Single neurons in human hippocampus and amygdala track the depth-of-processing elicited by visual representations of images

Aalap D Shah (aalap.shah@yale.edu)
Yale University
New Haven, CT 06520

Richard Xue (richard.xue@yale.edu)
Yale University
New Haven, CT 06520

Qi Lin (qi.lin@riken.jp)
RIKEN
Wako, 351-0106, Japan

Runnan Cao (r.cao@wustl.edu)
Washington University in St. Louis
St. Louis, MO 63130

Shuo Wang (shuowang@wustl.edu)
Washington University in St. Louis
St. Louis, MO 63130

Ilker Yildirim (ilker.yildirim@yale.edu)
Yale University
New Haven, CT 06520

Abstract

The spontaneous processing of visual information plays a significant role in shaping memory, sometimes even overshadowing voluntary efforts to encode specific details. What are the neurocomputational mechanisms that underlie the transformation of percepts to memories in the brain? To address this, we analyzed single neuron recordings in hippocampus and amygdala, two important structures in the medial temporal lobe (MTL), collected while human participants viewed sequences of object images. We hypothesize that the activity of single neurons in these MTL structures track the depth-of-processing of incoming visual information, thereby supporting the perception to memory interface, with more deeply processed images leading to stronger memory traces. Inspired by recent work, we derived a computational signature for the depth-of-processing of visual representations based on the iterative reconstruction loop in a sparse coding model. Consistent with our hypothesis, we found that the firing rates in both hippocampus and amygdala correlate with the number of iterations required for reconstruction — and do so in complementary ways. Moreover, single neurons that are more strongly associated with the number of model iterations also fire more. Our results provide an algorithmic account for how MTL might support the adaptive interface between perception and memory.

Keywords: depth-of-processing; sparse coding; hippocampus; amygdala

Introduction

Although memory can result from intentional selection, much of what we remember simply follows from the spontaneous processing of incoming visual inputs (Bainbridge, 2020; Broers, Potter, & Nieuwenstein, 2018; Isola, Xiao, Parikh, Torralba, & Oliva, 2013; Goetschalckx, Moors, & Wagemans, 2018). Extensive neuroscience research has pointed to the Medial Temporal Lobe (MTL) as the locus interfacing perception and memory. The hippocampus, for instance, not only sits at the top of the visual hierarchy according to an influential characterization of the visual system (Felleman & Van Essen, 1991), but has also been implicated in the formation of episodic memories from visual inputs (Scoville & Milner, 1957; Cao et al., 2024). Critically, not all visual inputs are equally memorable and MTL modulation reflects this non-uniformity (Bainbridge, Dilks, & Oliva, 2017). Yet, the neurocomputational principles of how these brain structures might implement such an adaptive interface, modulating the strength of memories as individual images are encountered, remain unknown.

Here, we hypothesize that the MTL supports the perception-to-memory interface by modulating, in an online fashion, the 'depth-of-processing' of incoming inputs (Craik & Lockhart, 1972). This influential depth-of-processing theory states that the strength of memory traces are dictated by the depth of perceptual processing that incoming visual inputs elicit. In a recent quantitative realization of this theory, Lin, Li, Lafferty, and Yildirim (in press) provided a computational signature for modulating the level of perceptual processing on an image basis: compression-based reconstruction error of the visual representations of an image. They implemented this idea by training a sparse coding model to compress and reconstruct activations of a pre-trained deep convolutional neural network (DNN). They found that images with harder-to-reconstruct representations lead to stronger memory traces. Using this computational signature, we set out to test the depth-of-processing theory in the firing rates of single hippocampus and amygdala neurons in humans, two important structures in the MTL.

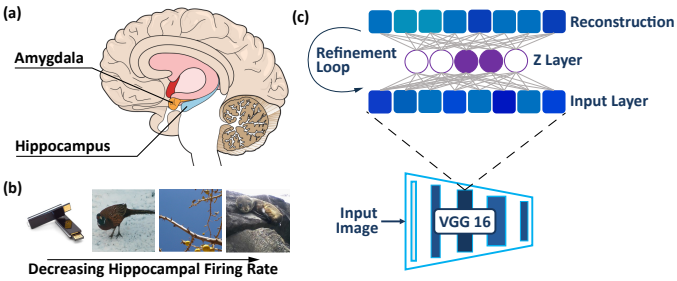

Figure 1: (a) Hippocampus and amygdala indicated on a sketch of the human brain. (b) Four example images from the data collected by Cao et al. (2024) arranged in the decreasing order of the average hippocampus firing rate evoked. (c) A schematic of the sparse coding model performing iterative refinement to more accurately reconstruct visual representations of a layer in a pre-trained DNN.

To do so, we analyze the firing rate of individual cells recorded in the hippocampus and amygdala from human patients viewing 500 sequentially presented images of objects. We test whether the stimulus-driven variability in the spiking rates of these neurons correlate with a computational signature of reconstruction difficulty — the number of iterations needed in the sparse coding model to arrive at convergence-level reconstruction error over the stimulus. Remarkably, we find that the firing rates of both hippocampus and amygdala neurons track the number of iterations until convergence, and they do so in complementary ways relative to both the direction of this effect, and the DNN layer that the visual representations originate in. These results provide support for the hippocampus and amygdala implementing the core interface of perception and memory via the adaptive depth-of-processing mechanism.

Methods

WUSTL dataset: single neurons in human hippocampus and amygdala The WUSTL dataset (Cao et al., 2024) contained activations of 808 single neurons in the human hippocampus ($n = 362$) and amygdala ($n = 446$). This data was collected from 15 human participants viewing a total of 500 natural object images, uniformly distributed across 50 categories from ImageNet (Deng et al., 2009). The stimuli were presented for 1000 ms (with an inter-stimulus-interval of 500 to 750 ms) and neural activity was recorded from 250 to 1250 ms after stimulus onset. Following (Cao et al., 2024), we an-

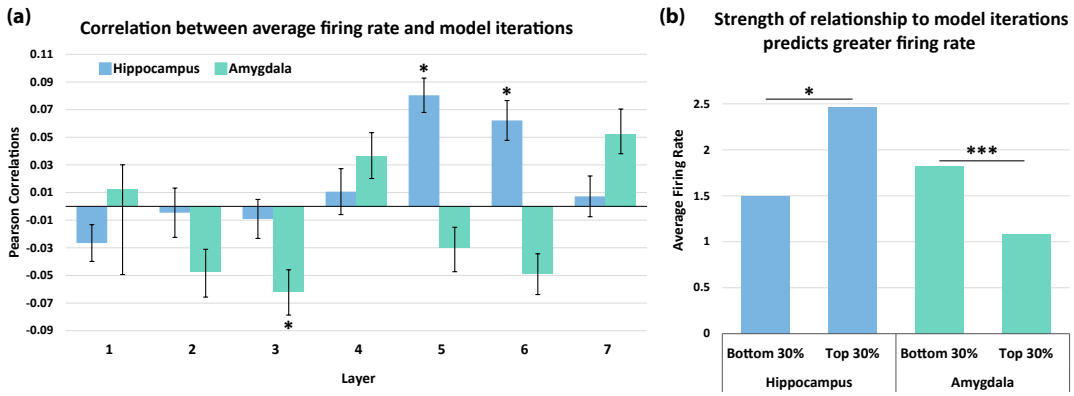

Figure 2: (a) Bootstrapped correlations between the iteration counts of the sparse coding model (each trained on a different layer of the DNN, from layer 1 to 7) and the firing rates of neurons in the human hippocampus and amygdala. While the iteration counts based on later layers (layers 5, 6) correlate positively with hippocampus, the iterations based on an earlier layer (layer 3) correlate negatively with amygdala. (b) The average firing rates of the top and bottom 30% neurons with respect to the strength of the relationship between the individual neurons' firing rates across images and the best-match sparse coding layer. Error bars indicate standard deviations; "*" for $p < .05$; "***" for $p < .001$

alyzed "background-subtracted" firing activity in each neuron — 1.5 standard deviations above the mean neural response in the -500 to 0 ms window preceding the stimulus onset.

A computational signature of depth-of-processing: Iterations to reconstruct in sparse coding Our computational model builds on the work of Lin et al. (in press) who proposed that compression-based reconstruction error — implemented using a sparse coding model (Olshausen & Field, 1996) — modulates depth-of-processing (Craik & Lockhart, 1972) during the spontaneous processing of visual inputs, thereby impacting their associated memory strengths. The sparse coding model is trained to reconstruct the activations evoked by a DNN (VGG-16 (Simonyan & Zisserman, 2014)) pre-trained on ImageNet (Deng et al., 2009). Given an input vector of DNN activations, sparse coding involves an iterative optimization process to reconstruct the input as accurately as possible from its 'codewords', a learned basis set of vectors spanning the input space, until it approaches a stable solution. Convergence is achieved when either a threshold reconstruction error (of 0.01) is achieved or the maximum allowable iteration is reached (1000 iterations). We therefore use the number of iterations to reach convergence-level reconstruction as our computational signature for depth-of-processing: Just as the larger iterations in the sparse coding model indicate an increased computational demand for accurately reconstructing inputs, we suggest that in the brain, deeper processing implies more firing at the neuron level (and likely more cognitive resources overall) incurred during visual processing.

Model vs. data comparisons We followed Lin et al. (in press) to train a separate sparse coding model for each of the 5 maxpool and 2 dense layers of a pretrained VGG-16. 1000 randomly sampled units from these layers served as the input to the corresponding model instance. All model instances had an intermediate z-layer (Fig. 1c) of size 500 and an output reconstruction layer of 1000 units. For each of the 500 images in the WUSTL dataset, we ran it through the 7 model instances and record its respective iteration numbers. During analysis, for each model instance, we correlated the 500-dimensional model prediction vector with the 500-dimensional average firing rates of the hippocampus and amygdala neurons.

Results

We find that images that require more iterations to reconstruct lead to more intense firing in human hippocampus neurons (Fig. 2a). The sparse coding model based on layers 5 and 6 correlate significantly positively with the average hippocampus firing rates ($r = .08$ and $r = .06$, both $p < .05$). Moreover, the sparse coding model based on layer 3 correlates significantly negatively with the amygdala firing rates ($r = -.06$, $p < .05$). This suggests a complementary impact of depth-of-processing across hippocampus and amygdala, both in inputs and the direction of the relationship.

We also find that the activity of individual neurons are modulated iterations-to-reconstruct (Fig. 2b). Based on the analysis above, we selected the model instances that best correlated with hippocampus (with layer 5) and amygdala (with layer 3). Then for each brain region, we identified two distinct clusters of neurons — the top 30% and bottom 30% of the neurons in their correlations with the corresponding model instance. In hippocampus, we find that the top neurons elicit significantly higher firing rate than the bottom neurons. And, we observe the opposite trend with the amygdala neurons. These results establish that the strength of the relationship to model iterations predicts firing rates in single neurons (higher for hippocampus and lower for amygdala).

Discussion

Taken together, these results suggest an algorithm-level mechanism for how the MTL neurons might support the perception to memory interface, based on the implementation of the Craik & Lockart's depth-of-processing theory via the iterative reconstruction process in sparse coding.

References

Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia*, *141*, 107408.

Bainbridge, W. A., Dilks, D. D., & Oliva, A. (2017). Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage*, *149*, 141–152.

Broers, N., Potter, M. C., & Nieuwenstein, M. R. (2018). Enhanced recognition of memorable pictures in ultra-fast rsvp. *Psychonomic bulletin & review*, *25*, 1080–1086.

Cao, R., Brunner, P., Chakravarthula, P. N., Wahlstrom, K., Inman, C., Li, X., ... others (2024). A neuronal code for object representation and memory in the human amygdala and hippocampus.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671–684.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 ieee conference on computer vision and pattern recognition* (pp. 248–255).

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, *1*(1), 1–47.

Goetschalckx, L., Moors, P., & Wagemans, J. (2018). Image memorability across longer time intervals. *Memory*, *26*(5), 581–588.

Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2013). What makes a photograph memorable? *IEEE transactions on pattern analysis and machine intelligence*, *36*(7), 1469–1482.

Lin, Q., Li, Z., Lafferty, J., & Yildirim, I. (in press). Images with harder-to-reconstruct visual representations leave stronger memory traces. *Nature Human Behaviour*.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, *20*(1), 11.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.