A Grid Cell-Place Cell Scaffold Allows Rapid Learning and Generalization at Multiple Levels on Mental Navigation Tasks

Jaedong Hwang* **Sujaya Neupane*** **Mehrdad Jazayeri**† **Ila Fiete**†

MIT

*: Equal Contribution
†: Equal Correspondence

Abstract

We investigate the role of cognitive maps and the hippocampal-entorhinal architecture in mental navigation (MNAV) by building a neural network model. The model uses a continuous-time recurrent neural network (CTRNN) for action decisions and a hippocampal-entorhinal model network, MESH (Memory network with Scaffold and Heteroassociation), for encoding and learning maps. The model is trained on a navigation-to-sample (NTS) task and tested on NTS in a MNAV setting (no sensory feedback) in five different environments (image sequences). The CTRNN with MESH tackles MNAV by reconstructing the next image via path integration and vastly outperforms the CTRNN alone in both tasks, showing better generalization to unseen pairs within each environment and faster adaptation to new environments. The study demonstrates the importance of hippocampal cognitive maps in enabling data-efficient and generalizable learning in the brain.

Keywords: visual learning; mental navigation; grid cell; generalization

Introduction

Cognition involves organizing experiences into retrievable knowledge for novel mental computations, which is achieved through cognitive maps encoding spatial, temporal, and abstract relationships. Spatial contexts have been extensively studied, with sensory experiences driving spatially selective responses in the hippocampus and entorhinal cortex (O'Keefe & Dostrovsky, 1971). On the other hand, when animals encounter a new task that conceptually matches a previously seen task, it is common to observe rapid learning even if surface-level details and inputs differ. Such learning is believed to involve a transfer of conceptual understanding to the new task, but neural models of such generalization are lacking.

When humans and animals are trained to navigate through spatially laid out landmarks, they can learn the spatial map of the environment and reuse this learned knowledge of the spatial structure to rapidly generalize in a novel environment (Behrens et al., 2018). Conventional recurrent networks do not succeed in these types of generalizations. We hypothesize that a structured neocortical-entorhinal-hippocampal circuit, the Memory Scaffold with Heteroassociation (MESH) adapted from Sharma et al. (2022) with grid cell modules can enable such generalizations. We build a multi-region brain model, using a continuous-time recurrent neural network (CTRNN) to decide actions and the hippocampal MESH network to encode and learn maps. The learning rules in MESH are online and associative, based on velocity input and external cues. The outputs of MESH drive the action network. We sequentially trained the model on five image sequences in the visual navigation setting. Our model achieved the same performance in visual and mental navigation tasks while CTRNN without MESH failed at mental navigation. The model exhibited better generalization to unseen pairs in each environment and adapted to new environments faster than the baseline. Our work is thus a step toward a whole-system understanding of how the brain performs highly data-efficient and generalizable learning.

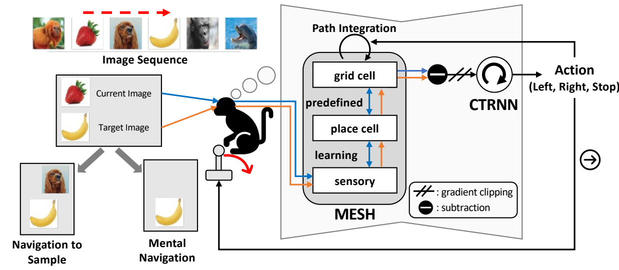

Figure 1: The agent explores the sequence of images with blank intervals. On each trial, the input to the network is a pair of start and target images. The images are encoded as grid codes by MESH and fed into CTRNN to decide an action; move in one direction or stop the trial. As in the animal experiment, the start image is continuously updated in navigation-to-sample task but masked in mental navigation task.

Mental Navigation

Neupane et al. (2022) developed a mental navigation task for monkeys. Monkeys are trained on a sequence of six images (landmarks) (Figure 1). Given a start and a target image, they must use a joystick to move between them (navigate-to-sample or NTS task). After reaching a performance criterion in NTS, the monkeys were introduced to the mental MNAV version of the task (MNAV). In MNAV, the image sequence was occluded and only the start and the target landmark were visible before joystick deflection. The sequence was hidden throughout the trial, including during joystick navigation. To solve the task, the animals had to rely on their memory of relative landmark positions and navigate without sensory feedback. The monkeys successfully learned to perform the MNAV task, and the produced vectors closely matched the actual vectors in terms of magnitude and direction.

Method

We built a multi-region brain model, using a CTRNN-based network to decide actions, and the entorhinal-hippocampal MESH network (Sharma et al., 2022) to encode and learn maps that associate observations with the grid cell scaffold. The learning rules for map formation in MESH run online, based on velocity (action) input and external cues (sensory inputs). MESH is composed of three layers; sensory (input), place cell, and grid cell layers. The grid code (phase code) is formulated as a k-hot vector imposed by local recurrent inhibition, where k is the number of modules and each module has a different period. All codes are paired with place cell activation before training and the sensory input is associated with each grid code by pseudo-inverse learning between the input and place cells. Please refer to Sharma et al. (2022) for more details. The association between the current image and the

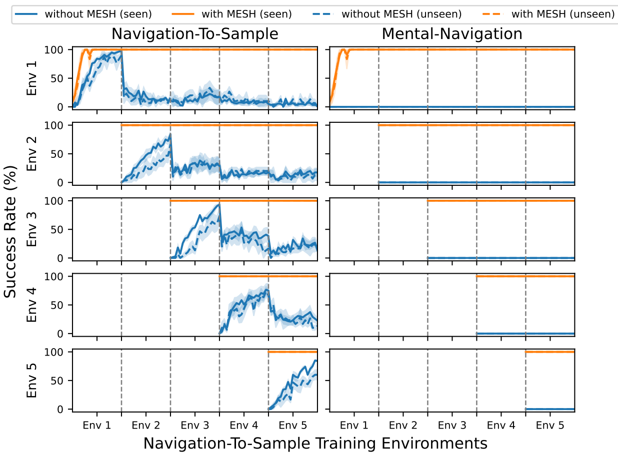

Figure 2: Success rate of each environment while sequentially training five different environments. The lines and shades denote average and standard error, respectively.

grid code is made via this learning mechanism during NTS. Given the grid codes of both current and target images, the CTRNN predicts the actions; move left/right or stop.

In MNAV, MESH first associates both the start and the target images with corresponding grid codes. Upon taking action, MESH then infers the subsequent code for every current image via path integration without referring to the visual stimuli. The output of MESH drives the action network and the action is fed into the grid cell layer as a velocity input. Consequently, the proposed model can retrieve the correct grid representation for each image during the mental navigation task. Figure 1 shows the entire architecture of the model.

Experiments

We sequentially trained the model in five different environments only in navigation-to-sample (NTS) using ground-truth actions. Each environment has six different images with the same size of intervals similar to Neupane et al. (2022). Among all possible pairs in each environment, 80% pairs are used in training (seen pairs) and the others are only used in testing (unseen pairs). The model was tested on both NTS and mental navigation (MNAV) using all pairs. We employed the success rate to evaluate performance in all conditions: both navigation with seen or unseen start-target pairs. All experiments were conducted five times with different random seeds.

Figure 2 shows success rate changes in each environment for NTS and MNAV conditions while sequentially training on each environment in the NTS manner. We evaluated the previous environments during training in a new environment to show catastrophic forgetting. MESH enables overcoming mental navigation while CTRNN without MESH fails to do it. Moreover, it makes the agent more robust to catastrophic forgetting and learning faster. This is because MESH can encode all observations into a much simpler space, grid space. We also analyzed the hidden states in CTRNN using principal component analysis (PCA) for the trajectories in Environment 1 in Figure 3. The states are clearly separated in the

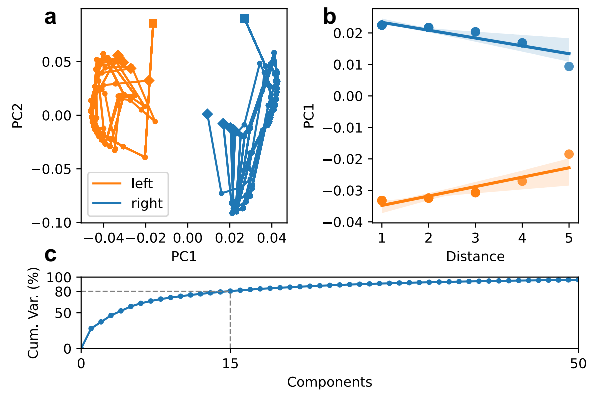

Figure 3: (a) PCA over hidden states (256-dimensional vector) of CTRNN with MESH in environment 1 The diamond and the denote the start and target location of each trajectory, respectively. (b) PC1 of the first hidden states on each trajectory plotted against the distance between the start and target images. (c) Cumulative variances (%) over the first 50 principal components. A small number of components can explain most of the variance.

relative direction of the target image from the start image (Figure 3a) and the hidden state for the first observation in each trajectory is linearly aligned with the distance for each direction. Figure 3c demonstrates that 15 components out of the 256 components explain 80% of the variance, indicating a low-dimensional representation.

Discussion

We modeled how monkeys solve mental navigation tasks (Neupane et al., 2022) by building multi-brain region neural network models with a hippocampal-entorhinal scaffold network, MESH (Sharma et al., 2022). We sequentially trained the model on five different environments and tested how quickly it could adapt to new environments and whether it could overcome catastrophic forgetting. The model could solve the purely mental version of NTS after training only on NTS, by using path integration to reconstruct the next grid state without referring to visual stimuli. It could also overcome catastrophic forgetting and learned new environments instantly (one-shot) by projecting observations into its structured embedding space based on the grid cell code. The hidden states of the model reveal that they are clearly separated by the direction from start to target but only slightly disentangled by the distance between the two. This model can suggest novel hypotheses about neural representations and behavior in monkeys trained on multiple images and on mental navigation.

Acknowledgments

SN is supported by NSERC PDF-516867-2018, FRQNT B3X-258512-2018. MJ is supported by NIH (NIMH-MH129046), Paul and Lilah Newton Brain Science Award, and the McGovern Institute. IF is supported by the Office of Naval Research, and the Howard Hughes Medical Institute (HHMI), and NIH (NIMH-MH129046).

References

Behrens, T. E., Muller, T. H., Whittington, J. C., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509.

Neupane, S., Fiete, I. R., & Jazayeri, M. (2022). Vector production via mental navigation in the entorhinal cortex. *bioRxiv*, 2022–12.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*.

Sharma, S., Chandra, S., & Fiete, I. (2022). Content addressable memory without catastrophic forgetting by heteroassociation with a fixed scaffold. In *Icml*.