More than meets the eye: Reconstructing lingering thoughts from visual long-term memories

Futing Zou (futingz@uoregon.edu)
Department of Psychology, University of Oregon,
Eugene, OR, USA

J. Benjamin Hutchinson (bhutch@uoregon.edu) *
Department of Psychology, University of Oregon,
Eugene, OR, USA

Brice A. Kuhl (bkuhl@uoregon.edu) *
Department of Psychology, University of Oregon,
Eugene, OR, USA
Institute of Neuroscience, University of Oregon,
Eugene, OR, USA

* Co-senior author

Abstract

Leading theories of memory propose that our experiences are embedded within slowly drifting representations that capture the passage of time (temporal context). When events from the past are remembered, temporal context representations are thought to also be reinstated. Here, using natural language processing methods and inverted fMRI encoding models, we developed a novel approach to directly measure the reinstatement of temporal context. Specifically, we show that when a previously-encountered stimulus is re-encountered, activity patterns in ventromedial prefrontal cortex reflect the semantic information that immediately preceded its original encounter. That is, re-encountering a stimulus reinstates semantic information that putatively 'lingered in mind' when the stimulus was originally encountered. This constitutes novel evidence of temporal context reinstatement and highlights the influence of past events on ongoing processing.

Keywords: temporal context reinstatement; inverted encoding model; semantic content; episodic memory

Introduction

Central to leading theories and computational models of episodic memory is the idea that individual events are embedded in a broader "temporal context" (Howard & Kahana, 2002; Norman et al., 2001). When a past event is retrieved from memory, this is thought to trigger reinstatement of the event's temporal context (Polyn et al., 2009; Sederberg et al., 2008). As such, remembering an event not only recovers the event itself, but also reinstates information related to other events that were encoded nearby in time. However, methods for directly measuring temporal context reinstatement are currently limited.

One way in which temporal context reinstatement has been measured is through reinstatement of neural activity patterns. This approach is motivated by the idea that temporal context is encoded within distributed, drifting patterns of neural activity. Indeed, converging evidence from human neuroimaging studies has shown that recalling a stimulus not only involves reinstatement of neural activity patterns that were initially engaged during the encoding of that stimulus, but also reinstatement of the activity patterns that preceded the encoding of that stimulus (Folkerts et al., 2018; Manning et al., 2011; Yaffe et al., 2014). While important, these neural activity patterns, on their own, do not indicate the nature of the information that is reinstated.

Here rather than measuring reinstatement of neural activity patterns, we developed and used a novel fMRI-based approach to directly test for reinstatement of *information* that preceded initial encoding of a stimulus. Specifically, we used natural language processing methods to characterize the content of natural scene images and inverted fMRI encoding models to test whether repetition of a stimulus reinstated the semantic content that preceded the stimulus' original encounter. Thus, temporal context reinstatement was operationalized as the degree to which neural patterns evoked by a stimulus' repetition contain information about the semantic content that lingered in mind when the stimulus was originally encountered. This approach allowed us to explicitly test predictions from leading theoretical models.

Methods

To test our hypothesis, we analyzed data from the massive Natural Scenes Dataset (Allen et al., 2022). Eight participants performed a continuous recognition task in which they viewed thousands of scene images that were repeatedly presented across many scan sessions, from which the first two encounters of each stimulus (E1, E2) were used for the present study. Analyses were restricted to E1 and E2 trials that were drawn from the same session (but different scan runs) and were each associated with correct behavioral responses (i.e., E1 = 'new' responses, E2 = 'old' responses) to avoid potential confounds.

We operationalized the temporal context of E1 as the semantic information of the stimulus that immediately preceded E1 (i.e., E1-1). The primary goal of our analyses was to reconstruct the semantic content of stimuli that preceded E1 based on activity patterns evoked during E2. To characterize the semantic content of scene images, we transformed independent annotations (text descriptions) of the scene images into 512-dimensional semantic embeddings using Google's Universal Sentence Encoder (USE) (Cer et al., 2018). We then applied principal component analysis (PCA) on the USE semantic embeddings and used the top 20 PCs as a 20-dimensional representation of the semantic content of each scene image. These PC scores were used for the inverted encoding model described below.

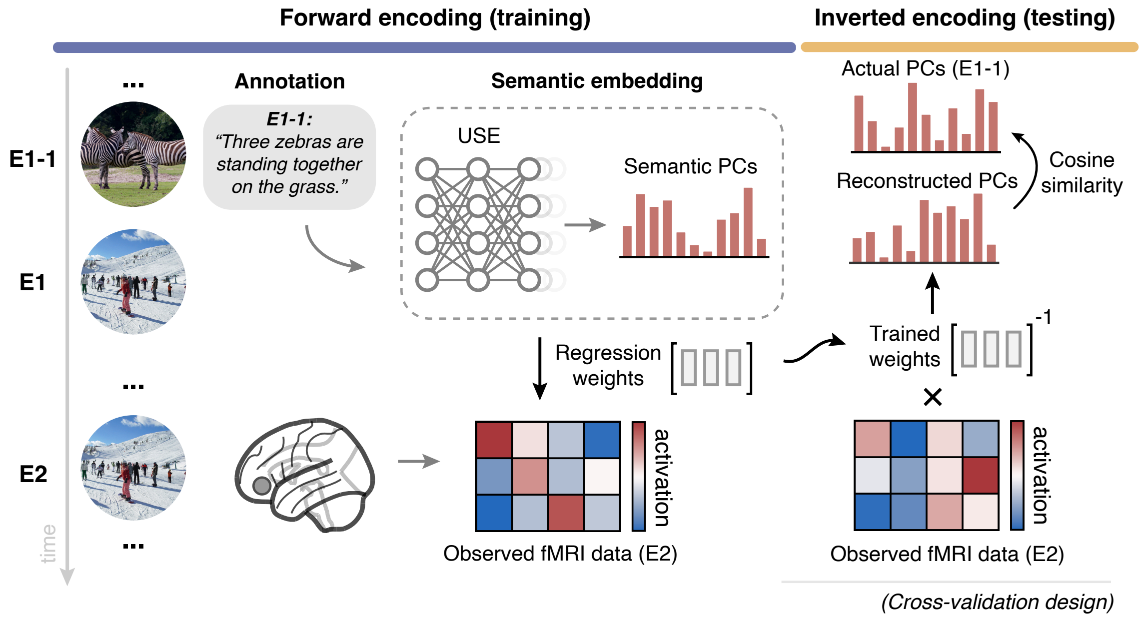

Figure 1: Inverted encoding models for reconstructing semantic components of scenes from fMRI activity patterns.

Reconstructions of semantic components were generated using a cross-validation approach (Figure 1). Specifically, using ridge regression, we first learned a direct linear mapping from the semantic components (of E1-1) to fMRI activity patterns evoked during E2. We then inverted this encoding model and applied it to held-out data (leave-one-session-out) in a cross-validated manner to reconstruct semantic components. Reconstruction accuracy was computed as the cosine similarity between the reconstructed and actual semantic components.

Results

We focused our analyses on three regions of interest (ROIs) that we predicted would be sensitive to the semantic content within scene images: (1) ventromedial prefrontal cortex (vmPFC) which forms high-level schemas (Gilboa & Marlatte, 2017), (2) angular gyrus (AG) which forms event-specific memory representations (Humphreys et al., 2021), and (3) lateral occipitotemporal cortex (LOTC) which is sensitive to current visual content (Konkle & Caramazza, 2013). We first tested whether activity patterns at E2 contained information about the current stimulus (E2). Consistent with previous studies (Lee & Kuhl, 2016; Wang et al., 2023), robust content reconstruction was obtained from all ROIs (*ps* $< .001$; permutation test).

Of critical interest was whether semantic information that preceded E1 (i.e., the semantic embeddings for E1-1) could be reconstructed from activity patterns elicited at E2. Successful reconstruction would indicate that participants reinstated the temporal context of E1 when E2 was encountered. Strikingly, we found that semantic embeddings of E1-1 were successfully reconstructed from E2 activity patterns in vmPFC (Figure 2A; $p = .035$; permutation test), with above-chance reconstruction for 3/8 individual participants ($p = .006$; binomial test). While AG and LOTC showed sensitivity to the content of the current stimulus (E2), neither region supported reconstruction of E1-1 (Figure 2B-C; *ps* $> .19$).

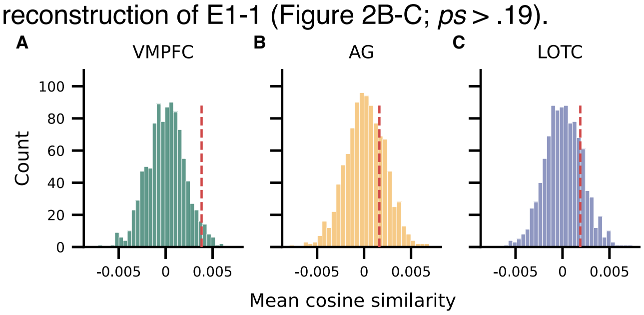

Figure 2: Mean cosine similarity between reconstructed and actual semantic components for each ROI. Red dashed lines depict actual similarity values.

Summary

In this study, we employed inverted semantic encoding models to measure temporal context reinstatement. We found that, when a stimulus was re-encountered, vmPFC supported successful reconstruction of information that 'lingered in mind' when the stimulus was first encountered. This represents novel evidence for temporal context reinstatement, specifically demonstrating reinstatement of semantic content from temporally-adjacent stimuli.

Acknowledgments

This work was supported by NIH-NINDS 1R01NS137608 and NIH-NINDS 2R01NS089729 (to B.A.K.).

References

Allen, E. J., St-Yves, G., Wu, Y., Breedlove, J. L., Prince, J. S., Dowdle, L. T., Nau, M., Caron, B., Pestilli, F., Charest, I., Hutchinson, J. B., Naselaris, T., & Kay, K. (2022). A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, *25*(1), Article 1. https://doi.org/10.1038/s41593-021-00962-x

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B., & Kurzweil, R. (2018). *Universal Sentence Encoder* (arXiv:1803.11175). arXiv. https://doi.org/10.48550/arXiv.1803.11175

Folkerts, S., Rutishauser, U., & Howard, M. W. (2018). Human Episodic Memory Retrieval Is Accompanied by a Neural Contiguity Effect. *Journal of Neuroscience*, *38*(17), 4200–4211. https://doi.org/10.1523/JNEUROSCI.2312-17.2018

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Howard, M. W., & Kahana, M. J. (2002). A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

Humphreys, G. F., Ralph, M. A. L., & Simons, J. S. (2021). A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition. *Trends in Neurosciences*, *44*(6), 452–463. https://doi.org/10.1016/j.tins.2021.01.006

Konkle, T., & Caramazza, A. (2013). Tripartite Organization of the Ventral Stream by Animacy and Object Size. *Journal of Neuroscience*, *33*(25), 10235–10242. https://doi.org/10.1523/JNEUROSCI.0983-13.2013

Lee, H., & Kuhl, B. A. (2016). Reconstructing Perceived and Retrieved Faces from Activity Patterns in Lateral Parietal Cortex. *Journal of Neuroscience*, *36*(22), 6069–6082. https://doi.org/10.1523/JNEUROSCI.4286-15.2016

Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences*, *108*(31), 12893–12897. https://doi.org/10.1073/pnas.1015174108

Norman, K. A., Detre, G., & Polyn, S. M. (2001). Computational Models of Episodic Memory. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (1st ed., pp. 189–225). Cambridge University Press. https://doi.org/10.1017/CBO9780511816772.011

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893–912. https://doi.org/10.1037/a0013396

Wang, Y., Lee, H., & Kuhl, B. A. (2023). Mapping multidimensional content representations to neural and behavioral expressions of episodic memory. *NeuroImage*, *277*, 120222. https://doi.org/10.1016/j.neuroimage.2023.120222

Yaffe, R. B., Kerr, M. S. D., Damera, S., Sarma, S. V., Inati, S. K., & Zaghloul, K. A. (2014). Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proceedings of the National Academy of Sciences*, *111*(52), 18727–18732. https://doi.org/10.1073/pnas.1417017112