Representational Overlap Triggers Reorganization of Memories

Anisha Babu (ababu@uoregon.edu)
Psychology Department, 1451Onyx Street
Eugene, OR 97403 USA

Zhifang Ye (zhifangy@uoregon.edu)
Psychology Department, 1451Onyx Street
Eugene, OR 97403 USA

Brice Kuhl (bkuhl@uoregon.edu)
Psychology Department, 1451Onyx Street
Eugene, OR 97403 USA

Abstract:

Remembering events from the past requires discriminating between similar memories. Theoretical and empirical work argues that when neural representations of memories overlap, this triggers adaptive changes that improve discriminability. Here, we tested this idea using (1) fMRI to measure initial representational overlap among memories for naturalistic scene images and (2) Natural Language Processing algorithms to quantify the structure of verbal recall. Across six runs of fMRI scanning, N=21 participants learned to discriminate 18 scene images (three categories * six similar exemplars). After scanning, participants verbally recalled each scene. Within the parahippocampal place area (PPA), we assessed the representational structure of the 18 images and the relationship of this structure to verbal recall. We found that PPA robustly reflected category-level information, which was preserved in verbal recall. Within categories, however, PPA representational structure of individual exemplars was negatively correlated with the structure of verbal recall. These results suggest that neural overlap triggered an adaptive reorganization that improved discriminability of recalled memories.

Keywords: fMRI; memory; Natural Language Processing; naturalistic images

Introduction

As we navigate the world, we experience many highly similar and overlapping events (e.g., two visits to the same restaurant). One of the primary challenges for the memory system is to avoid interference between overlapping events (Bakker et al., 2008; Colgin et al., 2008; Yassa & Stark, 2011). Recent theoretical and empirical work suggests that overlap among memories can trigger active mechanisms that reduce this overlap by exaggerating subtle differences between memories (Chanales et al., 2017; Chanales et al., 2021; Drascher & Kuhl, 2022; Hulbert & Norman, 2015; Ritvo et al., 2019; Wanjia et al., 2021; Zhao et al., 2021). However, there remains relatively little evidence directly linking the overlap of neural representations of memories to adaptive changes in how these overlapping memories are recalled (behavior).

In the current work, we sought to directly link neural measures of memory overlap to adaptive changes in verbal recall. We did this using fMRI to measure representational overlap as participants learned overlapping naturalistic scene images and a Natural Language Processing (NLP) algorithm to measure the representational structure of subsequent verbal recall of the scene images. Specifically, we tested whether overlap among scene representations in parahippocampal place area (PPA; Epstein & Kanwisher, 1998) would trigger adaptive changes in verbal recall that exaggerate subtle differences between similar images.

Figure 1: Sample stimuli from pool category

Methods

Task Design

N=21 participants studied 18 naturalistic scene images while being scanned in fMRI. Images were drawn from three categories (e.g., pools, libraries, etc.; Figure 1) with six exemplars each. Participants completed six functional runs during which they learned to associate each scene image with a unique face image via (1) study trials where they viewed a face image followed by a scene image, and (2) vividness trials where they practiced recalling scene images when presented with each face image. After learning these parings inside the scanner, participants completed a verbal recall task outside of the scanner. Here, they were presented with

each face image and asked to type a description of the associated scene image using at least 10 words.

Analysis

To quantify fMRI activity patterns, we applied a general linear model (GLM) to study trials for each image, separately for each functional run. This generated a map of t-statistics for each image and run that served as the "activity pattern."

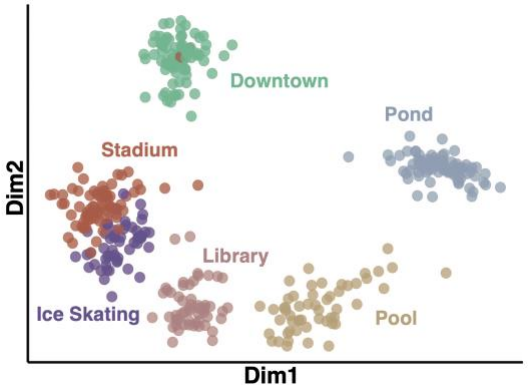

Figure 2: MDS plot of text embeddings of verbal recall data. The six categories are separated by color

To quantify the representational structure of verbal recall descriptions, we used the NLP algorithm MPNet (Masked and Permuted Network; Song et al., 2020) to transform descriptions into 768-dimension text embeddings. We confirmed that these text embeddings are highly sensitive to category-level structure, as demonstrated by clear separation of the 6 categories when embeddings were visualized using multidimensional scaling (MDS; Figure 2).

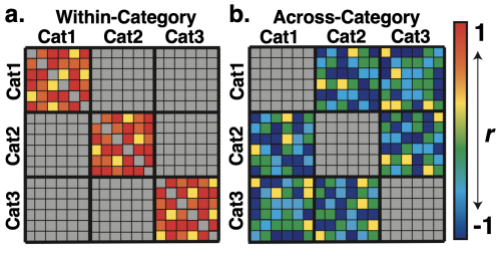

Figure 3: Example RSM

In order to relate the representational structure of verbal recall to the representational structure of fMRI data, we separately constructed representational similarity matrices (RSMs; Kriegeskorte et al., 2008) for the verbal recall data (text embeddings) and for the fMRI data. For the fMRI data, we generated RSMs for each pair of consecutive scan runs (referred to as timepoints 1-5), to account for potential changes cross learning. For each participant, we then correlated the verbal recall RSM with the fMRI RSMs (for each timepoint) to test whether representational structure in PPA predicted (correlated with) the representational structure of verbal recall. Importantly, we separately considered representational structure for pairs of images that were "within category" (e.g., two pools; Figure 3a) and pairs of images that were across-category (e.g., pool – library; Figure 3b). In addition to PPA—which we predicted would be sensitive to high-level scene information—we also considered early visual cortex (EVC) as a control region representing low-level visual information.

Results

Figure 4 shows the similarity between the fMRI (neural) RSMs and the verbal recall RSMs, separated by timepoint during learning and by within- vs. across-category similarity. Within PPA, we found that the fMRI and recall RSMs for across-category similarity were positively correlated across all timepoints (T1: $p=0.027$, T2: $p=0.002$, T3: $p<0.001$, T4: $p=0.032$, T5: $p<0.001$; Figure 4a). The pattern was qualitatively similar in EVC (Figure 4b). These results indicate that differences between exemplars from different categories that were captured in the fMRI-based measures of similarity were preserved in verbal recall.

Next, we considered the critical question of whether neural overlap between exemplars from the same category (i.e., overlap among similar memories) was related to the structure of verbal recall. In PPA, representational structure during early time points in learning was *negatively* correlated with the representational structure of verbal recall (T1: $p=0.046$, T2: $p=0.072$, T3: $p=0.028$, T4: $p=0.073$, T5: $p=0.055$; Figure 4a). For EVC, however, these correlations were all numerically positive (T1: $p=0.238$, T2: $p=0.250$, T3: $p=0.061$, T4: $p=0.039$, T5: $p=0.236$; Figure 4b).

Finally, we directly compared the neural-to-recall correlations within- vs. across categories via a 2-way ANOVA (with factors of timepoints and representational level: within- vs. across-category). This revealed a main effect of representational level in PPA ($p<0.001$) but not in EVC ($p=0.242$). Thus, whereas differences between categories (category-level structure) was preserved from PPA to verbal recall, similarity within category (exemplar-level structure) was inverted from PPA to recall. In other words, greater overlap among highly

similar scenes in PPA predicted *less similar* verbal recall.

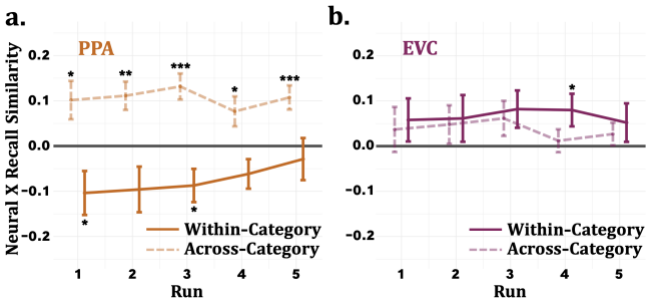

Figure 4: Within- and Across-Category RSM Correlations in PPA and EVC

In summary, we report novel evidence that representational overlap among memories triggers adaptive reorganization in the content of those memories, specifically increasing the discriminability of highly similar memories. Our findings also specifically implicate PPA—which represents high-level scene information (Epstein & Kanwisher, 1998)—in representing features that trigger these changes in memory. More broadly, our findings are consistent with the idea that interference between similar memories is resolved via adaptive reorganization of memory content.

Acknowledgments

#1F31MH135686-01 funding to AB and R01 #NS107727 and R01 #NS089729 funding to BK

References

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern Separation in the Human Hippocampal CA3 and Dentate Gyrus. *Science*, *319*(5870), 1640–1642. https://doi.org/10.1126/science.1152882

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Current Biology*, *27*(15), 2307-2317.e5. https://doi.org/10.1016/j.cub.2017.06.057

Chanales, A. J. H., Tremblay-McGaw, A. G., Drascher, M. L., & Kuhl, B. A. (2021). Adaptive Repulsion of Long-Term Memory Representations Is Triggered by Event Similarity. *Psychological science*, *32*(5), 705–720. https://doi.org/10.1177/0956797620972490

Colgin, L. L., Moser, E. I., & Moser, M.-B. (2008). Understanding memory through hippocampal remapping. *Trends in Neurosciences*, *31*(9), 469–477. https://doi.org/10.1016/j.tins.2008.06.008

Drascher, M. L., & Kuhl, B. A. (2022). Long-term memory interference is resolved via repulsion and precision along diagnostic memory dimensions. *Psychonomic bulletin & review*, *29*(5), 1898–1912. https://doi.org/10.3758/s13423-022-02082-4

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

Hulbert, J. C., & Norman, K. A. (2015). Neural Differentiation Tracks Improved Recall of Competing Memories Following Interleaved Study and Retrieval Practice. *Cerebral cortex (New York, N.Y. : 1991)*, *25*(10), 3994–4008. https://doi.org/10.1093/cercor/bhu284

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends in cognitive sciences*, *23*(9), 726–742. https://doi.org/10.1016/j.tics.2019.06.007

Song, K., Tan, X., Qin, T., Lu, J., & Liu, T. Masked and Permuted Pre-training for Language Understanding. *arXiv, 2004.09297*. https://doi.org/10.48550/arXiv.2004.09297

Wanjia, G., Favila, S. E., Kim, G., Molitor, R. J., & Kuhl, B. A. (2021). Abrupt hippocampal remapping signals resolution of memory interference. *Nature Communications*, *12*(1), 4816. https://doi.org/10.1038/s41467-021-25126-0

Yassa, M. A., & Stark, C. E. (2011). Pattern separation in the hippocampus. *Trends in neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Zhao, Y., Chanales, A. J. H., & Kuhl, B. A. (2021). Adaptive Memory Distortions Are Predicted by Feature Representations in Parietal Cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *41*(13), 3014–3024. https://doi.org/10.1523/JNEUROSCI.2875-20.2021