of episodic memory for choosing actions—may enable RL agents to learn from small amounts of data (i.e., few-shot learning)[3,4]. Typically, episodic control utilizes explicitly saved "event tuples" of the form $(\boldsymbol{s}, \boldsymbol{a}, \hat{Q})$, where $\boldsymbol{s}$ represents a previously encountered environmental state, $\boldsymbol{a}$ represents the action taken in that state, and $\hat{Q}$ approximates how "rewarding" taking action $\boldsymbol{a}$ in state $\boldsymbol{s}$ was. RL agents empowered by episodic control can recall and use these personal experiences when deciding which actions to take.

Motivation: Research in neuroscience on episodic memory is oftentimes "carried out in settings that are far removed from ecological contexts of decision making"[1]. An example of an ecologically realistic context would be the fire drill detailed above: it requires few-shot, continual learning of perceptually rich data (spatial/navigational data, which often requires CA3 involvement[2]). The behavior of neuroscience models of episodic memory in ecologically realistic contexts is not fully understood.

RL research on episodic memory is frequently carried out in more complex environments (e.g., video games), however existing models[3,4] are memory intensive. They explicitly save "tables" of event tuples, writing each one into its own memory region. This *tabular* saving procedure differs from that of *associative* memory, wherein patterns are written into overlapping memory regions (i.e., into a distributed network of neurons). Associative algorithms can save patterns with greater memory efficiency[5].

Methodology: To simultaneously address these problems, I propose the deployment of a hippocampus-inspired associative memory into an RL environment that requires few-shot, continual learning of perceptually rich (navigational) data.



Intellectual Merit: *My core hypothesis is that sparsity will enable an associative memory to perform as well as a tabular memory while utilizing less storage space*. If true, this work may provide a foundation for the development of more capable episodic memory systems in RL agents. Additionally, this project may demonstrate the effectiveness, or lack thereof, of neuroscience models in ecologically realistic environments.

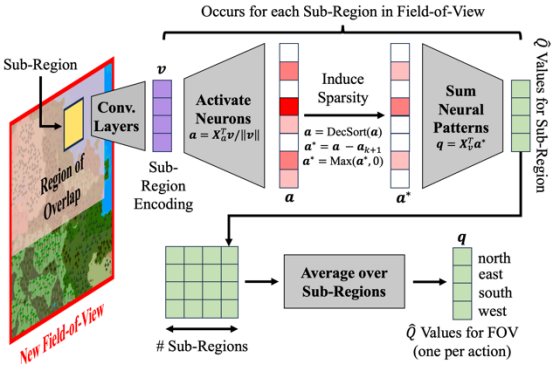
Task Design: For an initial study, I propose a novel task called Cache Recovery, which is inspired by a well-studied phenomenon in scrub jays (*Aphelocoma coerulescens*). Remarkably, scrub jays remember both where and when they cached their food, leading scientists to believe they possess a form of "episodic-like" memory[6]. Similarly, Cache Recovery will test RL agents on their ability to navigate to an artificial cache.

Cache Recovery will be built on a two-dimensional, procedurally generated map. Upon bootup, the task will choose an arbitrary location on the map to serve as the cache. The cache will be brightly colored and will emit a reward signal into its immediate environment. This will enable the agent to learn to "recover" the cache (navigate to it) from various nearby locations. Upon recovery (or failure), the agent will begin a new episode at a new location on the same map. After an initial learning period, Cache Recovery will begin placing the agent further and further from the cache. Due to partial observability, the increasing placement distance will require the agent to make navigational decisions without being able to directly see the cache. Instead, it will need to recognize previously encountered geographical regions that partially overlap with its current field of view. Cache Recovery is challenging due to three factors:

(1) Few-shot Learning: To keep up with the increasing placement distance, the agent will need to learn new geographical regions after only a small number of exposures to each one.
(2) Continual Learning: The agent must learn new geographical regions without forgetting old ones.
(3) Perceptual Richness: As the model encounters more geographical regions, it must recall more details about each one to distinguish between them. The memories thus need to be perceptually rich.

Architecture: I plan to use a Sparse Distributed Memory (SDM) architecture, which has proven effective for continual learning in non-RL settings[7]. Like CA3, SDM is an associative memory that uses sparsity to enable pattern separation and completion.

I propose the integration of SDM into a Deep Q Network (DQN)[8], a non-sparse neural network architecture. To test both the DQN and my proposed Sparse DQN (S-DQN) architectures for their few-shot, continual learning capabilities, my experiment will not provide either network with a memory table (which is atypical for a DQN). Instead, each model will adjust its weights to more accurately estimate the sum of rewards ($\hat{Q}$ values) from the most recent episode.



The DQN architecture encodes an observation $\boldsymbol{o}$ using a convolutional neural network (CNN) and applies a linear transformation to output one $\hat{Q}$ value for each possible action. I plan to replace one CNN filter with an SDM. When the SDM is "convolved" over a previously seen sub-region, it will recall the $\hat{Q}$ values previously learned for that sub-region. Then, S-DQN will average the $\hat{Q}$ values over all sub-regions to obtain $\hat{Q}$ values for the entire field of view.

Experiments: For my baseline, I plan to compare S-DQN against Neural Episodic Control (NEC)[5], a tabular method for episodic control, and a DQN[8]. Additionally, I will examine ablated and hybridized versions of all three architectures to isolate which components drive performance. I hypothesize that the DQN (with no memory table) will not perform well on Cache Recovery due to its lack of sparsity and resultant inability to perform pattern separation and completion. In contrast, I hypothesize that NEC will perform well on Cache Recovery but will demand far greater memory to do so than S-DQN.

By manipulating the rate at which the placement distance increases, I plan to directly compare the learning effectiveness and efficiency of these three architectures on episodic-like tasks. I also plan to measure performance decay as a function of placement distance, which will yield insights regarding the memory capacity of S-DQN on ecologically realistic data.

Broader Impacts: This multidisciplinary research project holds significant potential for building capable, interpretable AI systems. For instance, I am currently collaborating with the U.S. Army to develop autonomous vehicles that can effectively communicate their navigational decisions to passengers. A self-driving AI equipped with episodic-like memory may learn navigational information more effectively. Additionally, such an AI will retain more human-like memories of the world, which may seem more familiar (and thus more easily interpretable) when communicated to a human passenger.

References: [1]Zhou, C., et al. 2023, *Preprint*. [2]Rolls, E.T. 2010, *Behav. Brain Res*. [3]Blundell, C., et al. 2016, *Preprint.* [4]Pritzel, A., et al. 2017, *ICML.* [5]Sugandha, S., et al. 2022, *ICML.* [6]Clayton, N.S., & Dickinson, A. 1998, *Nature*. [7]Bricken, T., et al. 2023, *ICLR.* [8]Mnih, V., et al. 2013, *NIPS.*