Spatially- and non-spatially tuned hippocampal neurons exhibit distinct linear and non-linear representations

Maxime Daigle (maxime.daigle@mail.mcgill.ca)
Department of Physiology, McGill University, Canada

Benjamin W. Corrigan (bcorrig2@uwo.ca)
Robarts Research Institute, Western University, Canada

Julio C. Martinez-Trujillo (julio.martinez@robarts.ca)
Robarts Research Institute, Western University, Canada

Pouya Bashivan (pouya.bashivan@mcgill.ca)
Department of Physiology, McGill University, Canada

Abstract

The hippocampus is known to flexibly represents spatial and non-spatial features of the environment in a task-dependent manner. However, the underlying neural mechanisms governing this contextual adaptability remain elusive. To investigate this, we trained artificial neural networks (ANNs) to perform a navigation-dependent associative memory task mirroring the one performed by macaque monkeys. Using the unit activities of these models, we constructed predictive models of macaque monkey CA3 neurons and measured which types of model better capture the neural computation within the hippocampus. Our results reveal that spatially-tuned neurons predominantly code linear feature combinations, while non-spatially tuned neurons are better explained by non-linear spatiotemporal feature combinations. Moreover, we show that an ANN trained for a navigation-dependent associative memory task learns non-linear spatiotemporal representations that are substantially more aligned with those in the hippocampus compared to alternative models. Altogether, our results shed light on the nature of selectivity across multiple feature dimensions by revealing that the linear and non-linear mixing of features by distinct hippocampal neurons matches surprisingly well with their tendency to be spatially-tuned or not.

Keywords: Hippocampus; Associative memory; Navigation

Introduction

The hippocampus is known for its role in memory and spatial navigation (Eichenbaum & Cohen, 2014). Recent studies have promisingly shown that a range of its specialized spatial cells—such as place, and border cells—can emerge in ANNs (Whittington et al., 2020). However, how spatial and non-spatial information interact within the hippocampus remains a subject of debate. To this end, we constructed predictive models of hippocampal neurons and investigated which model better captures the neural computation governing those dynamics.

Methods

Neuronal Dataset. Gulli et al. (2020) recorded the activity of single hippocampal neurons in macaque monkeys' area CA3 while they performed a navigation-dependent associative memory task (Fig 1). The task required monkeys to learn an association between goal-objects appearing at the end of a maze and a context cue determining the reward values associated with those objects. The animals aims to navigate to the object offering the highest reward.

Neural network models. We tested neural networks employing two distinct memory mechanisms. Both models were trained using reinforcement learning, Actor-Critic, within a virtual environment replicating the associative memory task performed by the macaque monkeys (Fig. 1 top). They also both
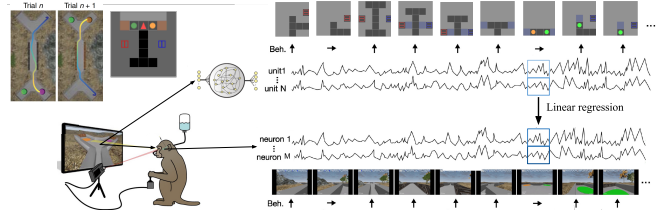

Figure 1: Associative memory task and neural predictivity analysis

received egocentric visual input through a convolutional neural network (CNN). The first model, Episodic Planning Network (EPN), was proposed to rapidly adapt to novel environments during navigational and spatial memory tasks (Ritter et al., 2020). To achieve this, EPN utilizes an episodic memory via iterative application of a multi-head attention layer. The second model, Reward-optimized artificial recurrent network (ROARN), includes a LSTM and utilizes its memory through recurrent connections. Both models achieve an optimal performance on the task

Neural predictivity. To predict hippocampal neural activity, we employed a previously used approach for comparing ANN activations to the neural activity of the visual (Yamins et al., 2014), motor (Sussillo et al., 2015), and auditory cortices (Kell et al., 2018) (Fig. 1 right). We subjected the artificial agent to the same trial sequences experienced by the animal subjects. We then performed a linear regression analysis using a linear SVM (Support Vector Machine) to predict the neurons' firing rates from the model's unit activations. The neural predictivity score was obtained by computing the Pearson correlation, ρ, between predicted and actual firing rates. The neural predictivity gain of a model is the neural predictivity added upon the neural prediticivity obtained from an untrained ROARN as baseline.

Results

Task-specific spatial tuning emerges in both hippocampus and ANN. We found that $52 \pm 1\%$ of hippocampal neurons exhibit at least one spatial response field (Fig 2a). Quantifying the coincidence of spatial response fields in neurons with a minimum of two spatial response fields reveals a tendency for these neurons to exhibit significantly elevated firing rates at multiple locations that hold functional equivalence within the task structure, manifesting as symmetrical tunings for this task (Fig 2b). This task-dependant spatial distribution, previously observed by Gulli et al. (2020), also emerges naturally within ANN models, without any constraint or module designed to replicate this phenomenon (Fig 2b,c).

While $100 \pm 0\%$ of units with random weights exhibit at least one spatial response fields, only $12 \pm 0\%$ of those untrained units are specifically selective for locations holding task-equivalent functions (Fig 2a). In contrast, $56 \pm 2\%$ of spatially-tuned neurons are tuned to those task-equivalent locations. Optimizing ROARN to perform the associative mem-

ory task more than doubles the percentage of spatially-tuned neurons that are tuned at task-equivalent locations, increasing from $12 \pm 0\%$ of spatially-tuned untrained units to $27 \pm 4\%$ of spatially-tuned units (Fig 2a). Moreover, learning the task flattens the distribution of spatial response fields per unit and breaks the complete absence of non-spatially tuned untrained units ($0 \pm 0\%$) by converting $30 \pm 1\%$ of these units into non-spatially tuned units (Fig 2d). In short, while learning the navigation-dependent associative memory task, the units naturally remapped into task-specific spatial tunings similar to those observed in hippocampal neurons (Fig 2d).
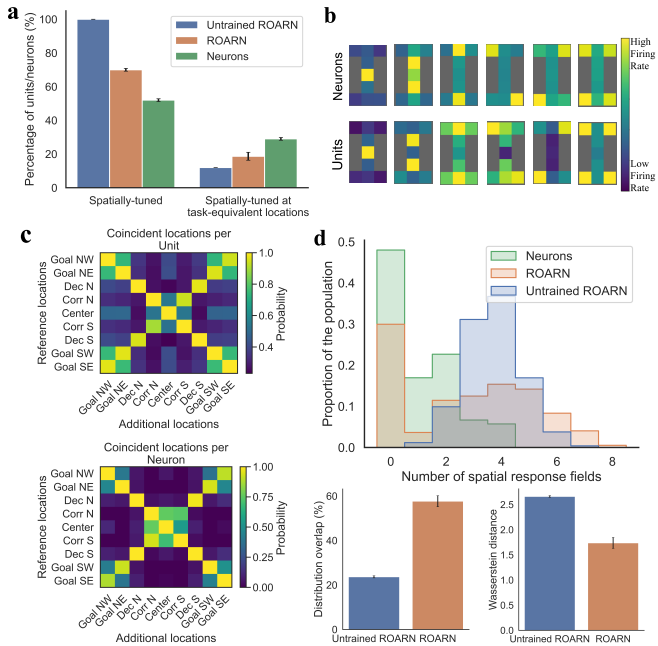

Figure 2: Spatial tuning. a) Percentage of units/neurons with at least one spatial responses fields or with spatial responses fields in locations that are task-wise equivalent. b) Examples of six neurons and units spatially-tuned at task-equivalent locations across the maze. c) Locations of coincident spatial response fields. d) Number of spatial response fields per unit/neuron. Similarity and distance between the spatial response field count distribution in models and neurons. N, north; S, south; W, west; E, east. Mean $\pm$ Std over 3 random seeds.

Neurons tendency to encode linearly or non-linearly is related to their spatial tunings. The hippocampus serves as a hub for integrating a variety of highly processed information, encompassing spatial, temporal, visual, olfactory, and auditory stimuli (Eichenbaum, 2017; Itskov et al., 2012; Save et al., 2000) and these highly refined representations are thought to be key to the hippocampus representations. To better capture this role of hippocampus, we propose an ideal observer model that incorporates all relevant variables, including location, direction, visual cues, decision, and reward. With this model, we predicted the neural responses of each individual neuron and measured the contribution of each type of feature

through its corresponding fitted weight. While non-spatially tuned neurons rely significantly more on non-spatial features (Fig 3a non-spatial features, $P < 1 \times 10^{-7}$, Wilcoxon rank-sum test), spatially-tuned neurons exhibit strong encoding of task phases (Fig 3a task phases, $P < 1 \times 10^{-35}$, Wilcoxon rank-sum test) which are partly associated with spatial locations. Interestingly, we observed that spatial features exhibit similar weights, with slight higher average for non-spatially-tuned neurons (Fig 3a spatial features, $P = 0.011$, Wilcoxon rank-sum test).
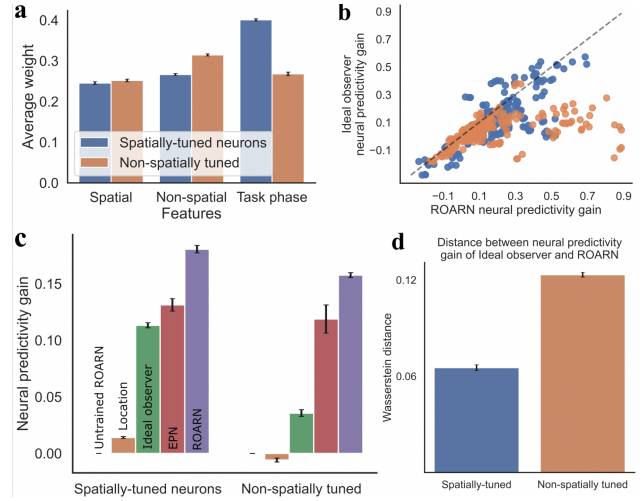

Figure 3: Neural predictivity. a) Average weights for linear regression from Ideal observer. b) Neural predictivity gain per model grouped by spatially- and non-spatially tuned neurons. c) Average neural predictivity gain. d) Distance between neural predictivity gain of ideal observer and ROARN. Mean $\pm$ Std over 3 random seeds.

The ideal observer consists of linear combinations of task-relevant features, where as ROARN receives essentially the same information, but employs multiple non-linear layers. Surprisingly, linear combinations of task-relevant features predicts responses of spatially-tuned neurons similarly to ROARN (Fig 3b spatially-tuned neurons, $\rho = 0.76 \pm 0.01, P < 1 \times 10^{-16}$), while comparatively performing poorly with non-spatially tuned neurons (Fig 3b non-spatially tuned neurons, $\rho = 0.43 \pm 0.02, P < 1 \times 10^{-3}$). Consequently, the gap between ROARN and ideal observer widens considerably for non-spatially tuned neurons, indicating a greater reliance on non-linear computations for those neurons (Fig 3c,d). In summary, we observed that neurons with spatial tuning are often explained well by linear combinations of task-relevant features, whereas neurons without such tuning often encode non-linear mixtures of factors, which can still include spatial information. We hypothesize that this dichotomy may stem from established spatial tuning being obscured by non-linear computations, thereby promoting the two distinct profiles: spatially-tuned neurons characterized by linear representations and non-spatially tuned neurons characterized by non-linear representations.

Acknowledgments

References

Eichenbaum, H. (2017). On the integration of space, time, and memory. *Neuron*, *95*(5), 1007–1018.

Eichenbaum, H., & Cohen, N. J. (2014). Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, *83*(4), 764–770.

Gulli, R. A., Duong, L. R., Corrigan, B. W., Doucet, G., Williams, S., Fusi, S., & Martinez-Trujillo, J. C. (2020). Context-dependent representations of objects and space in the primate hippocampus during virtual navigation. *Nature neuroscience*, *23*(1), 103–112.

Itskov, P. M., Vinnik, E., Honey, C., Schnupp, J., & Diamond, M. E. (2012). Sound sensitivity of neurons in rat hippocampus during performance of a sound-guided task. *Journal of neurophysiology*, *107*(7), 1822–1834.

Kell, A. J., Yamins, D. L., Shook, E. N., Norman-Haignere, S. V., & McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, *98*(3), 630–644.

Ritter, S., Faulkner, R., Sartran, L., Santoro, A., Botvinick, M., & Raposo, D. (2020). Rapid task-solving in novel environments. *International Conference on Learning Representations*.

Save, E., Nerad, L., & Poucet, B. (2000). Contribution of multiple sensory information to place field stability in hippocampal place cells. *Hippocampus*, *10*(1), 64–76.

Sussillo, D., Churchland, M. M., Kaufman, M. T., & Shenoy, K. V. (2015). A neural network that finds a naturalistic solution for the production of muscle activity. *Nature neuroscience*, *18*(7), 1025–1033.

Whittington, J. C., Muller, T. H., Mark, S., Chen, G., Barry, C., Burgess, N., & Behrens, T. E. (2020). The tolman-eichenbaum machine: unifying space and relational memory through generalization in the hippocampal formation. *Cell*, *183*(5), 1249–1263.

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the national academy of sciences*, *111*(23), 8619–8624.