

Formalizing a Perceptual-Mnemonic Theory of the Medial Temporal Lobe

Tyler Bonnen (bonnen@stanford.edu)

Department of Psychology, Stanford University
Stanford CA, 94305, USA

Daniel L. K. Yamins (yamins@stanford.edu)

Department of Psychology and Computer Science, Wu Tsai Neurosciences Institute, Stanford University
Stanford CA, 94305, USA

Anthony D. Wagner (awagner@stanford.edu)

Department of Psychology, Wu Tsai Neurosciences Institute, Stanford University
Stanford CA, 94305, USA

Abstract:

Animals rapidly transform sensory experience into memory. How the mammalian brain supports these transformations has been subject to an enduring debate: Do medial temporal lobe (MTL) structures, typically implicated in memory-related behaviors, also play a role in perception? A rich experimental literature exists, yet reliance on descriptive accounts of stimulus properties (e.g. “feature ambiguity”) has made it difficult to synthesize results. In order to formalize perceptual demands on the MTL, in particular, the role of perirhinal cortex (PRC), here we adopt a combination of meta-analytic and computational approaches. We begin by designing a null model of PRC function in visual discrimination tasks, building from a computational proxy for the primate ventral visual system (VVS). With this model, we identify stimuli from previous studies that may not be diagnostic of PRC’s role in perception. We then demonstrate a striking correspondence between model and PRC-lesioned behavior across ten experiments ($r=.80$). Critically, the model and PRC-lesioned subjects fail on the same visual discrimination tasks, unlike controls. This approach formalizes the MTL’s role in perception by providing a tractable, stimulus-computable proxy for visual discrimination tasks in a PRC-lesioned state.

Keywords: perirhinal cortex; perceptual-mnemonic; MTL

Introduction

Animals rapidly transform perceptual experience into memory, allowing this information to guide future behavior. Here, we focus on how sensory information from the ventral visual system (VVS) is transformed within the medial temporal lobe (MTL). Traditionally, the MTL has been characterized as a system dedicated to supporting memory-related behaviors (Squire & Wixted 2011). In contrast, the Perceptual-Mnemonic Theory (PMT) argues that the MTL is also necessary for certain perceptual tasks (Bussy & Saksida 2007). This theory has centered on the role of perirhinal cortex (PRC), a MTL structure situated at the apex of the VVS (Suzuki & Naya 2014). Accordingly, PRC is thought to support “configural visual representations”, enabling subjects to perform “complex” visual discrimination tasks that rely on configural properties of objects (Barense et al. 2007). In line with PMT, patients with perirhinal lesions have shown deficits on tasks

designed to assess PRC’s role in perception (Lee et al. 2005, Barense et al. 2007). By contrast, in other work, PRC lesioned patients are unimpaired on similar visual discrimination tasks (Buffalo et al. 1998, Levi et al. 2005, Knutson et al. 2013), leading some to argue that observed deficits reflect memory impairment. Unfortunately, reliance on informal, descriptive accounts of stimulus properties (e.g. visual “complexity,” “feature ambiguity,” or “high-level” perceptual demands) have made it difficult to compare results across experiments.

We begin by formalizing a null model of PRC function in accordance with PMT. As subjects with focal lesions to PRC still have an intact VVS, we suggest that visual discrimination behaviors of lesioned subjects should reflect those discriminations that are supported by the VVS. In particular, we should expect PRC-lesioned behavior to rely on Inferior Temporal cortex (IT) when performing object discrimination tasks. If a visual discrimination task is “IT computable,” we suggest that the task is nondiagnostic of PRC’s role in perception, as no perceptual processing beyond the VVS is necessary. Moreover, if PRC plays a role in perception, then PRC-intact control subjects should be able to perform non-IT computable visual discriminations; PRC-related deficits would only be evident for these tasks.

Methods

To implement this formalization of PMT, we borrow from a model class that makes quantitative predictions about neural activity within the primate visual system: task-optimized convolutional neural networks (CNNs). Given an image as input, these models largely recover the response patterns observed across the VVS (Yamins et al. 2014, Schrimpf et al. 2018). For simplicity, we refer here to a single instance of this model class optimized to perform object recognition (VGG16: Simonyan & Zisserman 2014). As primate visual behaviors have been shown to track a linear readout of population-level neural representations in IT (Majaj et al. 2015) we use linear separability as our metric of IT computability.



We obtained stimulus sets from previously published studies (Buffalo et al. 1998, Barense et al. 2007, Stark & Squire 2000, Knutson et al. 2012) where human subjects with perirhinal lesions and PRC-intact controls performed visual discrimination tasks. Tasks involved detecting the “oddity” with a choice array of relatively similar objects. Stimuli from each experiment were originally formatted as concurrently presented choice arrays, with between 3 and 8 objects on each screen. When appropriate, we used a kmeans clustering approach to identify the centroid of each object, otherwise the stimuli were split into quadrants of uniform size. Each stimulus screen was segmented into N object-centered images of equal size. We formatted each experiment such that the inputs to the model reflect the stimuli shown to subjects--the visual statistics of each image, as well as the distribution of objects. For each trial, we passed these N trial images to the model and extracted feature vectors of length F from an intermediate layer, resulting in an $F \times N$ response matrix.

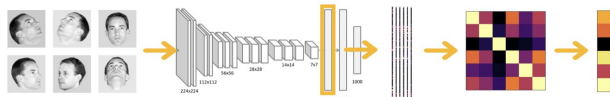


Figure 1. Computational proxy for PRC-lesioned behavior: N objects in each trial are segmented and passed to a Convolutional Neural Network; responses are extracted from an intermediate layer, yielding an $F \times N$ array; pairwise correlations between objects result in an $N \times N$ covariance matrix; the lowest mean off-diagonal correlation is that trial’s predicted oddity; results are averaged across trials.

We defined a protocol (Figure 1) to determine the oddity from each trial’s $F \times N$ array, extracted from the first fully connected layer. To operationalize linear separability within the model, we first used Pearson’s correlation as a uniform linear readout, computing the covariance matrix between model responses to each item within a trial, resulting in an $N \times N$ array. We identified the item with the lowest mean off-diagonal covariance as the oddity. If the resulting model-selected oddity matched the ground truth, then that trial was classified as correct. This was repeated for each trial, resulting in an average accuracy for each experiment.

Results

First, this analytic approach identified published stimulus sets that are nondiagnostic of PRC’s involvement in perception: Stimuli are perfectly separable in the model (e.g. model accuracy of 100%) and, by proxy, should not require visual processing beyond the VVS. This includes stimuli from studies on both sides of the debate (Knutson et al 2012, Experiment 2 in Barense et al. 2007). PRC-lesion deficits (or lack thereof) on these tasks may be for reasons unrelated to the perceptual demands placed on PRC.

Second, we observed a striking correspondence between PRC-lesioned patient and model performance ($r=.80$). We analyzed ten visual discrimination experiments from two studies, where each study has experiments with putatively PRC-relevant and -irrelevant stimuli. After computing the model accuracy for each experiment, we standardized

results into a common metric space, which is agnostic to stimuli used, the number of items in the choice array, the number of trials, etc., across all experiments, we computed the correlation between model performance and PRC-lesioned patients (Figure 2).

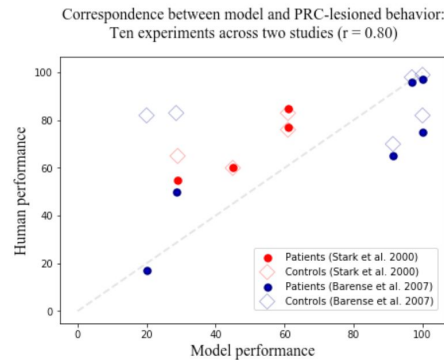


Figure 2. Correspondence between model and PRC-lesioned behavior. Solid dots along the diagonal suggest the model predicts the behavior of PRC-lesioned patients across experiments. While patient and control behaviors are well predicted by the model in Stark et al. 2000, the model poorly predicts control behavior in Barense et al. 2007. This lack of correspondence is driven by “high feature ambiguity” conditions that require putatively PRC-dependent representations.

Finally, we note a qualitative divergence between model and control behaviors within these two studies. In Stark et al. (2000) model performance was well matched with control as well as lesioned behaviors ($r = .77$ and $.93$, respectively). For Barense et al. (2007), the model was far worse at predicting control relative to lesioned behavior ($r=.28$ and $.89$, respectively). The lack of correspondence was driven by the “high feature ambiguity” conditions, designed to rely on representations putatively unique to PRC. For these experiments, controls significantly outperformed model behavior, as well as lesioned subjects (Figure 2; open blue squares, upper left quadrant). PRC-lesioned subject and control behaviors diverge only when experimental stimulus sets are not IT computable, as determined by our computational approach.

Summary

We have leveraged computational and meta-analytic approaches to formalize the role of the MTL in perception. With this formalization, we suggest that not all existing experimental stimulus sets are diagnostic of the role that PRC may play in perception. When a stimulus set is “IT computable” it should not require perceptual processing beyond the VVS. Consequently, a lack of impairment for PRC-lesioned patients in these tasks would be expected, given that they still have an intact VVS. Additionally, across ten visual experiments, we have demonstrated a correspondence between the model and human subjects in a PRC-lesioned state. Critically, it appears that the behavior of PRC-lesioned and -intact subjects only diverges when a visual discrimination task is not IT computable, as indicated by the model. These results offer tentative support for the PMT, as the debate is currently construed, and suggest that

theoretical disagreements may be due to differing perceptual demands across experiments.

Acknowledgments

This work is supported by a National Science Foundation Graduate Research Fellowship under Grant No. DGE-1656518, the Marcus and Amelia Wallenberg Foundation (MAW2015.0043), and Stanford's Center for Mind Brain Behavior and Technology. We thank the original authors for sharing their stimuli; Morgan Barense, Craig Stark, Elizabeth Murry, Tim Bussey, Mona Hopkins, as well as Jennifer Frascino for helpful correspondence.

References

- Barense, M. D., Gaffan, D., & Graham, K. S. (2007). The human medial temporal lobe processes online representations of complex objects. *Neuropsychologia*, 45(13), 2963-2974.
- Buffalo, E. A., Reber, P. J., & Squire, L. R. (1998). The human perirhinal cortex and recognition memory. *Hippocampus*, 8(4), 330-339.
- Bussey, T. J., Saksida, L. M., & Murray, E. A. (2002). Perirhinal cortex resolves feature ambiguity in complex visual discriminations. *European Journal of Neuroscience*, 15(2), 365-374.
- Knutson, A. R., Hopkins, R. O., & Squire, L. R. (2012). Visual discrimination performance, memory, and medial temporal lobe function. *Proceedings of the National Academy of Sciences*, 109(32), 13106-13111.
- Knutson, A. R., Hopkins, R. O., & Squire, L. R. (2013). A pencil rescues impaired performance on a visual discrimination task in patients with medial temporal lobe lesions. *Learning & Memory*, 20(11), 607-610.
- Lee, A. C., Buckley, M. J., Pegman, S. J., Spiers, H., Scahill, V. L., Gaffan, D., ... & Graham, K. S. (2005). Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus*, 15(6), 782-797.
- Levy, D. A., Shrager, Y., & Squire, L. R. (2005). Intact visual discrimination of complex and feature-ambiguous stimuli in the absence of perirhinal cortex. *Learning & memory*, 12(1), 61-66.
- Majaj, N. J., Hong, H., Solomon, E. A., & DiCarlo, J. J. (2015). Simple learned weighted sums of inferior temporal neuronal firing rates accurately predict human core object recognition performance. *Journal of Neuroscience*, 35(39), 13402-13418.
- Murray, E. A., & Bussey, T. J. (1999). Perceptual-mnemonic functions of the perirhinal cortex. *Trends in cognitive sciences*, 3(4), 142-151.
- Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., ... & Yamins, D. L. (2018). Brain-Score: which artificial neural network for object recognition is most brain-like?. *BioRxiv*, 407007.
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Suzuki, W. A., & Naya, Y. (2014). The perirhinal cortex. *Annual Review of Neuroscience*, 37, 39-53.
- Stark, C. E., & Squire, L. R. (2000). Intact visual perceptual discrimination in humans in the absence of perirhinal cortex. *Learning & Memory*, 7(5), 273-278.
- Squire, L. R., & Zola-Morgan, J. (1991). The cognitive neuroscience of human memory since HM. *Annual review of neuroscience*, 34, 259-288.
- 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.