Hippocampal-guided reconstruction of an event's prior temporal context

Futing Zou[1,#], J. Benjamin Hutchinson[2,*] & Brice A. Kuhl[2,3,*]

[1]Department of Cognitive and Psychological Sciences, Brown University, Providence, RI, USA

[2]Department of Psychology, University of Oregon, Eugene, OR, USA

[3]Institute of Neuroscience, University of Oregon, Eugene, OR, USA

[*]Co-senior author

[#]Corresponding author: futing_zou@brown.edu

Abstract

Events are thought to be encoded into memory in the context of temporally-adjacent events. Here, using fMRI, we show that when a visual stimulus from the past is re-encountered, lateral occipitotemporal cortex (LOTC) reinstates visual content from stimuli that were adjacent to the original encounter. This LOTC reinstatement effect was selective to stimuli that were subjectively remembered and was mediated by stimulus-specific activity patterns within the hippocampus.

Main

Influential computational models argue that individual events are encoded into episodic memory in the context of temporally-adjacent events[1]. These 'temporal context representations' are thought to later be reinstated when events are remembered[2]—a process putatively supported by the hippocampus[3]. However, directly measuring temporal context reinstatement in the brain is surprisingly challenging. The most compelling evidence comes from intracranial electrophysiological studies which show that when a stimulus is remembered, it evokes neural activity patterns that resemble activity patterns that were temporally-adjacent to the initial encoding of that stimulus[4–6]. However, an open question is whether reinstatement actually includes information about the content of temporally-adjacent events. Here, we tested this idea using data from a massive human fMRI dataset (N = 8, 30-40 fMRI sessions per subject). We predicted that when a stimulus is re-encountered, high-level visual cortical areas would carry information about *other stimuli* that were previously encoded nearby in time. Importantly, we also predicted that this temporal context reinstatement would be guided by the hippocampus[7–9].

During fMRI scanning, subjects performed a continuous recognition memory task which required, on each trial, indicating whether a natural scene image was 'old' or 'new.' Behavioral memory performance for this dataset has previously been described[10]. Here, we focused on the first two encounters with each stimulus (1st encounter = E1, 2nd encounter = E2). We operationalized the temporal context of E1 as the content of temporally-adjacent visual scenes (E1-1 and E1+1; 'temporal context images').

To measure temporal context reinstatement, we implemented an inverted encoding model in which fMRI activity patterns elicited by E2 were used to reconstruct the visual content of the E1 temporal context images (**Fig. 1**). Specifically, we extracted visual features of images from the Contrastive Language-Image Pre-Training (CLIP) image encoder (ViT-32 transformer backbone[11])—a deep neural

network that captures high-level visual representations in the human brain[12]. We then performed principal component analysis (PCA) on the extracted CLIP image embeddings and used the top 20 principal components (PCs) to define the information contained within each image. Next, we constructed voxel-wise encoding models to learn a direct linear mapping from the PCs (image content) to fMRI activity patterns. The encoding models used cross-validated ridge regression, separately for each subject. For each held-out fold, we inverted the model to reconstruct the image PCs from fMRI activity patterns. Reconstructed PCs were compared to actual PCs using cosine similarity.

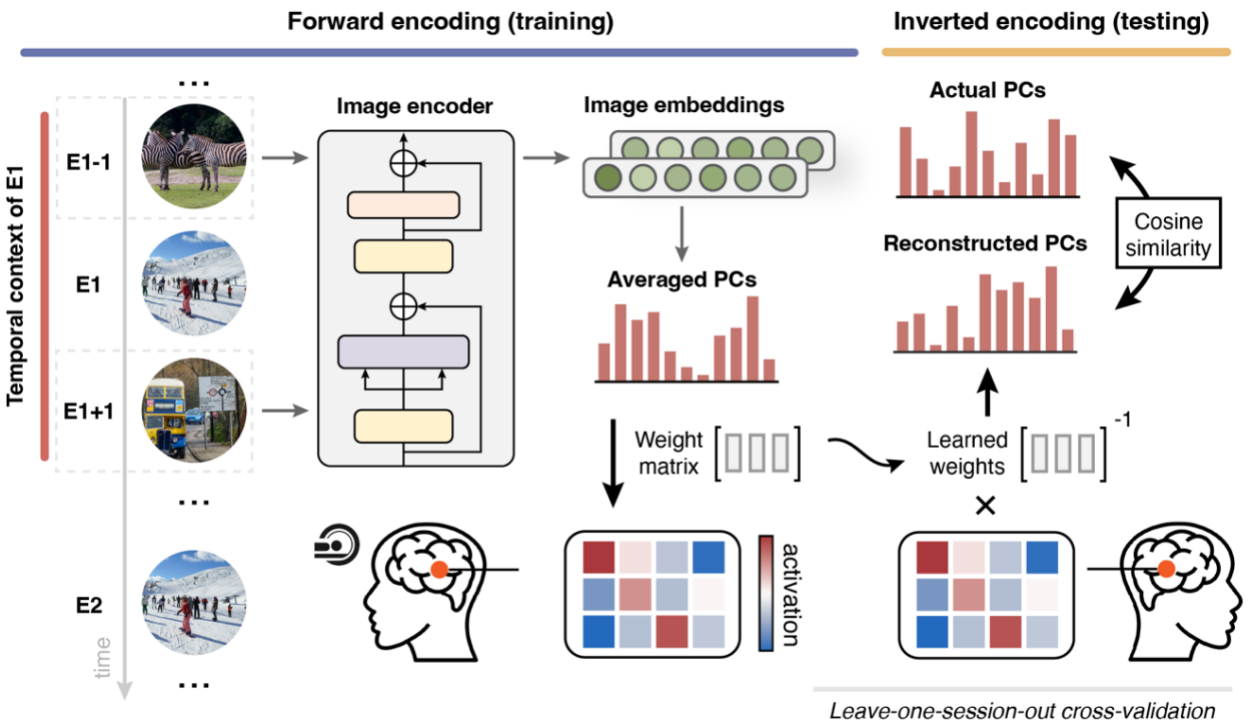

Fig 1. An fMRI inverted encoding model for temporal context reconstruction. We operationalized the temporal context of a stimulus' original encounter (E1) as the content of temporally-adjacent visual scenes (E1-1 and E1+1). Scene content was measured by extracting embeddings from the CLIP visual encoder. PCA was then applied to identify the top 20 PCs. PC scores were averaged across the E1-1 and E1+1 images and these average scores (E1's temporal context) were used in voxel-wise encoding models to learn a linear mapping to fMRI activity evoked at E2. The encoding models used ridge regression with leave-one-session-out cross-validation, separately for each subject. For each held-out fold, the encoding model was inverted such that the image PC scores were reconstructed from the fMRI activity patterns. Reconstruction accuracy was evaluated by computing the cosine similarity between the reconstructed and actual image PC scores.

We tested for temporal context reinstatement in three *a priori* regions of interest (ROIs) that are known to support memory reinstatement across different representational levels (**Fig. 2b**): angular gyrus (AG)[13], lateral occipitotemporal cortex (LOTC)[14], and early visual cortex (V1)[15]. As a control region, we also included primary motor cortex (M1). Our initial analyses were restricted to E1/E2

stimuli that were (1) correctly identified as "new" at E1 (correct rejection) and correctly recognized as "old" at E2 (hit) and (2) were encountered in the same fMRI session, but in different scan runs.

Before testing for temporal context reinstatement, we first validated our approach by testing whether fMRI activity patterns at E2 contained information about the current image (E2). In line with prior work[12], we found that visual content information was robustly reflected in AG, LOTC, and V1 (**Fig. 2a**; permutation tests: *P*s < 0.001, corrected), but not in M1 (*P* = 0.07, corrected). These results confirm that the ROIs were sensitive to content within natural scene images and validate our approach for reconstructing image content from fMRI activity patterns.

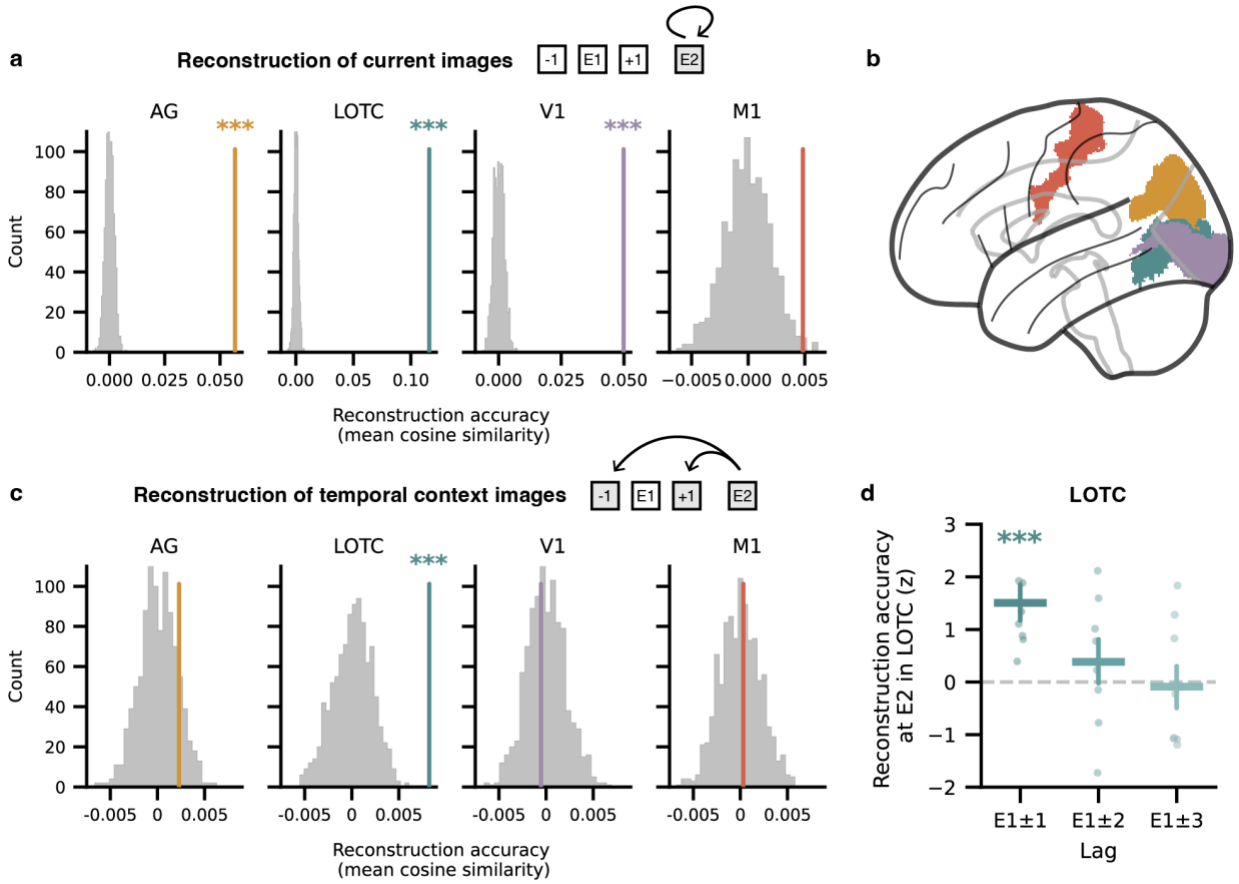

Fig 2. Temporal context reconstruction in LOTC. a, Reconstruction accuracy (from inverted encoding model) for PC scores of current images (E2) from fMRI activity patterns at E2. Reconstruction of current images was above chance in AG, LOTC, and V1 (*P*s < 0.001), but not in M1 (*P* = 0.07). Statistical significance was determined by permutation test (shuffling reconstructed image PC scores within subjects to generate a null distribution of the group-level average similarity between actual and reconstructed scores). **b,** *A priori* ROIs included bilateral angular gyrus (AG; yellow), lateral occipitotemporal cortex (LOTC; green), early visual cortex (V1; purple), and primary motor cortex (M1; red). **c,** Reconstruction accuracy of temporal context images (E1-1 and E1+1) from fMRI activity patterns at E2. Reconstruction of temporal context images was above chance in LOTC (*P* < 0.001), but not in AG, V1, or M1 (*P*s > 0.15). In **a** and **c**, density plots show the across-subject mean of the observed reconstruction accuracy (colored vertical line) relative to the null distribution (gray histograms; 1,000 iterations). Reported P-values are Bonferroni corrected to account for

multiple comparisons. **d,** Temporal context reconstruction accuracy in LOTC decreased as a function of the temporal context images' distance to E1 ($F_{2,14}$ = 3.79, P = 0.048; repeated-measures ANOVA). Dots represent effects for each of the 8 subjects (*z*-scores from subject-specific permutation tests; see Methods) and error bars depict s.e.m.; ***P < 0.001.

Of critical interest was whether the fMRI activity patterns at E2 contained information about images that were temporally-adjacent to E1 (PCs from the E1 temporal context images). Importantly, because we only analyzed fMRI data from E2, we avoided any issues related to fMRI autocorrelation across E1 and E1 temporal context images. We found that information within the E1 temporal context images was successfully reconstructed from E2 activity patterns in LOTC (**Fig. 2c**; permutation test: P < 0.001, corrected), but not in AG or V1 (Ps > 0.15, corrected). As expected, reconstruction was not above chance in M1 (P = 0.64, corrected). Notably, reconstruction in LOTC was not significant when images were not recognized at E2 (permutation test: P = 0.90, uncorrected). Thus, we observed robust LOTC-based reconstruction of E1's temporal context, but only when subjects had the phenomenological experience of remembering a stimulus at E2. This reinstatement effect is striking when considering that subjects were only deciding whether each E2 stimulus was 'old' or 'new'; in other words, reinstatement of stimuli that neighbored E1 was entirely incidental to the task.

Behavioral studies have shown that when a list of words is recalled from memory, stimuli tend to be clustered according to their temporal proximity during encoding[16]. These 'temporal contiguity effects'—which are a primary inspiration for theories of temporal context reinstatement—have a characteristic shape wherein recall clustering is strongest for stimuli that were immediately adjacent during encoding (± 1) and falls off dramatically for stimuli further apart during encoding (effects typically level-off for stimuli separated by ± 3 positions)[2,17,18]. To test for a similar function in our reinstatement effects, we ran additional reconstruction analyses that defined temporal context images as E1 ± 2 and, separately, as E1 ± 3. In contrast to the significant reconstruction effect in LOTC for E1 ± 1, reconstruction was numerically lower and not above chance at E1 ± 2 (P = 0.15) or E1 ± 3 (P = 0.57; permutation test). An analysis of variance (ANOVA) with lag to E1 as a factor confirmed a main effect of lag on reconstruction accuracy (**Fig. 2d**; $F_{2,14}$ = 3.79, P = 0.048). Thus, reconstruction accuracy decreased as a function of the temporal context images' distance to E1, strongly mirroring temporal contiguity effects in free recall.

Finally, we tested whether temporal context reinstatement in LOTC was mediated by the hippocampus. In prior work using this same fMRI dataset, we found that re-expression of hippocampal activity patterns across stimulus encounters predicted subsequent behavioral expressions of temporal memory (memory for when E1 occurred)[19]. While this finding is compatible with an account wherein the hippocampus reinstates past temporal contexts (thereby benefiting temporal memory), it is also open to other interpretations.

To test whether the hippocampus guided temporal context reinstatement, we first ran linear mixed-effects models using stimulus-specific E1-E2 pattern similarity (in medial temporal lobe regions) to predict temporal context reconstruction accuracy in LOTC. These models controlled for additional factors that might co-vary with temporal context reinstatement (E1 onset, E1-E2 lag; see Methods). Among the medial temporal lobe ROIs we considered, only hippocampal subfields CA1 and CA2/3/dentate gyrus (CA23DG) predicted temporal context reinstatement in LOTC (**Fig. 3a**; CA1: β = 0.024, P = 0.007; CA23DG: β = 0.027, P = 0.005; linear mixed-effects models, uncorrected). Specifically, when hippocampal activity patterns were more consistent across stimulus encounters (greater E1-E2 similarity), LOTC expressed stronger reinstatement of the information that *surrounded* the initial stimulus encounter.

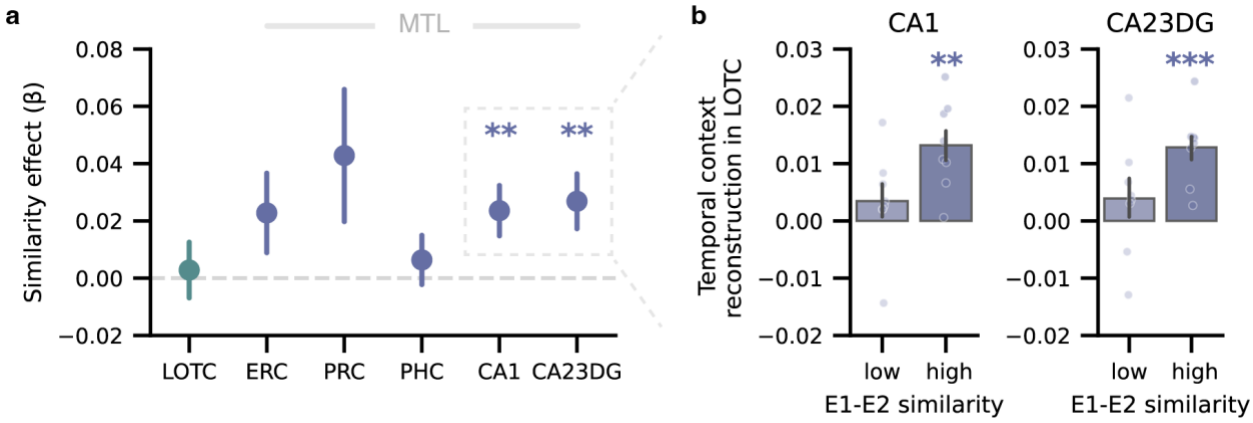

Fig 3. Hippocampus-mediated temporal context reconstruction. a, Temporal context reconstruction accuracy in LOTC was predicted by stimulus-specific pattern similarity (E1-E2 similarity) in hippocampal subfields CA1 and CA23DG (CA1: $\beta = 0.024$, $P = 0.007$; CA23DG: $\beta = 0.027$, $P = 0.005$; other ROIs: Ps > 0.11; linear mixed-effects models, uncorrected). **b,** LOTC temporal context reconstruction accuracy was significantly greater than chance (zero) for images with high E1-E2 pattern similarity in the hippocampus (CA1: $t_7 = 4.75$, $P = 0.002$; CA23DG: $t_7 = 5.56$, $P < 0.001$; two-tailed one-sample t tests, uncorrected). Dots indicate individual subject data and bars represent the across-subject mean. In all panels, error bars depict s.e.m.; $^{**}P < 0.01$; $^{***}P < 0.001$.

As a complementary analysis, we binned images with high versus low E1-E2 similarity (subject-specific median splits). We found that LOTC reconstruction accuracy did not differ from zero for images with low CA1 or CA23DG similarity (**Fig. 3b**; Ps > 0.30) but was significantly greater than zero for images with high similarity (CA1: $P = 0.002$; CA23DG: $P < 0.001$; two-tailed one-sample t tests, uncorrected). Reconstruction accuracy was also significantly greater for high vs. low similarity trials (CA1: $P = 0.015$; CA23DG: $P = 0.025$; permutation tests, uncorrected).

Notably, E1-E2 similarity in LOTC did not predict temporal context reconstruction accuracy in LOTC ($P = 0.77$, uncorrected), suggesting that the hippocampus and LOTC played complementary, but distinct roles in temporal context reinstatement. Specifically, our findings are consistent with the idea that the hippocampus guided cortical reinstatement (in LOTC)[20–22].

Collectively, our findings provide unique evidence that remembering a stimulus elicits incidental reinstatement of the temporal context in which the stimulus previously occurred. In contrast to prior human studies which only measured reinstatement of neural activity patterns[4,5,19], we explicitly show reinstatement of visual content from temporally-adjacent events. This finding validates temporal context theories which argue that temporally-adjacent stimuli become part of a stimulus' temporal context[1]. Moreover, by specifically linking trial-by-trial measures of cortical reinstatement to pattern similarity within the hippocampus, we also identify a division of labor wherein the hippocampus contains stimulus-specific codes that are retained across stimulus encounters and that guide reinstatement of high-level visual information in neocortex.

Methods

Overview

All analyses were performed using the Natural Scenes Dataset (NSD), a large-scale dataset of high-resolution (7T) fMRI data from eight subjects who each performed a continuous recognition task on thousands of unique natural scene images over the course of 30–40 scan sessions. The results reported in this study are based on data from all of these sessions, from all subjects who participated in the NSD study. A detailed description of the dataset is reported in the original data publication[10]. Below we describe the specific methods relevant to the current study.

Dataset

Subjects

Eight subjects (six females and two males; 19-32 years old) recruited from the University of Minnesota community participated in the NSD study. All subjects were right-handed with no known cognitive deficits or color blindness and with normal or corrected-to-normal vision. Subjects were not involved in the design or planning of the NSD experiment. Informed written consent was obtained from all subjects before the start of the study, and the experimental protocol was approved by the University of Minnesota Institutional Review Board.

Stimuli

All images used in NSD were drawn from the Microsoft Common Objects in Context (COCO) database[23]. A total of 73,000 unique images of natural scenes were prepared with the intention that each subject would be exposed to 10,000 distinct images (9,000 unique images and 1,000 shared images across subjects) three times each across the 40 scan sessions. Image presentations were pseudo-randomly distributed across sessions over the course of almost a year. The presentation structure was determined in advance and fixed across subjects so that difficulty of the recognition task was roughly similar across subjects. To provide a sense of the overall experimental design, the mean number of distinct images shown once, twice, or three times within a session was 437, 106, and 34, respectively.

Design and procedure

A detailed description of the task design is reported in the original data publication[10]. Briefly, subjects performed a continuous recognition task in which they decided, on each trial, whether the current image had been seen at any previous point in the experiment ('old') or if they had not seen it before ('new'). Subjects completed up to 40 scan sessions in which they viewed up to 10,000 different color natural scene images. Each scene image was presented three times (if subjects completed all 40 sessions), with presentations pseudo-randomly spaced over the course of the entire experiment. Each scanning session consisted of 12 runs (750 trials per session). Each trial consisted of the presentation of an image for 3 s followed by a 1-s gap. Subjects were able to respond during the entire 4-s period and were also permitted to make multiple responses per trial (if they changed their mind). Because trials with multiple responses potentially captured a mixture of cognitive operations / decision making, we opted to exclude these trials from our analyses.

Four of the subjects completed the full set of 40 NSD scan sessions. Due to constraints on subject and scanner availability, two subjects completed 30 sessions, and two subjects completed 32 sessions. Accordingly, each subject viewed a total of 9,209-10,000 unique natural scene images across 22,500-30,000 trials.

fMRI data acquisition and preprocessing

MRI data was collected at the Center for Magnetic Resonance Research at the University of Minnesota. Imaging was performed on a 7T Siemens Magnetom passively-shielded scanner with a single-channel-transmit, 32-channel-receive RF head coil. Functional images were acquired using whole-brain gradient-echo echo-planar imaging (EPI) at 1.8-mm resolution and 1.6-s repetition time.

Details of the preprocessing of anatomical and functional data are reported in the original data publication[10]. Briefly, functional data were preprocessed by performing one temporal resampling to correct for slice time differences and one spatial resampling to correct for head motion within and across scan sessions, EPI distortion, and gradient non-linearities. Informed by the original data publication, the current study used the 1.0-mm volume preparation of the functional time-series data and 'version 2' of the NSD single-trial betas.

Data Analysis

The current study focused on the first two encounters with each stimulus (1st encounter = E1, 2nd encounter = E2). Our main analyses were restricted to E1/E2 stimuli that were (1) correctly identified as "new" at E1 (i.e., correct rejection) and correctly recognized as "old" at E2 (i.e., hit) and (2) were encountered in the same fMRI session, but in different scan runs. This resulted in a total of 721-1,875 stimuli that were used in the main analyses for each subject.

Regions of interest

Cortical ROIs, including bilateral angular gyrus (AG), lateral occipitotemporal cortex (LOTC), early visual cortex (V1), and primary motor cortex (M1) were drawn from the surface-based Human Connectome Project multimodal parcellation (HCP-MMP) atlas of human cortical areas[24]. All medial temporal lobe (MTL) subregions were manually drawn on the high-resolution T2 images obtained for each subject. These MTL ROIs included two subfields of the hippocampus (CA1 and CA2/3/dentate gyrus), along with entorhinal cortex (ERC), perirhinal cortex (PRC), and parahippocampal cortex (PHC).

Temporal context reconstruction analysis

To measure temporal context reinstatement (**Fig. 1**), we first extracted visual features of NSD scene images from the image encoder in OpenAI's pretrained CLIP model with vision transformer backbone (ViT-32)[11]. We used the PyTorch implementation of this model. We then performed principal component analysis (PCA) on the extracted image embeddings to reduce the dimensionality of the visual features[25]. The first 20 principal components (PCs) were used to define the visual content of each image. For the primary temporal context reinstatement analyses, PC scores were averaged across E1-1 and E1+1 and these average scores were used in the encoding models.

To learn a direct linear mapping from the image to neural activity, we built voxel-wise encoding models to predict fMRI responses from image PC scores using cross-validated ridge regression (using scikit-learn Python package), separately for each subject and each ROI. The regularization parameter was determined by grid search over 11 values logarithmically spaced from 10^{-5} to 10^{5} to minimize prediction error. We then inverted this encoding model and applied it to held-out test data (leave-one-session-out) to reconstruct the image PC scores from fMRI activity patterns. Reconstruction accuracy was assessed by comparing the reconstructed and actual image PC scores using cosine similarity.

Pattern similarity analysis

Pattern similarity (within a given ROI) was calculated as the Pearson correlation between fMRI activity patterns evoked during different image encounters. Correlations were Fisher's z-transformed before further analyses were performed. To avoid potential contamination from BOLD signal autocorrelation, all pattern similarity analyses were performed by correlating activity patterns for stimuli encountered across different runs (i.e., correlations were never performed within the same scanning run).

As in prior work[19], we computed a stimulus-specific measure of E1-E2 pattern similarity. Specifically, for each target image, we computed 'within-image' pattern similarity (E1 and E2 = same stimulus) and 'across-image' pattern similarity (E1 and E2' = different stimuli). The E2' images used to compute across-image similarity were chosen from the same set of sessions (but different scanning runs) as the target image's E1 and E2, thus controlling for differences in lag between image encounters. Further, E2' images were chosen such that they had the same memory outcomes at the first two encounters as the target image. For target images with multiple possible E2' images based on the criteria above, the median value of the across-image similarity scores was used. The across-image similarity was then subtracted from within-image similarity to yield a stimulus-specific measure of E1-E2 similarity for each image.

Statistics

Statistical analyses were performed using a combination of permutation tests, repeated-measures ANOVA, one-sample *t* tests, and mixed-effects regression models. A threshold of $P < 0.05$ was used to assess statistical significance.

To assess the statistical significance of reconstruction accuracy (**Fig. 2a**, **c**), we used non-parametric permutation tests (1,000 iterations). For each iteration of the permutation test, we shuffled the reconstructed image PC scores within each subject, then recomputed the average reconstruction accuracy (cosine similarity) across subjects. The 1,000 iterations yielded a null distribution of group-level reconstruction accuracy. We then computed one-tailed *P* values by determining how many samples from the null distribution exceeded the actual group-level reconstruction accuracy. For analyses shown in **Fig. 2a** and **2c**, *P* values were Bonferroni-corrected for multiple comparisons across 4 ROIs (AG, LOTC, V1, M1).

To assess reconstruction accuracy as a function of temporal context images' lag to E1 (**Fig. 2d**), permutation tests were performed within subjects (1,000 iterations) to obtain a subject-specific null distribution for each lag (E1 ± 1, E1 ± 2, E1 ± 3). The actual reconstruction accuracy for each subject, at each lag, was compared to the null distribution and expressed as a *z* score. Repeated-measures ANOVA was then used to test whether reconstruction accuracy (mean *z* score) varied by lag.

For analyses linking temporal context reconstruction accuracy to pattern similarity (**Fig. 3**), we used two approaches. The first approach (**Fig. 3a**) used mixed-effects regression models. Specifically, we ran mixed-effects linear regression models that predicted trial-by-trial reconstruction accuracy (cosine similarity) from stimulus-specific pattern similarity while including (controlling for) several additional factors. These additional factors of no interest included the lag between the beginning of the first trial in the experiment and E1 onset and the E1-E2 lag. All temporal lags were quantified in seconds and transformed with the natural logarithm to correct for non-normality in the distribution. All mixed-effects models included random intercepts per subject. The second approach (**Fig. 3b**) compared reconstruction accuracy for trials associated with high versus low pattern similarity. Specifically, we median split trials within each subject according to stimulus-specific pattern similarity (high vs. low) in a given ROI. We first compared reconstruction accuracy to chance (zero), separately for each bin (high vs. low pattern similarity), using two-tailed one-sample *t* tests.

Separately, we compared reconstruction accuracy between the two bins (high vs. low pattern similarity) using permutation tests (1,000 iterations). For each permutation, we randomly shuffled each subject's pattern similarity labels across trials and computed the mean difference, across all subjects, in reconstruction accuracy for high vs. low pattern similarity trials. These permuted group-level differences served as the null distribution. Actual differences in reconstruction accuracy for high vs. low pattern similarity trials was expressed as one-tailed *P* values by determining how many samples from the null distribution exceeded the actual group-level difference. For these analyses, *P* values were not corrected for multiple comparisons because we had a strong *a priori* interest in the hippocampus (CA1 and CA23DG).

Data availability

The data used in this study are openly available as part of the Natural Scenes Dataset, publicly available at http://naturalscenesdataset.org.

Code availability

The code used for data analysis will be made available upon publication at https://github.com/futingzou/nsdTemporalContext.

Competing interests

The authors declare no competing interests.

Acknowledgements

This work was supported by NIH-NINDS 2R01NS089729 to B.A.K.

References

1. Howard, M. W. & Kahana, M. J. A Distributed Representation of Temporal Context. *Journal of Mathematical Psychology* **46**, 269–299 (2002).
2. Polyn, S. M., Norman, K. A. & Kahana, M. J. A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review* **116**, 129–156 (2009).
3. Hsieh, L.-T., Gruber, M. J., Jenkins, L. J. & Ranganath, C. Hippocampal Activity Patterns Carry Information about Objects in Temporal Context. *Neuron* **81**, 1165–1178 (2014).
4. Manning, J. R., Polyn, S. M., Baltuch, G. H., Litt, B. & Kahana, M. J. Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences* **108**, 12893–12897 (2011).
5. Folkerts, S., Rutishauser, U. & Howard, M. W. Human Episodic Memory Retrieval Is Accompanied by a Neural Contiguity Effect. *J. Neurosci.* **38**, 4200–4211 (2018).
6. Yaffe, R. B. *et al.* Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proceedings of the National Academy of Sciences* **111**, 18727–18732 (2014).
7. Eichenbaum, H. Time cells in the hippocampus: a new dimension for mapping memories. *Nature Reviews Neuroscience* **15**, 732–744 (2014).
8. Davachi, L. & DuBrow, S. How the hippocampus preserves order: the role of prediction and context. *Trends in Cognitive Sciences* **19**, 92–99 (2015).
9. Eichenbaum, H. Memory: Organization and Control. *Annual Review of Psychology* **68**, 19–45 (2017).
10. Allen, E. J. *et al.* A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat Neurosci* **25**, 116–126 (2022).
11. Radford, A. *et al.* Learning Transferable Visual Models From Natural Language Supervision. Preprint at https://doi.org/10.48550/arXiv.2103.00020 (2021).
12. Wang, A. Y., Kay, K., Naselaris, T., Tarr, M. J. & Wehbe, L. Better models of human high-level visual cortex emerge from natural language supervision with a large and diverse dataset. *Nat Mach Intell* **5**, 1415–1426 (2023).
13. Humphreys, G. F., Ralph, M. A. L. & Simons, J. S. A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition. *Trends in Neurosciences* **44**, 452–463 (2021).
14. Konkle, T. & Caramazza, A. Tripartite Organization of the Ventral Stream by Animacy and Object Size. *J. Neurosci.* **33**, 10235–10242 (2013).
15. Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K. & Gallant, J. L. A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage* **105**, 215–228 (2015).
16. Howard, M. W., Youker, T. E. & Venkatadass, V. S. The persistence of memory: Contiguity effects across hundreds of seconds. *Psychonomic Bulletin & Review* **15**, 58–63 (2008).
17. Sederberg, P. B., Howard, M. W. & Kahana, M. J. A context-based theory of recency and contiguity in free recall. *Psychol Rev* **115**, 893–912 (2008).

18. Polyn, S. M. & Kahana, M. J. Memory search and the neural representation of context. *Trends in Cognitive Sciences* **12**, 24–30 (2008).

19. Zou, F. *et al.* Re-expression of CA1 and entorhinal activity patterns preserves temporal context memory at long timescales. *Nat Commun* **14**, 4350 (2023).

20. Eichenbaum, H. A cortical–hippocampal system for declarative memory. *Nat Rev Neurosci* **1**, 41–50 (2000).

21. Teyler, T. J. & DiScenna, P. The hippocampal memory indexing theory. *Behavioral Neuroscience* **100**, 147–154 (1986).

22. O'Reilly, R. C. & Norman, K. A. Hippocampal and neocortical contributions to memory: advances in the complementary learning systems framework. *Trends in Cognitive Sciences* **6**, 505–510 (2002).

23. Lin, T.-Y. *et al.* Microsoft COCO: Common Objects in Context. in *Computer Vision – ECCV 2014* (eds Fleet, D., Pajdla, T., Schiele, B. & Tuytelaars, T.) 740–755 (Springer, Cham, 2014).

24. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).

25. Wang, Y., Lee, H. & Kuhl, B. A. Mapping multidimensional content representations to neural and behavioral expressions of episodic memory. *NeuroImage* **277**, 120222 (2023).