COGNITIVE NEUROSCIENCE

Transformative neural representations support long-term episodic memory

Jing Liu[1], Hui Zhang[2], Tao Yu[3], Liankun Ren[4], Duanyu Ni[3], Qinhao Yang[1], Baoqing Lu[1], Liang Zhang[1], Nikolai Axmacher[1,2†], Gui Xue[1*†]

Memory is often conceived as a dynamic process that involves substantial transformations of mental representations. However, the neural mechanisms underlying these transformations and their role in memory formation and retrieval have only started to be elucidated. Combining intracranial EEG recordings with deep neural network models, we provide a detailed picture of the representational transformations from encoding to short-term memory maintenance and long-term memory retrieval that underlie successful episodic memory. We observed substantial representational transformations during encoding. Critically, more pronounced semantic representational formats predicted better subsequent long-term memory, and this effect was mediated by more consistent item-specific representations across encoding events. The representations were further transformed right after stimulus offset, and the representations during long-term memory retrieval were more similar to those during short-term maintenance than during encoding. Our results suggest that memory representations pass through multiple stages of transformations to achieve successful long-term memory formation and recall.

INTRODUCTION

Memory has long been conceived as a dynamic process. As Bartlett (*1*) noted, "Remembering is not the re-excitation of innumerable fixed, lifeless and fragmentary traces, but rather an imaginative reconstruction or construction." Using multivariate decoding and representational similarity analysis (RSA), an increasing number of studies have examined neural representations during encoding and retrieval. Despite notable overlap between representational patterns of individual items during encoding and retrieval (*2*–*6*), increasing evidence suggests that memory representations undergo substantial transformations [for recent reviews, see (*7*, *8*)]. First, while encoding relies on sensory cortical areas such as the occipital lobe and the ventral visual stream (in the case of visual stimuli), retrieval involves representations in the lateral parietal cortex (*9*, *10*). Second, the similarity of item-specific representations between encoding and retrieval is systematically lower than the similarity within each of these stages (*10*). Third, two very recent studies found different tuning functions of functional magnetic resonance imaging (fMRI) voxels during perception and memory (*11*, *12*), putatively reflecting a semanticization process from perceptual to conceptual representations over time (*13*, *14*). Last, neural representations during retrieval are temporally compressed as compared to those during encoding (*15*, *16*).

In light of these substantial transformations between encoding and retrieval, a critical question concerns the representational formats during encoding that promote the successful formation and retrieval of event-specific memory traces. Previous studies used deep neural networks (DNNs) as a quantitative model to characterize the distinct representational steps during perception [e.g., (*17*, *18*)]. They demonstrated that within the first few hundred milliseconds, brain activities gradually and progressively change from representing low-level visual information to higher-order categorical and semantic information (*19*, *20*). During this process, sensory inputs are being transformed into internally interpretable representations via bottom-up and top-down interactions (*21*, *22*). A very recent study revealed that transformed, abstract, semantic representational formats contributed to stable short-term memory maintenance (*20*). Do such representational transformations during encoding also contribute to successful long-term memory formation? Previous studies have shown that during encoding, higher fidelity of representations (*23*, *24*) and, in particular, representations from a late encoding window predict better subsequent memory (*25*, *26*). Nevertheless, no studies so far have linked this dynamic, transformative encoding process to representational fidelity and subsequent memory performance.

A second question concerns the onset and temporal dynamics of the representational transformations after encoding, e.g., during consolidation, since memory consolidation not only results in the strengthening of memory traces but also involves a transformation of memory traces and integration with existing knowledge (*27*). Recent studies suggested that memory consolidation may start much earlier than previously thought, i.e., at the end of encoding. For example, event boundaries during encoding, which reflect the end of an event, trigger hippocampal activity and cortical replay, a hallmark of memory consolidation (*28*–*30*). Activity during such post-encoding periods is predictive of subsequent memory (*31*, *32*). It is further posited that during this post-encoding period, items, their contexts, and unified event representations are maintained in working memory via a hypothetical episodic buffer (*33*). These representations are then transformed into a longer-lasting format that relies on both the neocortex and hippocampus (*34*). Consistently, short-term memory maintenance contributes to long-term memory performance (*35*, *36*), and neural activity during maintenance predicts long-term memory performance (*37*, *38*). Nevertheless, no existing studies have systematically tracked the dynamic

[1]State Key Laboratory of Cognitive Neuroscience and Learning and IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China. [2]Department of Neuropsychology, Institute of Cognitive Neuroscience, Faculty of Psychology, Ruhr University Bochum, Bochum 44801, Germany. [3]Beijing Institute of Functional Neurosurgery, Xuanwu Hospital, Capital Medical University, Beijing 100053, China. [4]Comprehensive Epilepsy Center of Beijing, Department of Neurology, Xuanwu Hospital, Capital Medical University, Beijing 100053, China.
*Corresponding author. Email: gxue@bnu.edu.cn
†These authors contributed equally to this work.

transformations of neural representations from perception via short-term memory maintenance to long-term memory retrieval.

The current study aimed to unveil the transformative nature of episodic memory using a comprehensive experimental design that integrates multiple memory stages from encoding to short-term maintenance and long-term memory recall. Leveraging the high spatiotemporal resolution of intracranial electroencephalogram (EEG) recordings and the analytical power of RSA and DNN modeling, we examined the impact of representational transformations on subsequent long-term memory performance. We also systematically compared item-specific neural representations across encoding, short-term maintenance, and long-term memory retrieval to examine the representational characteristics, temporal dynamics, and functional role of transformations across different memory stages. Our results demonstrate the time course and functional relevance of representational transformations during memory processing and help advance our understanding of the generative and constructive nature of episodic memory.

RESULTS

Behavioral results

Sixteen patients (mean age ± SD: 27.13 ± 6.78 years, seven females) with pharmacoresistant epilepsy who were implanted with stereotactic EEG electrodes for clinical purposes performed a combined short- and long-term memory paradigm (see Materials and Methods, Fig. 1A). A delayed matching-to-sample (DMS) task was used to investigate encoding and short-term memory maintenance, and a cued-recall task was used to probe long-term memory retrieval. Briefly, participants encoded a word-picture association for 3 s, followed by a 7-s maintenance interval during which the word remained on the screen, and the participants were required to maintain information about the associated picture. A probe picture then appeared, and participants indicated whether this was the target item or a similar lure. Participants studied 14 word-picture associations that were repeated three times in each run. After a 1-min countback task and a 1- to 4-min rest period, participants were tested for their long-term memory of these associations.

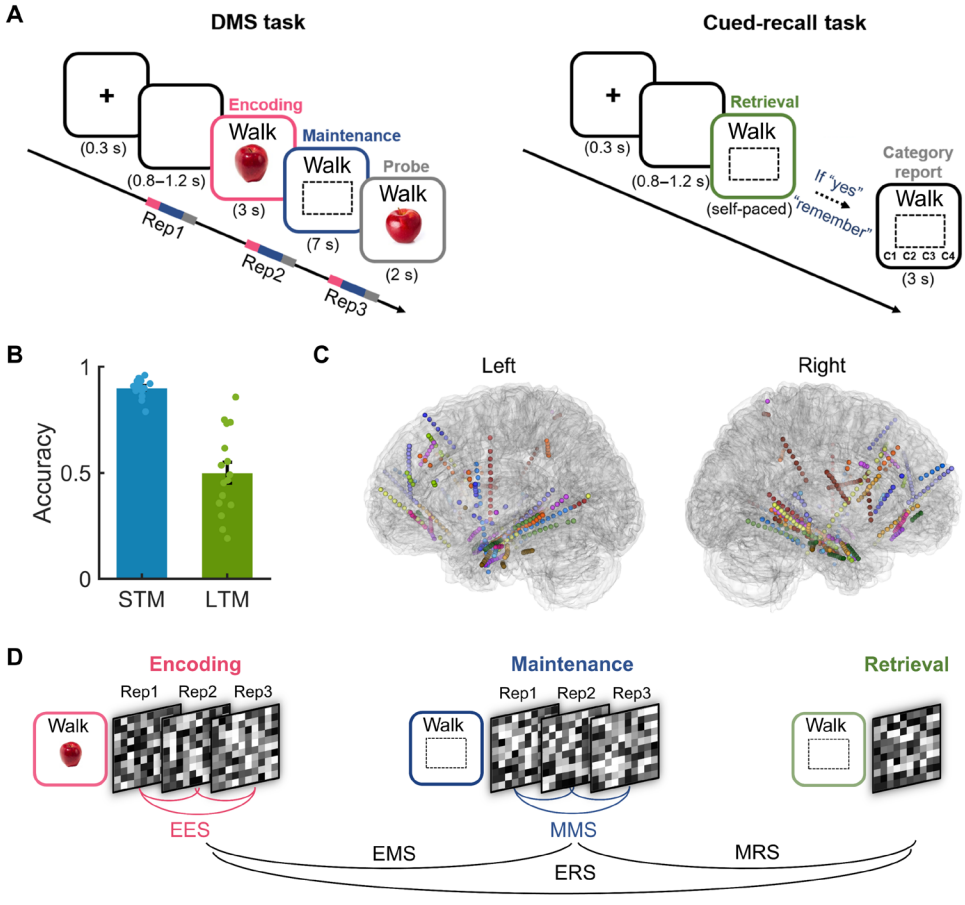

Fig. 1. Experimental protocol, behavioral performance, and analysis strategy. (**A**) Experimental paradigm. Associations between words and pictures were first encoded and maintained in a short-term memory task (DMS task; left). After each run, participants performed a long-term memory cued-recall task (right). (**B**) Accuracy during short-term memory (STM) and long-term memory (LTM) retrieval. (**C**) Localization of channels across participants. Channels were widely distributed across brain regions, including the lateral temporal lobe, the medial temporal lobe, the frontal lobe, and the parietal lobe. (**D**) Schematic overview of the pipeline for RSA. RSA was performed both within and across memory stages. For within-memory stage analyses, RSA was performed between repetitions (Rep) of the same associations during encoding and maintenance, resulting in measures of encoding-encoding similarity (EES) and maintenance-maintenance similarity (MMS). RSA was also performed across different memory stages, resulting in encoding-maintenance similarity (EMS), maintenance-retrieval similarity (MRS), and encoding-retrieval similarity (ERS). Note that all RSA was performed within runs.

During the cued-recall task, participants were presented with a cue word and asked to indicate whether they could recall the picture that was paired with this specific word by pressing buttons for "remember" or "don't remember." If they indicated to remember the picture, they were further asked to report the category of this picture by pressing one of four buttons. Participants completed four to eight runs (means ± SD: 6.25 ± 1.44). Results from the short-term memory part of this paradigm were published before (*20*).

During the short-term memory task, participants were accurate in 90.03 ± 4.22% of all trials (means ± SD; Fig. 1B). During long-term memory retrieval, participants responded with "remember" and also indicated the correct category in 50.06 ± 19.75% (means ± SD) of all trials (remembered trials). In 30.69 ± 18.97% of the trials, participants responded with "remember" but reported an incorrect category, and in 19.25 ± 21.73%, they responded with "don't remember." The latter two types of trials were jointly labeled as forgotten. The mean reaction time during long-term memory retrieval was 4.00 ± 2.27 s (means ± SD) for remembered items, significantly shorter than that for forgotten items [means ± SD: 10.43 ± 8.06 s, $t(15) = -3.942$, $P = 0.001$].

Higher encoding dynamicity predicts subsequent memory

We recorded from overall 592 artifact-free channels in the 16 patients, which were widely distributed across brain regions (means ± SD, 37.0 ± 12.98; Fig. 1C). In the first analysis, we examined the representational transformations during memory encoding and their relationship to subsequent memory. Existing studies have shown that encoding visual objects depends on a series of processing steps that range from the representation of early perceptual features such as edges or colors via intermediate steps that involve representations of complex forms to the extraction of conceptual and semantic information (*19*, *39*). However, it is still an open question whether the representational dynamics during encoding predict subsequent memory performance.

To address this question, we first performed an RSA during encoding (see Materials and Methods). Spectral power analysis revealed stimulus-locked power changes in a broad frequency range (fig. S1), which also showed item-specific tuning (fig. S2). As a result, spectral power across a broad frequency range (from 2 to 29 Hz in 1-Hz steps and from 30 to 120 Hz in 5-Hz steps) and across all channels was used as features in the RSA. We correlated the representational patterns between repetitions of the same word-picture association for all pairs of encoding time windows (Fig. 1D), resulting in a temporal map of within-item (WI) encoding-encoding similarity (WI EES; fig. S3). Representational dynamicity, which reflects the degree of representational change for the same item across time, was measured according to the framework of dynamic coding (*40*). This assumes that neural representations of a specific item are processed in a time-specific manner, resulting in reduced across-temporal correlations $r(t_1, t_2)$ as compared to correlations between corresponding time points, i.e., $r(t_1, t_1)$ and $r(t_2, t_2)$ (*41*). Following a previous study, the representational dynamicity index (*di*) is 1 if $r(t_1, t_1)$ and $r(t_2, t_2)$ are both greater than $r(t_1,t_2)$ and 0 otherwise. We thus obtained a temporal map of representational dynamicity indexes between every two time points for WI EES in individual participants (see Materials and Methods). The *di* values were then averaged across the two encoding dimensions of the WI EES map, resulting in time-resolved *di* values across encoding time windows. Applying a linear fit to *di* values and then comparing the coefficients across subjects against zero, we found that the representational dynamicity decreased with encoding time [$t(15) = -3.424$, $P = 0.004$; fig. S3], suggesting more dynamic representational changes during the early encoding period [see also (*20*)]. We then separately computed the representational *di* of subsequently remembered and forgotten items (Fig. 2, A and B). The *di* values of both remembered and forgotten items significantly decreased across time (all $P < 0.001$; Fig. 2C). However, the *di* values of remembered items were significantly greater than those for forgotten items (420 to 710 ms, $P_{corr} = 0.039$; 1250 to 1550 ms, $P_{corr} = 0.018$; Fig. 2C). These results indicate larger representational transformations during encoding for subsequently remembered than forgotten items.

Semantic representational formats support long-term memory formation

While the results so far demonstrate that more pronounced transformations during encoding benefit long-term memory, they do not provide information about the specific representational formats that are extracted during these transformations. Thus, we next examined the role of representational formats during encoding for successful memory formation. We compared the neural representations during different encoding windows to the representations in a visual DNN, the "AlexNet" (*42*). The "AlexNet" assigns labels to a multitude of objects and serves as an approximative quantitative model of neural representational structures during visual perception (*43*). We computed pairwise representational similarities of the stimuli used in our study by correlating the activations of the artificial neurons in each DNN layer. To simplify these representations, we averaged the similarity matrices across the first five convolutional layers and across the three fully connected layers ("early" and "late" visual representational similarity, respectively; see Materials and Methods). In addition, we compared these neural representations to representations in a semantic model, the Chinese word2vector model (*44*) (Fig. 2D). This semantic model allowed us to convert the label of each picture into a vector of 200 semantic features, which were used to compute the semantic similarity between the labels of all pairs of pictures. We found that the semantic similarity matrix was highly correlated with the late visual representational similarity from deeper layers of the DNN (fig. S4), suggesting that the word2vector model captures both cognition-derived and sensory-derived semantic information (*45*). To obtain the abstract, cognition-derived semantic similarities, we iteratively regressed out the visual similarity matrices of all DNN layers from the semantic similarity matrix. The four resulting similarity matrices of representational formats (early visual, late visual, semantic, and abstract semantic formats) were then correlated with neural similarity matrices during each encoding time window using Spearman's correlations.

When including all pictures regardless of subsequent memory performance, we found that representational formats changed dynamically across the encoding period, with a 340- to 590-ms cluster for early visual representations [$t(15) = 3.059$, $P_{corr} = 0.058$], a 290- to 650-ms cluster for late visual representations [$t(15) = 2.847$, $P_{corr} = 0.02$], a 370- to 1890-ms cluster for semantic representations [$t(15) = 4.899$, $P_{corr} < 0.001$], and a 530- to 1330-ms cluster for abstract semantic representations [$t(15) = 4.664$, $P_{corr} < 0.001$] (fig. S5). Direct comparisons across these formats revealed more pronounced semantic and abstract semantic formats during a late encoding period (630- to 1740-ms poststimulus, $P_{corr} < 0.001$; fig. S5).

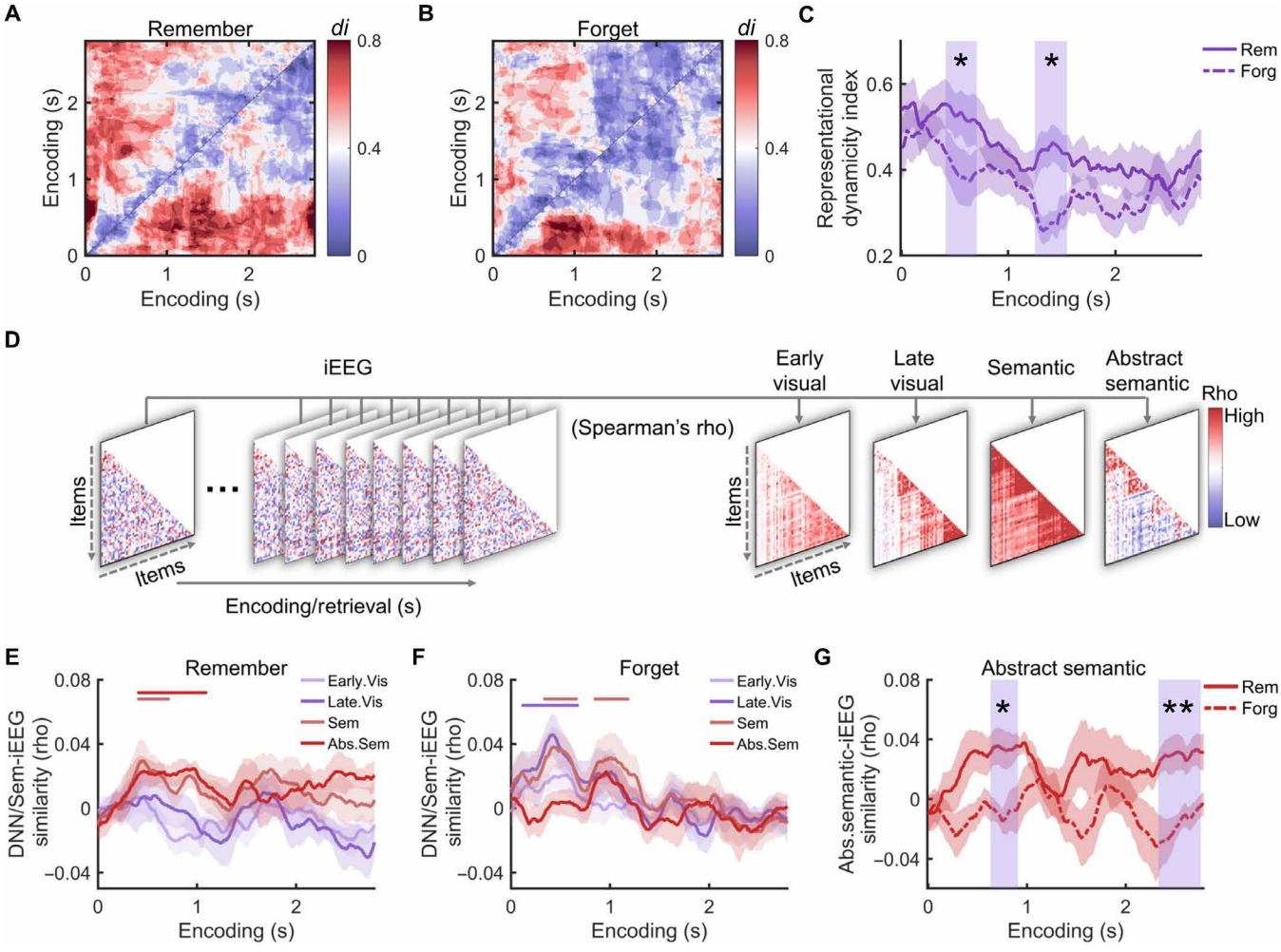

Fig. 2. Representational dynamics and formats during encoding predict long-term memory success. Representational dynamicity index (*di*) of subsequently remembered (**A**) and forgotten items (**B**). (**C**) The representational *di* was significantly greater for remembered than forgotten items within two encoding time windows (blue rectangular shaded time windows). (**D**) Analysis of representational formats: schematic depiction. During all encoding time windows, neural similarity matrices were created by pairwise correlations of neural representations of items. These neural matrices were correlated with different predictor matrices. The predictor matrices reflected early and late visual representations (derived from a DNN), semantic, and abstract semantic representations (extracted from word embedding analysis), respectively. Representational formats during encoding for subsequently remembered (**E**) and forgotten items (**F**). Colored horizontal bars indicate the time windows showing different representational formats. (**G**) Greater abstract semantic representations of subsequently remembered than forgotten items in two encoding clusters (blue rectangular shaded time windows). *$P_{corr} < 0.05$; **$P_{corr} < 0.01$.

To examine whether these representational formats predict encoding success, we performed the above analysis separately for subsequently remembered and forgotten items. Given that too few items would lead to unstable results (fig. S6), we only included subjects with at least 20 items into the analysis, resulting in 13 subjects for the remembered condition and 12 subjects for the forgotten condition. We found that for remembered items, the neural representations were significantly correlated with abstract semantic representations in a cluster between 420 and 1100 ms [$t(12) = 5.715$, $P_{corr} = 0.004$] and marginally correlated with semantic representations in a cluster between 420 and 720 ms [$t(12) = 2.349$, $P_{corr} = 0.07$; Fig. 2E]. They were not significantly correlated with early visual or late visual representations ($P_{corr} > 0.456$). For forgotten items, we found that the neural representations were significantly correlated with late visual representations in a cluster between 130 and 680 ms [$t(11) = 2.973$, $P_{corr} = 0.010$; Fig. 2F] and with semantic representations in a cluster between 350 and 670 ms [$t(11) = 3.175$, $P_{corr} = 0.030$]. They were also marginally correlated with semantic representations in a cluster between 860 and 1190 ms [$t(11) = 1.999$, $P_{corr} = 0.078$], but not with early visual or abstract semantic representations ($P_{corr} > 0.12$). Direct comparisons of these four representational formats revealed greater semantic and abstract semantic representations than visual representations for subsequently remembered items (fig. S7). In contrast, for subsequently forgotten items, late visual representations were more pronounced than early visual and abstract semantic representations (fig. S7). Very similar results were obtained when using other visual DNN models, including the VGG19 and the GoogLeNet model (fig. S8), suggesting that our results did not critically depend on the employment of a specific visual DNN model.

To directly compare the representational formats of subsequently remembered versus forgotten items, we included the nine subjects who had more than 20 items in both the remembered and the

forgotten conditions and matched the number of remembered and forgotten items within each subject. This analysis revealed two clusters that showed greater abstract semantic representations for remembered than forgotten items, including an earlier cluster between 640 and 920 ms (P_{corr} = 0.040) and a later cluster between 2340 and 2770 ms (P_{corr} = 0.004) (Fig. 2G). No significant differences were found for the other three representational formats (P_{corr} > 0.066). We also found that there was a significant increase of abstract semantic representations in the first 630 ms of the encoding period for subsequently remembered items [$t(13)$ = 3.172, P = 0.008], but not for forgotten items [$t(12)$ = 0.406, P = 0.693]. These results suggest that greater representational dynamicity of subsequently remembered items leads to a faster emergence of abstract semantic representations, which contributes to subsequent memory.

To understand the regional specificity of representational formats, we grouped all artifact-free channels into three brain regions: lateral temporal lobe (LTL), medial temporal lobe (MTL), and frontal-parietal lobe (FP). This grouping was based on the number of channels in each area (see Fig. 1C and table S1) and their functional relevance (*46*–*49*), resulting in a total of 185 clean channels from 15 participants in the LTL, a total of 116 clean channels from 16 participants in the MTL, and a total of 159 clean channels from 14 participants in the FP. Applying the same analysis to each region, we found that late visual representations were mainly observed in the LTL and the MTL, while semantic representations were observed in all three brain regions (fig. S9). In line with the whole-brain results, subsequently remembered items showed significantly greater semantic and abstract semantic representations than early visual representations in the LTL (fig. S10). In addition, greater semantic representations and weaker visual representation for subsequently remembered than forgotten items were found in the LTL, but not in the other two brain regions (fig. S11). Together, these results indicate that, compared with forgotten items, subsequently remembered items show more pronounced semantic and abstract semantic representations in the LTL.

Representational fidelity during encoding mediates the effect of encoding dynamicity on long-term memory

Why does greater encoding dynamicity contribute to better subsequent long-term memory? Previous studies have revealed that a higher fidelity of neural representations during encoding is associated with better subsequent memory (*25*, *26*). One possibility is that neural representations of those items that show a greater dynamicity are transformed into representational formats that show higher representational fidelity (i.e., contain higher amounts of item-specific information) and thus support short-term memory maintenance (*20*) and possibly also long-term memory formation.

To test this hypothesis, we examined representational fidelity, i.e., the amount of item-specific information during encoding. This was done separately for subsequently remembered and forgotten items. Representational fidelity was measured by contrasting EES between repetitions of the same word-picture associations (WI EES) versus different associations [between-item EES (BI EES)]. We found significant representational fidelity during encoding of both subsequently remembered [$t(15)$ = 4.224, P_{corr} = 0.009] and forgotten items [$t(15)$ = 6.142, P_{corr} = 0.023] (Fig. 3A). The temporal cluster of remembered items occurred much later than the cluster of forgotten items (460 to 2200 ms versus 240 to 1120 ms). Direct comparison revealed greater WI EES for subsequently remembered versus forgotten items in an encoding cluster between 590 and 1930 ms [$t(15)$ = 3.402, P_{corr} = 0.019; Fig. 3B]; no cluster showed the opposite effect (P_{corr} > 0.419). Within this cluster, we found a significant interaction between representational fidelity and memory [$F(1,15)$ = 6.102, P = 0.026], as indicated by significant representational fidelity (i.e., WI versus BI EES) for remembered items [$t(15)$ = 3.738, P = 0.002] but not for forgotten items [$t(15)$ = 1.168, P = 0.262] (Fig. 3C). Notably, this cluster overlapped with the time windows showing pronounced semantic and abstract semantic representations (fig. S5). To further probe the representational format during this time period, we correlated the neural similarity matrix averaged across this cluster with visual and semantic similarity matrices. This analysis revealed that the neural representations in this cluster were significantly correlated with semantic [$t(15)$ = 5.042, P < 0.001] and abstract semantic representations [$t(15)$ = 3.812, P = 0.002], but not with early or late visual representations (P > 0.186). Direct comparisons revealed greater semantic and abstract semantic representations than visual representations in this cluster (all P_{FDR} < 0.029; fig. S12). Together, these results show that representations of subsequently remembered items contain more item-specific information in semantic and abstract semantic formats.

We then investigated whether greater dynamicity improved long-term memory formation via enhancing representational fidelity during encoding. To this end, we performed a multilevel mediation analysis (see Materials and Methods), using the single-item *di* values as predictors, long-term memory performance as the outcome, and single-item representational fidelity (i.e., WI EES versus BI EES of each item) in the cluster shown in Fig. 3B as a mediator. This analysis showed that representational fidelity during encoding partially mediated the effect of representational dynamicity on subsequent long-term memory (indirect effect: 0.035, 95% confidence interval: 0.0006 to 0.070, P = 0.018; direct effect: 0.245, 95% confidence interval: −0.002 to 0.044, P = 0.054; Fig. 3D).

Significant item-specific EES were found in all three regions for subsequently remembered items (all $P_{corr} \leq 0.034$; fig. S13), but this effect was only significant in the LTL (P_{corr} = 0.008) and marginally significant in the MTL (P_{corr} = 0.072) for subsequently forgotten items. Direct comparisons revealed greater WI EES for subsequently remembered than forgotten items in the FP (P_{corr} = 0.002). Within this cluster, we found significant item-specific EES for remembered items [$t(13)$ = 4.690, P < 0.001], but not for forgotten items [$t(13)$ = −0.428, P = 0.675]. However, the representational dynamicity in none of the three regions predicted long-term memory performance (all P > 0.2). These results indicate that encoding dynamicity supports memory formation by increasing the amount of item-specific representations during encoding, and this effect was observed at the whole-brain level.

Lack of item-specific reinstatement of representations from encoding during memory retrieval

The above results indicate that greater representational transformations (i.e., encoding dynamicity) are associated with subsequent long-term memory. Do these transformed representations undergo further modifications before memory retrieval? To address this question, we compared representations during encoding with those during memory retrieval. First, we investigated whether the pattern of neural activity during encoding recurred during long-term memory retrieval. To this end, we performed an analysis of encoding-retrieval similarity (ERS) by pairing the same items during encoding and

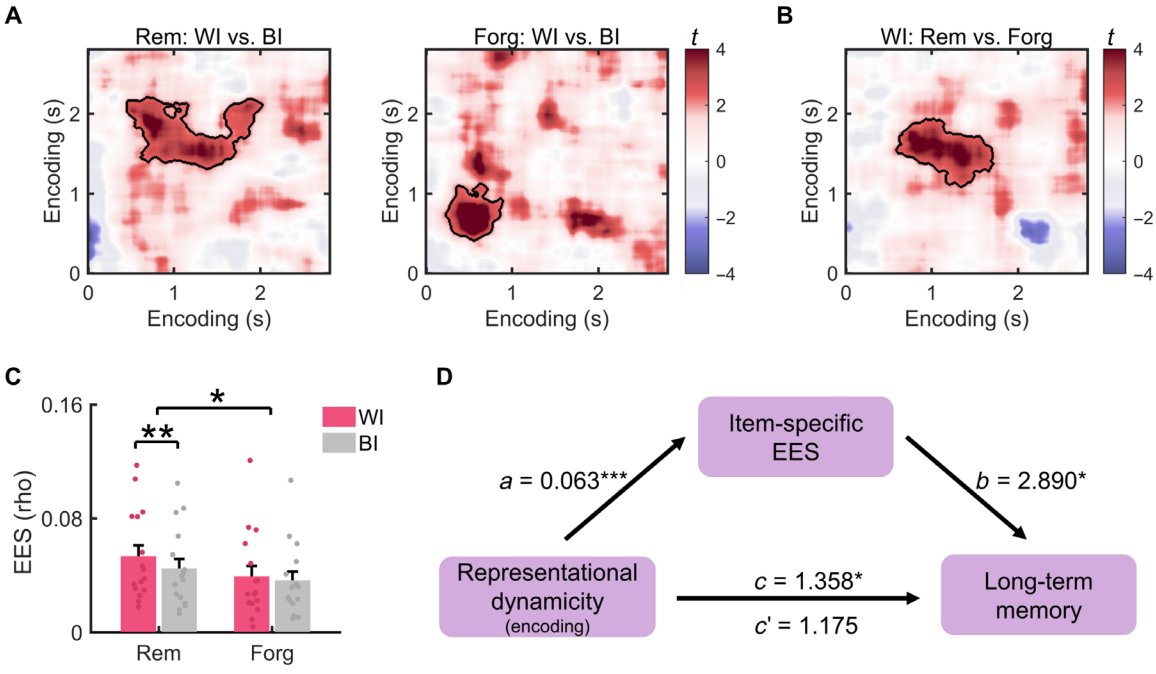

Fig. 3. Representational fidelity during encoding improves long-term memory and mediates the beneficial effect of encoding dynamicity. (**A**) Analysis of representational fidelity (i.e., WI EES versus BI EES) for subsequently remembered (left) and forgotten items (right). The black lines circle the clusters with significantly greater WI than BI EES for remembered and forgotten items. (**B**) Comparison of WI similarity between subsequently remembered and forgotten items during the encoding period. A significant cluster circled by the black line showed greater WI similarity across repetitions for subsequently remembered than forgotten items. (**C**) Representational fidelity within the cluster in (B). In this cluster, the representational fidelity of subsequently remembered items was significant, as well as the interaction between memory (remembered versus forgotten) and representational fidelity (WI versus BI). (**D**) Representational fidelity (i.e., item-specific EES) in the cluster in (B) mediates the effect of encoding dynamicity on subsequent long-term memory. *$P < 0.05$; **$P < 0.01$.

retrieval (WI ERS; Fig. 1D). Specifically, we correlated neural activity patterns during encoding with activity patterns before the retrieval response. Consistent with previous studies (*6*, *47*), we computed WI ERS for remembered and forgotten items (Fig. 4A). The direct comparison revealed a cluster that showed significantly greater WI ERS for remembered than forgotten items [$t(15) = 4.234$, $P_{corr} < 0.001$, clusters corrected for multiple comparisons; Fig. 4B].

To further characterize the temporal profile of pattern reinstatement, we examined WI ERS for each encoding time window after averaging ERS values across the entire retrieval time window (i.e., from 2000 ms before the retrieval response up to the response). This analysis revealed that WI ERS was significantly greater for remembered than forgotten items from 670 ms after stimulus onset to the end of encoding ($P_{corr} < 0.001$) (Fig. 4C, top). Consistent with previous studies (*46*, *50*), this result indicates that, at the whole-brain level, neural activity from a relatively late encoding stage is reinstated during successful retrieval. Reversely, we also examined WI ERS in each retrieval time window by averaging across all encoding time windows. This analysis showed that ERS was significantly greater for remembered than forgotten items from 1680 ms before the response up to the response ($P_{corr} < 0.001$; Fig. 4C, bottom).

Next, we examined whether reinstatement was item specific. We thus contrasted the similarity between encoding and retrieval of the same items (WI ERS) with the similarity between the encoding of one and retrieval of a different item (BI ERS; see Materials and Methods). This analysis did not reveal any significant clusters in which WI similarity values would exceed BI similarity values, for either remembered ($P_{corr} > 0.708$; Fig. 4D) or forgotten items ($P_{corr} > 0.543$; Fig. 4E). Only 0.55% of all time points showed greater WI than BI similarity values at an uncorrected level for remembered items. Focusing on the cluster that showed greater ERS for remembered than forgotten items again did not reveal any evidence for item-specific reinstatement, for either remembered [$t(15) = 1.077$, $P = 0.298$] or forgotten items [$t(15) = -1.830$, $P = 0.087$]. For each individual brain region, we found similar results of greater neural pattern reinstatement for subsequently remembered than forgotten items but a lack of item-specific reinstatement (fig. S14). These results indicate that pattern reinstatement during retrieval does not contain robust item-specific encoding information.

Reinstatement of short-term memory representations during long-term memory retrieval

Our results so far have shown that item-specific representations during encoding were not reinstated during long-term memory retrieval. By introducing a short-term memory stage, we next tested how neural representations during this post-encoding period were related to the representations during retrieval. First, we calculated maintenance-maintenance similarity (MMS) separately for remembered and forgotten items, which revealed pronounced item-specific representations of both item types (both $P < 0.05$; fig. S15), suggesting that the maintenance stage contained item-specific information for both remembered and forgotten items.

We then examined maintenance-retrieval similarity (MRS) by correlating neural representations in individual maintenance time windows with those during long-term memory retrieval and then averaging MRS values across the maintenance period. Again, item-specific

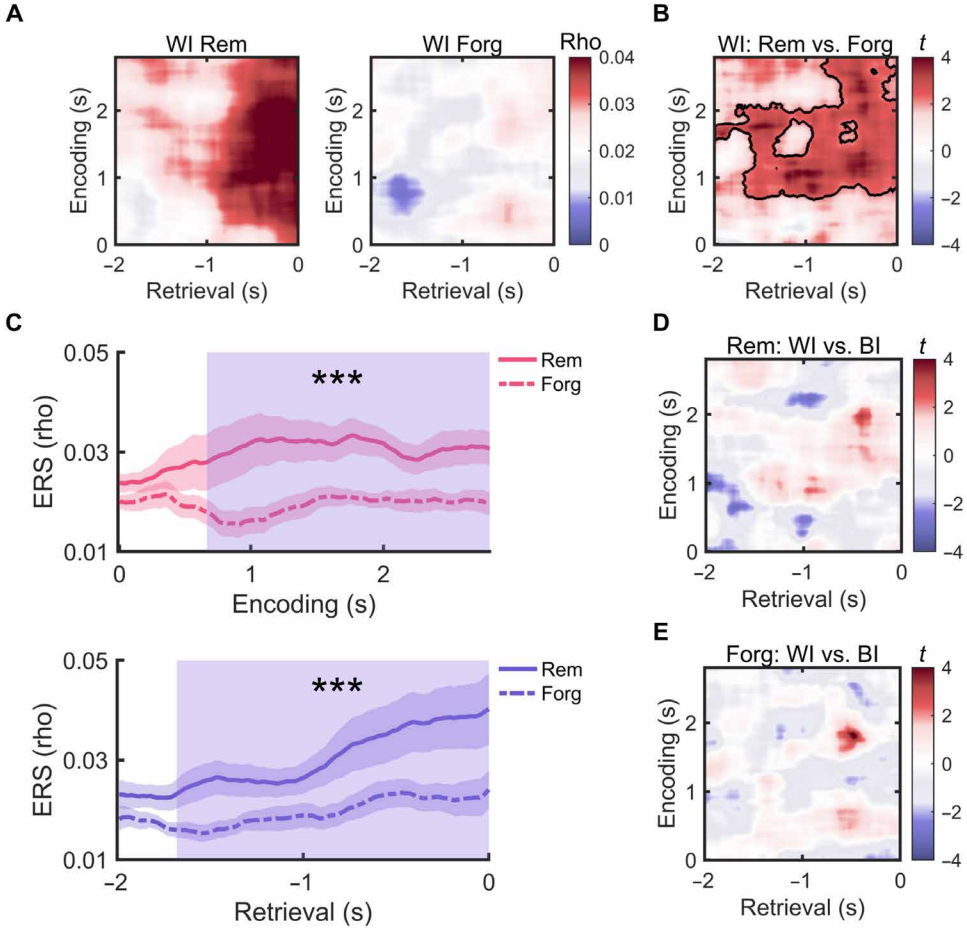

Fig. 4. Lack of item-specific encoding-retrieval similarity (ERS). (**A**) WI encoding-retrieval similarity (WI ERS) for remembered (left) and forgotten items (right), respectively. (**B**) Difference between WI ERS values for remembered versus forgotten items. The black line circles the significant cluster ($P_{corr} < 0.001$). (**C**) WI ERS across encoding (top) and retrieval (bottom) time periods for remembered and forgotten items. Blue rectangular areas indicate time periods where WI ERS of remembered and forgotten items differ. The shaded areas around the lines indicate 1 SEM. No significant clusters showed greater WI than BI ERS, for either remembered (**D**) or forgotten (**E**) items. ***$P_{corr} < 0.001$.

MRS was computed by contrasting WI versus BI MRS values. This was done in each retrieval time window, separately for remembered and forgotten items. We found a cluster with significantly greater WI than BI MRS values for remembered items [210- to 620-ms pre-retrieval response; $t(15) = 3.860$, $P_{corr} = 0.011$; Fig. 5A] and a similar cluster showing an opposite effect for forgotten items [180- to 630-ms pre-retrieval response; $t(15) = -2.729$, $P_{corr} = 0.022$; Fig. 5B]. A two-way analysis of variance (ANOVA) with "item specificity" (WI versus BI) and "memory" (remembered versus forgotten) as repeated measures revealed a cluster that showed both a significant interaction [150- to 840-ms pre-retrieval response; $F(1,15) = 23.933$, $P_{corr} = 0.003$] and a main effect of memory [$F(1,15) = 4.725$, $P = 0.046$] (Fig. 5C). Post hoc analyses in this cluster showed significantly greater WI than BI MRS for remembered items [$t(15) = 3.115$, $P = 0.007$] (Fig. 5D) while an opposite effect was found for forgotten items [$t(15) = -2.624$, $P = 0.019$]. We also found significantly greater WI MRS for remembered versus forgotten items [$t(15) = 3.097$, $P = 0.007$]. When analyzing these effects separately in the individual brain regions, the item-specific MRS did not reach significance in any region (fig. S16). This may reflect the more distributed nature of memory representations during short-term maintenance (*51*). These results indicate that item-specific representations during short-term memory maintenance were reinstated during successful, but not during unsuccessful long-term memory retrieval.

The presence of item-specific MRS and the lack of item-specific ERS for remembered items indicate that representations during retrieval did contain item-specific information, but their representational formats were substantially transformed from encoding. To support this idea, we further compared the neural representations in the retrieval time windows with the four types of representational formats (i.e., early visual, late visual, semantic, and abstract semantic representations), separately for remembered and forgotten items. Similar to the encoding period, this analysis only included subjects with more than 20 items in each condition, resulting in 13 subjects in the remembered condition and 12 subjects in the forgotten condition. We found that the neural representations of remembered items were correlated with abstract semantic representations (550- to 890-ms pre-retrieval response; $P_{corr} = 0.039$), but not with the other three representational formats (all $P_{corr} > 0.226$) (Fig. 5E). No significant correlations were found for forgotten items (all $P_{corr} > 0.165$; Fig. 5F). Consistent with the whole-brain results, the neural representations in the LTL also contained abstract semantic

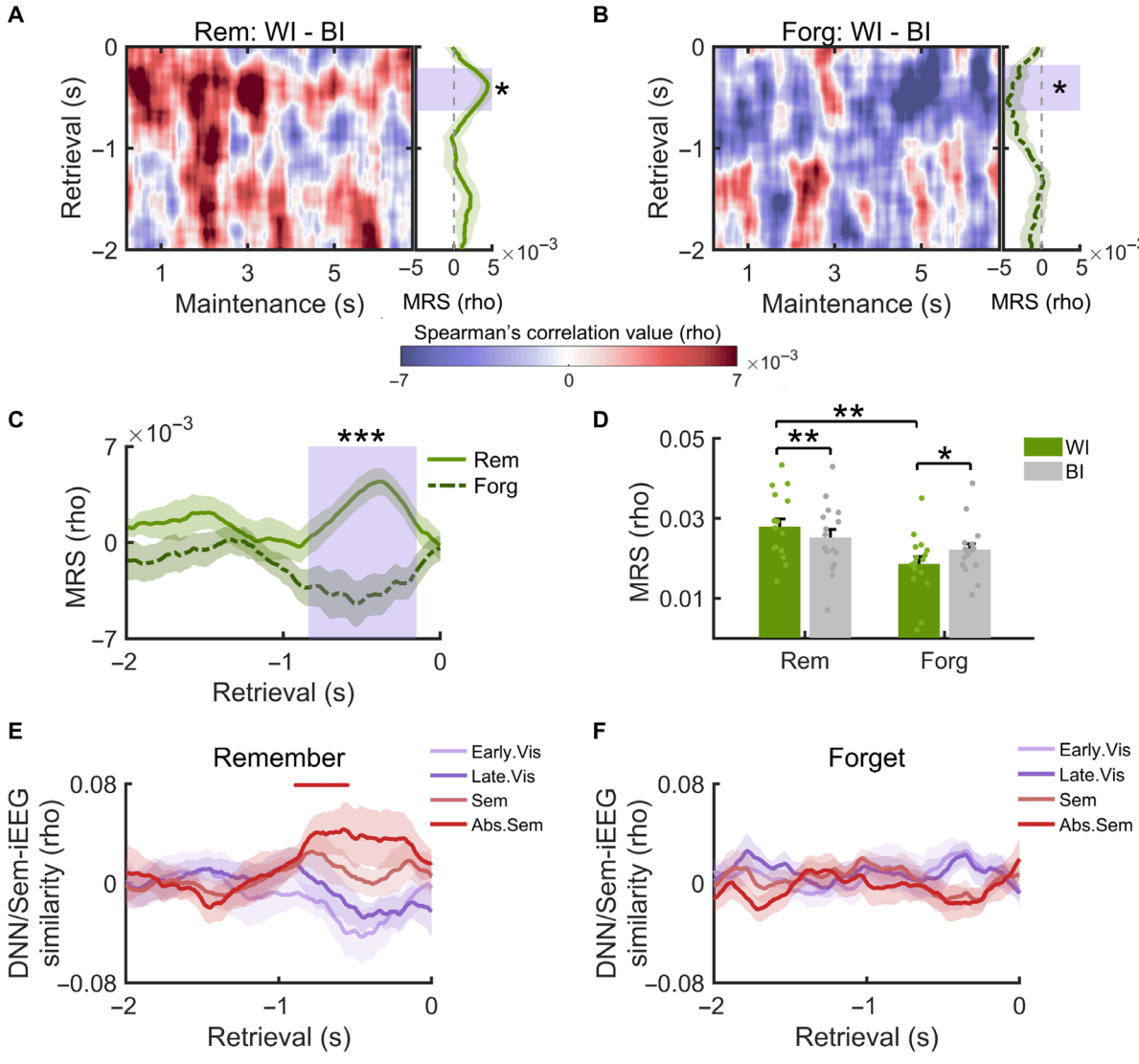

Fig. 5. Reinstatement of short-term memory representations during long-term memory retrieval. (**A**) Representational reinstatement of maintenance-related activity patterns during retrieval for remembered items (MRS). The line plots on the right show averaged item specificity values across the maintenance period for individual retrieval time windows. The shaded area shows a retrieval time window with significant reinstatement of item-specific representations for remembered items. (**B**) MRS for forgotten items. Here, the shaded area shows a time window with a significantly lower WI than BI MRS. (**C**) Comparison of item-specific MRS for remembered versus forgotten items. A significant cluster between 150- and 840-ms preretrieval response showed greater item-specific MRS for remembered than forgotten items (blue shaded rectangular area). (**D**) Averaged MRS values within the cluster found in (C). Post hoc comparisons indicate greater WI than BI similarities for remembered items, but an opposite effect for forgotten items. The result also shows greater WI MRS for remembered than forgotten items. Dots indicate MRS values of individual participants. (**E**) Representational formats of remembered items during retrieval. The colored horizontal bar indicates the time window showing a significant match with the representational format in the DNN/semantic model. (**F**) Representational formats of forgotten items during retrieval. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

information but not visual information during successful retrieval (fig. S17). Moreover, semantic and abstract semantic representations in the LTL were significantly more pronounced than early visual representations (fig. S17). These results indicate that semantic and abstract semantic representations were reinstated during successful retrieval.

Representational transformation from encoding to maintenance and retrieval

The above results revealed substantial representational transformations from encoding to retrieval. A further question concerns when this transformation occurred. To this end, we systematically examined the transformation of neural representations across different memory stages. We hypothesized that the representations went through multiples stages of transformation from encoding to maintenance and retrieval, including a transformation that occurred immediately after encoding, which may underlie fast memory consolidation (*28*). Hence, we predicted that, for remembered items, (i) there was a representational change right after stimulus offset, (ii) the item specificity of EES was greater than the item specificity of both EMS and ERS, and (iii) the item specificity of MRS was greater than the item specificity of ERS and EMS.

To test the first hypothesis, we calculated the *di* during the post-encoding maintenance period. This analysis showed a greater *di* for

remembered than forgotten items during the first 350 ms immediately following stimulus offset (P_{corr} = 0.043; Fig. 6, A and B). In a further analysis, we examined the neural pattern similarities between periods immediately before and after stimulus offset. A fast transformation of representations immediately after stimulus offset would imply a decrease in representational similarity between these periods. We used the representation in the last 10 encoding time windows as the benchmark (indicated by the green horizontal bar in Fig. 6C) and correlated it with the representations in different time periods before and after stimulus offset. The results revealed a significant decrease of similarity within the first second after stimulus offset (i.e., at the beginning of the maintenance period) for subsequently remembered items [$t(15) = -2.421$, $P = 0.029$], but not for forgotten items [$t(15) = -1.027$, $P = 0.321$] (Fig. 6C). Linear fits revealed a significantly greater decrease of similarity for subsequently remembered than forgotten items after stimulus offset [$t(15) = -2.277$, $P = 0.038$]. This decrease was also higher than that for both remembered and forgotten items before stimulus offset (both $Ps < 0.048$) (Fig. 6D). These results converge to suggest a prominent representational transformation for subsequently remembered items right after stimulus offset.

To test the second and third hypotheses, we first performed an analysis of encoding-maintenance similarity (EMS), correlating activities in each maintenance time window with activity in individual encoding time windows. The EMS was averaged across all maintenance time windows, and item-specific EMS was examined by contrasting WI versus BI EMS in individual encoding windows separately for remembered and forgotten items. This analysis revealed no significant item-specific EMS for remembered items (P_{corr} = 0.206) but a cluster with significantly greater WI than BI EMS (1180 to 1470 ms, P_{corr} = 0.043) for forgotten items (fig. S18), indicating that subsequently remembered items underwent substantial transformations from encoding to maintenance.

We then compared the item specificity of EES with the item specificity of EMS and ERS, respectively. Since the time periods in EES, EMS, and ERS analyses are of different lengths, we used two strategies to perform these comparisons. In the first strategy, we defined an equally sized temporal cluster around the center point of the cluster, which showed the largest item specificity (i.e., in which the sum of *t* values was the largest when comparing WI versus BI). We systematically varied the size of the cluster from 0.5 to 1.5 s with steps of 0.2 s. This analysis showed that the item specificity of EES was significantly greater than the item specificity of EMS in clusters with sizes between 700 and 1300 ms [500 ms: P_{FDR} = 0.063; 700 to 1300 ms: P_{FDR} = 0.0497; 1500 ms: P_{FDR} = 0.058; Fig. 6E; all tests are false discovery rate (FDR)–corrected for multiple comparisons]. We also found that the item specificity of EES was significantly greater than the item specificity of ERS in clusters with sizes above 500 ms (500 ms: P_{FDR} = 0.053; 700 to 1500 ms: P_{FDR} = 0.044; Fig. 6E). The second strategy was to select the time point showing the highest item specificity (time points with largest WI versus BI effects). We systematically selected the top 5, 10, 15, 20, and 25% of time points with the largest item specificity for EES, EMS, and ERS, respectively. Similar results were found, with significantly greater item specificity of EES than EMS (from 5 to 25%, P_{FDR} = 0.04) and marginally greater item specificity of EES than ERS (from 5 to 25%, all P_{FDR} = 0.06) (Fig. 6F).

Using the abovementioned cluster-based strategy, we found that the item specificity of MRS was significantly greater than the item specificity of ERS in clusters with sizes of 900 ms (P_{FDR} = 0.048) and 1100 ms (P_{FDR} = 0.048) (Fig. 6E) and by trend in clusters of other sizes (700 ms: P_{FDR} = 0.079; 1300 ms: P_{FDR} = 0.079). Similar results were found when including the time points with the greatest item specificity in the range from the top 5 to 25% (P_{FDR} = 0.049 for all five types) (Fig. 6F). Last, the item specificity of MRS was numerically although not significantly greater than the item specificity of EMS ($P_{FDR} > 0.125$), suggesting a trend of greater representational transformations from encoding to maintenance than from maintenance to retrieval.

In summary, the above results revealed a significant representational transformation immediately after encoding. In addition, the greater within- than across-stage similarity, in combination with the greater item specificity of MRS than ERS, further suggest that neural representations went through multiple stages of transformation from encoding to short-term memory maintenance and long-term memory retrieval.

DISCUSSION

Episodic memory has long been conceived as a dynamic process, but the exact nature of neural representations across memory stages and the processes that underlie the dynamic transformations of these representations have just begun to be elucidated. The present study systematically compared the neural representations across memory stages including encoding, short-term maintenance, and long-term retrieval. Our results revealed substantial representational transformations during memory encoding that involve both visual and semantic representations and showed that greater transformations predicted subsequent long-term memory. This effect is partially mediated by enhanced representational fidelity during encoding. Moreover, the encoded representations continued to be transformed across memory stages from short-term memory to long-term memory. These results clearly illustrate the transformative nature of human episodic memory.

Previous work has found that visual objects are dynamically processed across space and time. On the one hand, visual inputs are progressively processed along the hierarchical structure of the ventral visual stream within the first few hundred milliseconds, during which visual representations are transformed from low-level visual to superordinate categorical features (*19*, *52*). On the other hand, the representations also change dynamically across time within single regions. For example, a non-human primate study showed that V1 neurons respond to visual orientation bars during early encoding and exhibit effects of contour integration 50 to 100 ms later that likely reflect feedback from V4 (*21*). Similar evidence has been found in an intracranial electroencephalogram (iEEG) study, which showed that fusiform face area (FFA) first identifies face-specific information at ~50 to 75 ms after the onset of face images and then forms invariant face identities at 200 to 500 ms, which might result from recurrent top-down and bottom-up interactions (*53*).

In the current study, we observed that neural representations changed dynamically across the extended stimulus encoding period (with a length of 3 s). These representations corresponded to early visual formats 340 to 590 ms after stimulus onset, to higher-order visual formats between 290 and 650 ms, to semantic formats between 370 and 1890 ms, and to abstract semantic formats between 530 and 1330 ms. Several previous studies have shown early visual representations within the first 200 ms of encoding, followed by

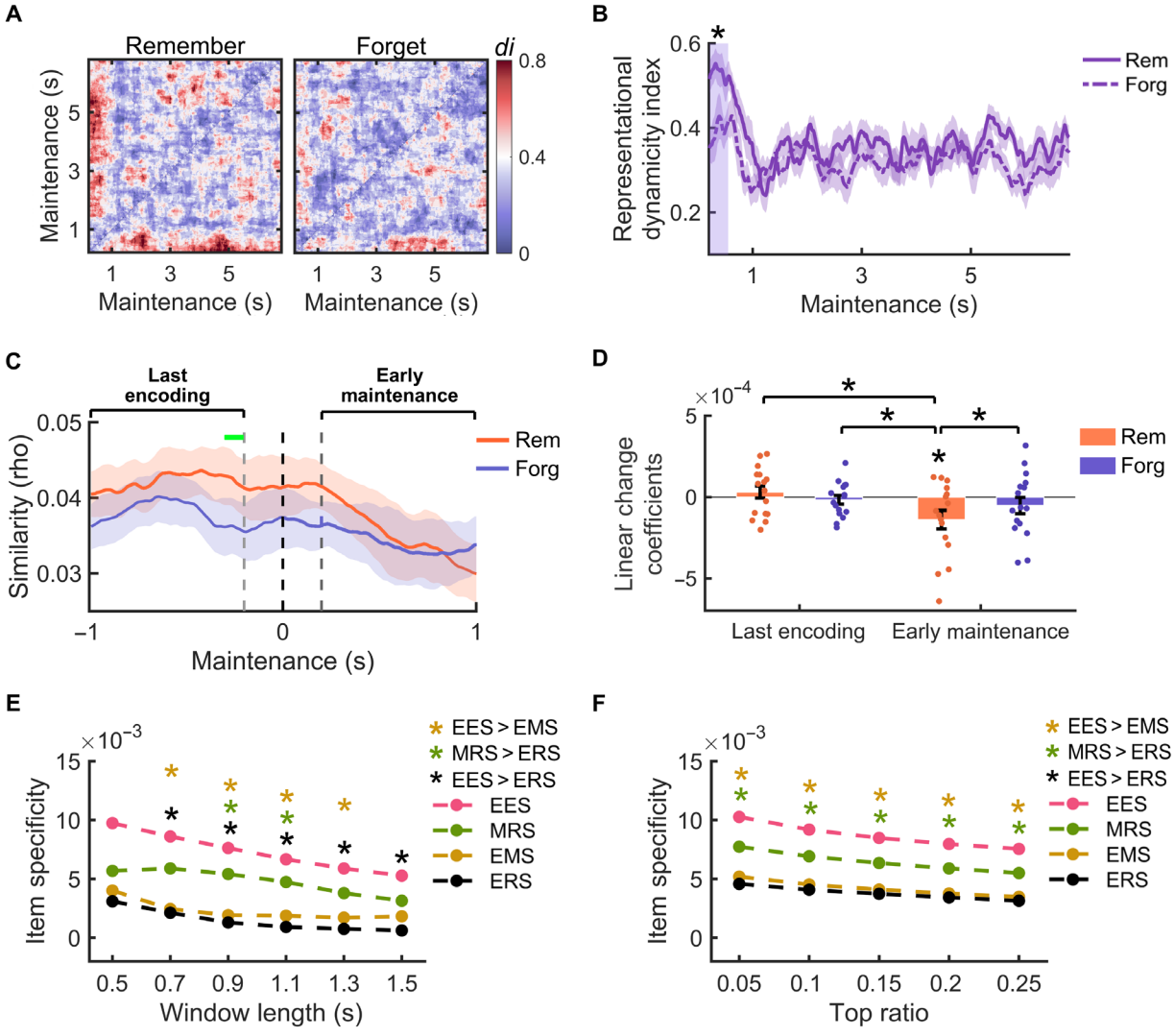

Fig. 6. Multistage transformation of representations from encoding to maintenance and retrieval. (**A**) Dynamicity index (*di*) map during maintenance for subsequently remembered and forgotten items. (**B**) Averaged *di* across the two maintenance temporal dimensions in (A). A greater dynamicity for subsequently remembered than forgotten items was found in the first 350 ms of the maintenance period (indicated by the shared area). (**C**) Correlation of the neural activity in the last 10 encoding time windows (central points of these time windows are indicated by the green horizontal bar) with neural activity in the last second of encoding and the first second of maintenance (i.e., early maintenance), separately for subsequently remembered and forgotten items. The black dashed line indicates the time point of stimulus offset. The time interval between the two gray lines reflects the sliding time windows with mixed time points from both encoding and maintenance periods, which were not included in the statistical analysis. A linear fit was applied to the correlation values during the last encoding interval and in the early maintenance interval with equal length (indicated by the horizontal black lines), respectively. (**D**) Comparisons of the coefficients of linear fit during the last second of encoding interval and the first second of maintenance. Dots indicate representational changing slopes of individual participants. (**E**) Comparisons between item specificity of EES, EMS, and ERS and between item specificity of MRS and ERS, within clusters of the same size. (**F**) Same comparisons when including identical percentages of top indexes showing greatest item specificity. Error bars indicate 1 SEM. *$P < 0.05$.

higher-order visual and semantic representations (*19*, *54*). The observed visual representations in our study occurred much later than in the abovementioned studies and did not differ from the timing of semantic representations. This may result from the absence of electrodes in the primary visual cortex, the fact that we examined oscillatory activities rather than spikes, and the sliding window approach, which may have further reduced temporal resolution. Nevertheless, our results extend previous studies on early visual processing by suggesting that neural representations experience temporally extended transformation periods and engage more complex representational formats such as abstract semantic representations.

Critically, the current study demonstrates that greater representational transformations during encoding are associated with better subsequent long-term memory performance. This representational transformation is characterized by the fast and prominent emergence of abstract semantic representations for subsequently remembered items during encoding. Compared to subsequently forgotten items, subsequently remembered items contained more semantic and abstract semantic information but less visual information. This result should not be simply attributed to the limited coverage of the early visual cortex, as we found pronounced late visual representations in both LTL and MTL for forgotten items, showing that, in principle,

our electrode coverage allowed us to detect this representational format. Furthermore, this result was not merely due to the task requirements and encoding strategy: Although we only asked participants to make category judgments during long-term memory retrieval, after encoding, they were first tested in a short-term memory task, which required access to fine visual details. Participants performed well in this task (accuracy: 0.90 ± 0.042), and their performance did not differ between subsequently remembered and forgotten items ($P > 0.18$). Thus, visual information was available to the participants for later remembered items as well. Corroborating the role of semantic processing in effective encoding, existing studies have shown that the semantic similarity among the study materials predicts later long-term memory (*55*, *56*), and a semantic encoding task (*57*) and semantical elaboration (*58*) could facilitate memory encoding. Nevertheless, our data cannot exclude the possibility that the visual representations in primary visual cortex could contribute to subsequent memory, particularly when a perceptual memory test was used (*59*).

Our results also fit very well with the idea that memory entails effective interactions between external inputs and preexisting concepts (*8*, *34*, *60*). Both rodent (*61*) and human studies (*62*) have shown that new information that is consistent with existing knowledge structures can be assimilated into cortical networks via a fast consolidation mechanism. Beyond this mechanism, we posit that long-term knowledge could also affect memory formation via a dynamic encoding process. During this process, sensory inputs are interpreted and comprehended, and perceptual activity patterns are transformed into internal representations that can be integrated with preexisting knowledge (*21*, *22*, *53*). In particular, the interaction between perceptual information and prior knowledge during these processes shapes the neural representations of perceived features across time (*63*, *64*). For example, one study has shown that prior knowledge facilitates new encoding by involving additional semantic and associative-binding processes (*65*).

We further showed that this effect is partially mediated by the fidelity of neural representation, extending previous results that show that greater fidelity of neural representations during a late encoding time window is associated with subsequent memory (*25*, *26*). These results indicate that the extraction of semantic and abstract semantic features during encoding benefits long-term memory via establishing more reliable item-specific representations across repetitions during a late encoding period. In addition, recent studies have shown that abstract semantic representations that occur relatively late after the stimulus presentation are more stably maintained during short-term memory maintenance (*20*) and more faithfully reinstated during long-term memory (*46*, *50*) than lower-level perceptual representations. Together, our results suggest that dynamic encoding through interaction with prior knowledge helps to form stable representations, which can be better integrated with existing long-term knowledge and more faithfully maintained and reinstated during long-term memory retrieval.

Beyond dynamic encoding, episodic memory also experiences post-encoding consolidation, during which neural representations are further transformed (*27*, *34*). Previous studies have shown that item-specific representations during memory retrieval depend primarily on frontoparietal cortices (*9*, *10*), which represent more abstract information (*66*). In the current study, we provide direct neural evidence to show that memory retrieval involves strong abstract semantic representations but lacks perceptual representations. We further show a lack of item-specific ERS across the whole brain as well as in several individual brain regions. Again, the absence of visual representations during retrieval and of item-specific ERS could not be merely attributed to the retrieval task, the limited spatial coverage of electrodes, or our analysis approach, because we did find significant late visual representations during encoding of forgotten items, significant item-specific representations during encoding (EES) and maintenance (MMS), and item-specific MRS. Corroborating a previous fMRI study using a similar paradigm, which found significant item-specific EES in the visual cortex and item-specific retrieval-retrieval similarity (RRS) in the parietal lobe, but no item-specific ERS (*10*), the current results suggest weak reactivation of visual representations and pronounced representational transformations from encoding to retrieval.

Several possibilities might account for the discrepancies between encoding-retrieval similarities found in previous work (*4*, *5*) and the substantial transformations observed in the current study. First, several previous studies using multivariate decoding or RSA either examined the neural pattern reinstatement at the category level (*67*, *68*) or did not test item-specific representational reinstatement (*69*). Second, several studies tested long-term memory via recognition memory tests, in which item-specific pattern similarity may be introduced by the common perceptual inputs (*4*, *47*). Third, some other studies used a cued-recall task and also found item-specific ERS (*46*). However, in these studies, words were used as materials, and the neural reinstatement was found in higher-order brain regions (e.g., hippocampus and anterior temporal lobe) where visually invariant, semantic representations might be shared between memory encoding and retrieval.

Our study found that transformations occur much faster than previously thought. Specifically, we found that the neural representations were substantially transformed in the first second of the post-encoding maintenance period. Corroborating our results, it has been shown that classifiers trained on perceptual data could better classify perceptual than imagery data (*70*), and an "imagery" classifier outperformed a "perceptual" classifier in classifying imagery data (*71*). Using a visual DNN, a recent study revealed that higher-order complex visual formats rather than early visual formats were shared between perception and visual imaginary (*72*). Our results showed that greater transformations immediately after encoding were associated with better long-term memory, consistent with the idea that this immediate post-encoding dynamics might reflect early memory consolidation processes that register information into long-term memory (*31*).

We further found that neural representations during retrieval were more similar to those during short-term maintenance than to representations during encoding. Consistently, an iEEG study combining encoding and off-line consolidation periods showed that late encoding components that were replayed during consolidation were also reinstated during successful long-term memory retrieval (*50*). Several reasons might contribute to the higher MRS than ERS in our study. First, we found that representations were transformed across memory stages and that short-term memory maintenance bridged encoding and retrieval. Second, both successful retrieval and short-term memory maintenance rely on internal representations from top-down processing. In particular, short-term memory reflects the temporary activation of long-term memory representation (*73*). In contrast, encoding involves both bottom-up and top-down processing (*74*). Consistently, studies have found shared

capacity limitations of visual short-term memory and visual information recall (*75*). Third, emerging evidence has shown that the same functional regions (i.e., those in frontoparietal cortices) are preferentially recruited during both short-term maintenance and long-term memory retrieval (*9*, *49*).

The representational transformation after encoding might result from memory reactivation, which has often been found during post-encoding waking periods (*76*, *77*). According to the theory of memory reconsolidation, whenever a memory trace is reactivated, it turns into a fragile state that is prone to undergo modifications (*78*). Recent studies have shown that this reactivation during waking state is prominently shaped by long-term knowledge, coreactivated competition of related memory representations, internal context information, goals, and reward history, which could result in memory strengthening, weakening, integration, and/or differentiation (*65*, *77*, *79*, *80*). Presumably, reactivation during rest and sleep is more strongly influenced by long-term knowledge than reactivation during waking state, due to the lack of external sensory input or cognitive control processes (*27*). Through this active and interactive process of reactivation and reconsolidation, memory representations are continuously transformed.

Last, memory representation could be further transformed during memory retrieval. Several previous studies have investigated the retrieval of information in different representational formats and found that the temporal order of these formats reversed from encoding to retrieval (*13*, *81*), suggesting a constructive process of memory retrieval. Compared to repeated learning, repeated retrieval practice not only strengthens target memory (*82*) but also transforms memory representations, resulting in greater reliance on the frontoparietal region than on the visual cortex (*80*) and greater effects in the differentiation of competing memory representations (*80*, *83*).

This transformative perspective of memory and the associated analytical tools in the current study could be further leveraged to advance our understanding of human episodic memory. First, the observed temporal change of representational formats could be accompanied by a representational transformation across brain regions (*84*). Future studies using magnetoencephalography with full coverage over various cortical regions could help examine in detail the time course of representational transformations within and across brain regions. Second, future studies should further examine how the nature of stimuli and task requirements could modulate the representational transformation and their effect on memory formation and retrieval. Third, our methods could be readily extended to examine how memory representations transform over days, months, and years, and how sleep transforms these representations. Last, future studies could examine how specific schemas shape dynamic encoding, memory consolidation, and representational transformations.

To conclude, our results provide important empirical evidence to emphasize the dynamic nature of human episodic memory. These transformative processes occur between multiple stages of memory processing, through interaction with existing long-term knowledge, and strongly influence memory performance. A better understanding of this dynamic nature of human episodic memory might not only help us understand how memory is constructed and reconstructed but also hold the potential to understand the role of episodic memory in supporting other cognitive functions, such as problem solving, creative thinking, and decision-making (*34*).

MATERIALS AND METHODS

Participants

A total of 16 patients with drug-resistant epilepsy who were implanted with stereotactic electrodes participated in the study (mean age ± SD: 27.13 ± 6.78 years, seven females). The implantation scheme, including the number and placement of electrodes, was exclusively based on diagnostic purposes and was unique for each patient. The study protocol was approved by the Institutional Research Board Committee of Xuanwu Hospital, Capital Medical University, Beijing, China, following the ethical standards of the Declaration of Helsinki. All patients had a normal or corrected-to-normal vision. Written informed consent was obtained from all patients.

Experimental design

Fifty-six pictures and 112 two-character Chinese verbs were used in this study. These objects were selected from four categories, including animals, electronics, fruits, and furniture, with 14 pictures in each category. Each picture was randomly paired with two different cue words, resulting in 112 word-picture associations overall. The word-picture associations that shared the same pictures were assigned to two consecutive runs.

The experiment consisted of a short-term memory and a long-term memory phase (Fig. 1A). During the short-term memory task—i.e., a DMS task—participants were asked to learn and maintain the word-picture associations. Each association was repeated three times, with an interrepetition interval between 5 and 10 trials to optimize learning (*25*). Each trial started with a fixation cross (300 ms), followed by a blank screen (800 to 1200 ms). A word-picture association was then presented on the screen for 3 s. This was followed by a 7-s delay period, during which only the cue word appeared on the screen, and participants were asked to maintain the associated picture as vividly as possible. Immediately after the maintenance period, a probe was presented, which was either identical (target, 50% of trials) or highly similar (lure) with the previously presented picture. Participants indicate via button press whether the probe was the target picture. Each run of the short-term memory task contained 14 unique associations and lasted for about 10 min.

After a 1-min countback task and a short break (1 to 4 min), the 14 associations were tested in a randomized order (without repetitions). A cued-recall paradigm was applied to the test phase. In each test trial, after a fixation cross (300 ms) and a blank screen (jittered between 800 to 1200 ms), a word cue was presented. Participants were asked to retrieve the associated picture and indicate their memory by pressing different keys for "remember" or "don't remember" responses. We encouraged participants to retrieve the pictures as vividly as possible. If a "remember" response was made, they were then asked to report the category of the retrieved picture within 3 s by pressing one of four buttons corresponding to animals, electronics, fruits, and furniture, respectively. If a "don't remember" response was made during the retrieval phase, the test would proceed to the next trial. Trials that were correct in the category report were coded as remembered, and trials associated with "don't remember" responses or incorrect category reports were coded as forgotten trials.

Data recordings and preprocessing

Intracranial EEG data were recorded in the Xuanwu Hospital, Beijing, China. Semirigid platinum or iridium depth electrodes were

implanted. There were 8, 12, or 16 contacts per electrode, with each contact 2 mm in length, 0.8 mm in diameter, and 1.5 mm apart. Data were sampled at 2500, 2000, or 2048 Hz on different recording systems, including Brain Products (Brain Products GmbH, Germany), NeuroScan (Compumedics Limited, Australia), and the Nicolet EEG system (Alliance Biomedica Pvt. Ltd., India), respectively. The online recording signals in all channels were referenced to a common contact placed subcutaneously on the scalp, which was recorded simultaneously. During the offline preprocess, data were then rereferenced to the common average of data across all clean channels. To eliminate 50-Hz line noise and its harmonic wave signal, a fourth-order Butterworth notch filter centered on the noise frequency with a stopband of 2 Hz was applied to the data. Channels located in epileptic loci and those severely contaminated by epileptic spikes were excluded from the analysis. Data epochs containing interictal spikes were identified by visual inspection and removed from the analysis as well. All these analyses were performed in EEGlab (www.sccn.ucsd.edu/eeglab/) and the Fieldtrip toolbox implemented in Matlab (MathWorks Inc.), using custom code written in Matlab.

Time-frequency analysis

Intracranial EEG data during the short-term memory task were first epoched into 16 s from 3 s before to 13 s after stimulus onset. Data during long-term memory retrieval were epoched into 9 s, from 6 s before to 3 s after retrieval responses. The time period of interest for the short-term memory was from 0 to 10 s relative to the presentation of the stimulus, with the encoding period from 0 to 3 s and the maintenance period from 3 to 10 s. During retrieval, the time period of interest was from 0 to 2 s before the retrieval response. The extended epochs before and after the time periods of interest account for edge effects resulting from time-frequency transformation, which were removed from subsequent analysis.

The epoched data were convolved with complex Morlet wavelets (six cycles) in the range of 2 to 120 Hz, with 1 Hz as a step. The spectral power was then obtained by squaring the magnitudes of the complex wavelet transform. The power was z-transformed separately for each frequency and each channel by using the mean and the SD of the power across the task period within a run. Note that this normalization process was done separately for the short-term memory task and the long-term memory task in each run. All spectral power data were subsequently downsampled to 100 Hz. Both the encoding and successful retrieval of visual objects elicited spectral power changes across a broad frequency range (2 to 120 Hz) (fig. S1), and the spectral power changes in these frequency bands showed item-specific tuning (fig. S2). Therefore, the spectral power in this broad range was used as the feature for the subsequent RSAs.

Electrode localization

High-resolution structural magnetic resonance images (3.0 T, Siemens) and computed tomography (CT) scans (Siemens) were acquired before and after the implantation of electrodes, respectively. The cortical surface was reconstructed, segmented, and parcellated on the basis of MRI data using the default Deskian Killiany atlas in the Freesurfer (https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/AnatomicalROI). We thus obtained coordinates and labels of individual anatomical regions for each participant. To identify the location of contacts, we coregistered postimplantation CT to preimplantation MRIs in Statistical Parametric Mapping (SPM12); https://www.fil.ion.ucl.ac.uk/spm/software/spm12/. The coordinates of individual contacts were acquired from the coregistered image in the 3D slicer (www.slicer.org/). The region in which individual contacts were located was then obtained by mapping the contacts to the closest anatomic brain area. All clean contacts across participants were projected to a standard Montreal Neurological Institute (MNI) space and plotted for visualization in the 3D slicer. Across participants, there were overall 592 clean channels (means ± SD, 37.0 ± 12.98), which were widely distributed across brain regions including the LTL, the MTL, the frontal lobe, and the parietal lobe (Fig. 1C).

Representational similarity analysis

Global RSA was performed between trials by correlating the spectral power across frequencies (2 to 120 Hz, with 1-Hz steps from 2 to 29 Hz and a 5-Hz step from 30 to 120 Hz) and across all clean channels, separately for each time window, using Spearman's correlation (*6*, *47*). To obtain representational patterns across time, RSA was computed in sliding time windows with lengths of 400 ms and increments of 10 ms.

To examine the regional specificity, the RSA was additionally performed in individual brain regions. We defined three regions of interest that contribute to episodic memory processes, including the LTL (*47*), the MTL (*48*), and the FP (*9*, *10*). In particular, channels in the fusiform gyrus, the inferior temporal cortex, the middle temporal cortex, and the superior temporal cortex were grouped into the LTL, with overall 185 clean channels across 15 participants (12.33 ± 9.08 channels per patient; range, 2 to 27). Channels in the hippocampus, the amygdala, and the parahippocampal cortex were grouped into the MTL, with overall 116 clean channels from 16 participants (7.25 ± 5.42 channels per patient; range, 1 to 22). Channels in the frontal lobe and the parietal lobe were grouped into the FP, with overall 159 clean channels from 14 participants (11.36 ± 6.92 channels per patient; range, 1 to 25) (for more details, see table S1). For each of these regions of interest, RSA was again calculated across frequencies and all clean channels in the respective region.

The current study examined neural representations within and across three memory stages, including encoding, maintenance (short-term memory), and retrieval (long-term memory). Thus, RSA was performed both within the same memory stage and between different memory stages. This resulted in two types of RSA within memory stages, EES and MMS, and three types of RSA across memory stages, EMS, MRS, and ERS (Fig. 1D). Note that the absence of retrieval-retrieval similarity was due to the fact that each word-picture association was only tested once during retrieval. All RSAs were performed across repetitions within the same experimental run, either within or across different memory stages. The WI similarity was defined as the similarity between trials that shared the same picture, while BI similarity was the similarity between trials with different pictures.

Correlating neural representations with visual and semantic models

Representational formats were examined by correlating the neural representations with visual and semantic representations obtained from a visual DNN, "AlexNet" (*42*), and a well-established semantic model, Directional Skip-Gram (*44*), respectively. The AlexNet consists of eight layers, five convolutional layers and three fully connected layers, which simulate the hierarchical structure of neurons along the ventral visual stream. In general, the early layers of the AlexNet process early visual features, including colors, contrasts, and frequencies, while the deeper layers process higher-order

visual representations, e.g., the surface structure of objects or body parts of animals. A well pretrained AlexNet via using the ImageNet (*85*) dataset was used in this study. Here, we applied the AlexNet to classify individual objects and then extracted the features from each layer of the AlexNet. We measured the similarity between the features of every two pictures via Spearman's correlations, resulting in a BI similarity matrix in each layer. All the similarity matrices were then Fisher Z-transformed before further analysis. To simplify the visual representations obtained from the DNN model, the visual similarity matrices in the first five DNN layers ($r > 0.675$, $P < 0.001$) were averaged and labeled as "early visual similarity," and the three fully connected layers ($r > 0.796$, $P < 0.001$) were averaged and labeled as "late visual similarity."

To ensure that our data did not depend on a specific visual DNN, we recruited another two DNN models, the VGG19 and GoogLeNet, which have been frequently used for understanding the neural representations during visual imagery and perception in both human (*72*) and non-human primates (*86*). The VGG19 consists of 16 convolutional layers and 3 fully connected layers. Following a previous study (*72*) and to facilitate the comparison with the five convolutional layers in the AlexNet, we simplified the model by averaging the first two convolutional layers as "grouped convolutional layer 1"; layers 3 and 4 as "grouped convolutional layer 2"; layers 5 to 8 as "grouped convolutional layer 3"; layers 9 to 12 as "grouped convolutional layer 4"; and layers 13 to 16 as "grouped convolutional layer 5." In addition to these five grouped convolutional layers, the three fully connected layers were kept separately, resulting in eight layers. The GoogLeNet consists of 22 layers, including two convolutional layers and nine inception modules. We extracted the image feature from the two convolutional layers, the output of each inception module, and the one last fully connected layer, resulting in 12 layers in total. The similarity matrices in all models and at all layers were created on the basis of the pairwise correlations of the artificial neural activities that were elicited by the stimuli used in our study.

Similar to the AlexNet model, we further simplified the DNN models by averaging the similarity matrices across the five grouped convolutional layers in the VGG19 and across the first nine grouped layers of GoogLeNet, respectively. These were termed the "early visual similarity matrix" of these DNNs. We also averaged the similarity matrices of the three fully connected layers of VGG19 and the last three layers of GoogLeNet and termed them "late visual similarity matrix," respectively.

The semantic similarity matrix was obtained by correlating semantic features between the labels of every two pictures. The labels were generated by five independent raters who did not participate in the experiment, and the most frequently generated label was selected for each picture (see exemplar labels of pictures in table S2). The Chinese word2vector (*44*) semantic model converted the label of each picture into a vector of semantic features, consisting of 200 values. Each value in the word vector indicates the meaning of a picture label in one semantic dimension, such as gender, animacy, etc. The semantic similarity matrices between items were obtained by calculating the cosine similarity of these word vectors. The semantic representational similarities correlated with perceptual similarities, suggesting that the word2vector model reflected representations both at the level of a sensory-derived cognitive system and at the level of a cognition-derived semantic system, which have been shown to coexist in the human brain (*45*). To disentangle these different representational formats, we regressed out early and late visual representations (DNN layers 1 to 5 and 6 to 8, respectively) from the semantic similarity matrix, which generated a matrix reflecting an "abstract semantic format" (fig. S4).

The neural similarity matrix was obtained by correlating neural activities of all pairs of pictures across frequencies and channels, using Spearman's correlations. A sliding time window of 400 ms, with a step size of 10 ms, was used to obtain the representational format across encoding or retrieval periods. Note that in this analysis, spectral power data were first averaged across repetitions of the same pictures. To remove potential confounds of commonly evoked power by stimulus onset, we further normalized the power spectral data across trials during each time window for each frequency and each channel during the encoding period.

Spearman's correlations were conducted to link neural representations to the different types of visual and semantic representations (i.e., early visual, late visual, semantic. and abstract semantic formats), separately for each time window. All Spearman's correlation values were then Fisher Z-transformed and tested against zero at the group level. The cluster-based permutation test (see below) was used to determine statistical significance.

Analysis of representational dynamics during memory encoding

Previous work suggests that if neural representations undergo dynamic transformations from one time point (t_1) to another time point (t_2), the pattern similarity between different time points $r(t_1, t_2)$ should be significantly lower compared with that at the same time point [e.g., $r(t_1, t_1)$, $r(t_2, t_2)$] (*41*). Notably, the similarity should be computed between repetitions of the same item to make this comparison nontrivial. Therefore, to quantify representational dynamics between t_1 and t_2, we computed the *di* according to the following equation

$$di(t_1, t_2) = \begin{cases} 1, \text{if } r(t_1, t_2) < r(t_1, t_1) \cap r(t_1, t_2) < r(t_2, t_2) \\ 0, \text{otherwise} \end{cases}$$

The *di* between t_1 and t_2 is 1 when $r(t_1, t_1)$ and $r(t_2, t_2)$ are both numerically greater than $r(t_1, t_2)$, indicating that the neural representational formats were transformed from t_1 to t_2. Otherwise, *di* is 0. The *di* was computed in a binary way at the cost of information about effect size. This is because the similarity between repeated presentations of the same item may change over time—for example, similarity within the early encoding time period may be greater than in the late encoding time period (see fig. S3), which makes the absolute difference or the ratio between $r(t_1, t_2)$ and $r(t_1, t_1)$ [or $r(t_2, t_2)$] less meaningful. The temporal map of *di* was obtained on either the subject level or the item level. We then averaged *di* values according to the two time dimensions that were used to compute the WI similarity, resulting in one time-resolved *di* at each time point.

Multilevel mediation analysis

We performed a multilevel mediation analysis to examine whether the effect of representational dynamicity on long-term memory performance was mediated by the representational fidelity, i.e., the amount of item-specific information during encoding. The representational dynamicity was extracted for individual items (word-picture associations) of each participant. Specifically, for the WI EES of each item, we contrasted the on-diagonal similarity versus off-diagonal similarity and obtained the temporal map of the *di* (see above) across

the encoding period. The *di* was then averaged across all encoding time windows, resulting in one *di* index for each item. Long-term memory performance was measured as a binary variable reflecting the retrieval success or failure of each item. The amount of item-specific information of each item was extracted by subtracting the similarity of this item with all other items from the WI similarity of this item, within the temporal cluster showing a subsequent memory effect (Fig. 3B). Given that the data of "representational *di*" and "item-specific representations" contain two levels, with items nested in subjects, a mediation analysis for multilevel data was performed in R by using the *mediation* package as well as the *lme4* package (*87*). Two models were built in the analysis: (i) a linear mixed-effects model to test the relationship between the representational dynamicity and the amount of item-specific information, with "representational dynamicity" as the fixed effect and "subject" as the random effect; and (2) a generalized linear mixed-effects model with "representational dynamicity" as the fixed effect, "subject" as a random effect, "item-specific EES" as the mediator, and "long-term memory performance" as the dependent variable. The direct and indirect effects were then obtained by contrasting these two models.

Multiple comparisons correction

A nonparametric cluster-based permutation test was applied for the correction of multiple comparisons in all contrasts where temporal clusters were examined (*88*). For this type of analysis, we first identified temporal windows that showed a significant difference (at $P < 0.05$) between conditions based on the empirical data. Adjacent significant temporal windows formed a temporal cluster, and the statistical value of the cluster was obtained by summing the *t* values within the cluster. Then, we created a null distribution by randomly shuffling the condition labels for each subject independently 1000 times and calculated the contrast between conditions across subjects after each permutation. We extracted the sum of the *t* values of the largest cluster from each permutation as surrogate statistical values. The corrected significance level was then obtained by comparing the empirical cluster value with the 1000 surrogate values. For examining the correlation between neural representational similarity and representational similarity in the DNN and semantic model, we created surrogate null distributions by randomly shuffling the labels of pictures in the neural similarity matrices 1000 times. Each surrogate neural similarity matrix was then correlated with the early visual, late visual, semantic, and abstract semantic matrix. This was done separately for each time window and then tested against zero across subjects, resulting in one of the 1000 surrogate statistical values in the null distribution. Again, empirical statistical values for each cluster and for the four representational formats were then compared with the statistical values in the respective null distribution to obtain the corrected significance level.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abg9715

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. F. C. Bartlett, *Remembering: A Study in Experimental and Social Psychology* (Cambridge Univ. Press, 1932).
2. J. F. Danker, A. Tompary, L. Davachi, Trial-by-trial hippocampal encoding activation predicts the fidelity of cortical reinstatement during subsequent retrieval. *Cereb. Cortex* **27**, 3515–3524 (2017).
3. A. M. Gordon, J. Rissman, R. Kiani, A. D. Wagner, Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cereb. Cortex* **24**, 3350–3364 (2014).
4. M. Ritchey, E. A. Wing, K. S. LaBar, R. Cabeza, Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cereb. Cortex* **23**, 2818–2828 (2013).
5. B. P. Staresina, R. N. A. Henson, N. Kriegeskorte, A. Alink, Episodic reinstatement in the medial temporal lobe. *J. Neurosci.* **32**, 18150–18156 (2012).
6. B. P. Staresina, S. Michelmann, M. Bonnefond, O. Jensen, N. Axmacher, J. Fell, Hippocampal pattern completion is linked to gamma power increases and alpha power decreases during recollection. *eLife* **5**, e17397 (2016).
7. S. E. Favila, H. Lee, B. A. Kuhl, Transforming the concept of memory reactivation. *Trends Neurosci.* **43**, 939–950 (2020).
8. G. Xue, The neural representations underlying human episodic memory. *Trends Cogn. Sci.* **22**, 544–561 (2018).
9. S. E. Favila, R. Samide, S. C. Sweigart, B. A. Kuhl, Parietal representations of stimulus features are amplified during memory retrieval and flexibly aligned with top-down goals. *J. Neurosci.* **38**, 7809–7821 (2018).
10. X. Xiao, Q. Dong, J. Gao, W. Men, R. A. Poldrack, G. Xue, Transformed neural pattern reinstatement during episodic memory retrieval. *J. Neurosci.* **37**, 2986–2998 (2017).
11. J. L. Breedlove, G. St-Yves, C. A. Olman, T. Naselaris, Generative feedback explains distinct brain activity codes for seen and mental images. *Curr. Biol.* **30**, 2211–2224.e6 (2020).
12. S. E. Favila, B. A. Kuhl, J. Winawer, Perception and memory have distinct spatial tuning properties in human visual cortex. *bioRxiv*, 811331 (2020).
13. J. Linde-Domingo, M. S. Treder, C. Kerrén, M. Wimber, Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nat. Commun.* **10**, 179 (2019).
14. J. Lifanov, J. Linde-Domingo, M. Wimber, Feature-specific reaction times reveal a semanticisation of memories over time and with repeated remembering. *Nat. Commun.* **12**, 3177 (2021).
15. S. Michelmann, B. P. Staresina, H. Bowman, S. Hanslmayr, Speed of time-compressed forward replay flexibly changes in human episodic memory. *Nat. Hum. Behav.* **3**, 143–154 (2019).
16. G. E. Wimmer, Y. Liu, N. Vehar, T. E. J. Behrens, R. J. Dolan, Episodic memory retrieval success is associated with rapid replay of episode content. *Nat. Neurosci.* **23**, 1025–1033 (2020).
17. N. Kriegeskorte, J. Diedrichsen, Peeling the onion of brain representations. *Annu. Rev. Neurosci.* **42**, 407–432 (2019).
18. A. Saxe, S. Nelli, C. Summerfield, If deep learning is the answer, what is the question? *Nat. Rev. Neurosci.* **22**, 55–67 (2021).
19. A. Clarke, B. J. Devereux, L. K. Tyler, Oscillatory dynamics of perceptual to conceptual transformations in the ventral visual pathway. *J. Cogn. Neurosci.* **30**, 1590–1605 (2018).
20. J. Liu, H. Zhang, T. Yu, D. Ni, L. Ren, Q. Yang, B. Lu, D. Wang, R. Heinen, N. Axmacher, G. Xue, Stable maintenance of multiple representational formats in human visual short-term memory. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 32329–32339 (2020).
21. H. Liang, X. Gong, M. Chen, Y. Yan, W. Li, C. D. Gilbert, Interactions between feedback and lateral connections in the primary visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8637–8642 (2017).
22. K. A. Paller, A. D. Wagner, Observing the transformation of experience into memory. *Trends Cogn. Sci.* **6**, 93–102 (2002).
23. G. Xue, Q. Dong, C. Chen, Z. Lu, J. A. Mumford, R. A. Poldrack, Greater neural pattern similarity across repetitions is associated with better memory. *Science* **330**, 97–101 (2010).
24. G. Xue, Q. Dong, C. Chen, Z.-L. Lu, J. A. Mumford, R. A. Poldrack, Complementary role of frontoparietal activity and cortical pattern similarity in successful episodic memory encoding. *Cereb. Cortex* **23**, 1562–1571 (2013).
25. K. Feng, X. Zhao, J. Liu, Y. Cai, Z. Ye, C. Chen, G. Xue, Spaced learning enhances episodic memory by increasing neural pattern similarity across repetitions. *J. Neurosci.* **39**, 5351–5360 (2019).
26. Y. Lu, C. Wang, C. Chen, G. Xue, Spatiotemporal neural pattern similarity supports episodic memory. *Curr. Biol.* **25**, 780–785 (2015).
27. Y. Dudai, A. Karni, J. Born, The consolidation and transformation of memory. *Neuron* **88**, 20–32 (2015).
28. A. Ben-Yakov, N. Eshel, Y. Dudai, Hippocampal immediate poststimulus activity in the encoding of consecutive naturalistic episodes. *J. Exp. Psychol. Gen.* **142**, 1255–1263 (2013).
29. I. Sols, S. DuBrow, L. Davachi, L. Fuentemilla, Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. *Curr. Biol.* **27**, 3499–3504.e4 (2017).

30. A. Jafarpour, S. Griffin, J. J. Lin, R. T. Knight, Medial orbitofrontal cortex, dorsolateral prefrontal cortex, and hippocampus differentially represent the event saliency. *J. Cogn. Neurosci.* **31**, 874–884 (2019).
31. A. Ben-Yakov, Y. Dudai, Constructing realistic engrams: Poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *J. Neurosci.* **31**, 9032–9042 (2011).
32. A. Tambini, L. Davachi, Persistence of hippocampal multivoxel patterns into postencoding rest is related to memory. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19591–19596 (2013).
33. A. Baddeley, The episodic buffer: A new component of working memory? *Trends Cogn. Sci.* **4**, 417–423 (2000).
34. M. Moscovitch, R. Cabeza, G. Winocur, L. Nadel, Episodic memory and beyond: The hippocampus and neocortex in transformation. *Annu. Rev. Psychol.* **67**, 105–134 (2016).
35. P. Khader, C. Ranganath, A. Seemüller, F. Rösler, Working memory maintenance contributes to long-term memory formation: Evidence from slow event-related brain potentials. *Cogn. Affect. Behav. Neurosci.* **7**, 212–224 (2007).
36. A. S. Souza, K. Oberauer, Time to process information in working memory improves episodic memory. *J. Mem. Lang.* **96**, 155–167 (2017).
37. N. Axmacher, D. P. Schmitz, I. Weinreich, C. E. Elger, J. Fell, Interaction of working memory and long-term memory in the medial temporal lobe. *Cereb. Cortex* **18**, 2868–2878 (2008).
38. C. Ranganath, M. X. Cohen, C. J. Brozinsky, Working memory maintenance contributes to long-term memory formation: Neural and behavioral evidence. *J. Cogn. Neurosci.* **17**, 994–1010 (2005).
39. U. Güçlü, M. A. J. van Gerven, Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *J. Neurosci.* **35**, 10005–10014 (2015).
40. M. G. Stokes, 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends Cogn. Sci.* **19**, 394–405 (2015).
41. E. Spaak, K. Watanabe, S. Funahashi, M. G. Stokes, Stable and dynamic coding for working memory in primate prefrontal cortex. *J. Neurosci.* **37**, 6503–6516 (2017).
42. A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks. *Commun. ACM* **60**, 84–90 (2012).
43. G. Giari, E. Leonardelli, Y. Tao, M. Machado, S. L. Fairhall, Spatiotemporal properties of the neural representation of conceptual content for words and pictures—An MEG study. *Neuroimage* **219**, 116913 (2020).
44. Y. Song, S. Shi, J. Li, H. Zhang, in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)* (Association for Computational Linguistics, 2018); http://aclweb.org/anthology/N18-2028, pp. 175–180.
45. X. Wang, W. Men, J. Gao, A. Caramazza, Y. Bi, Two forms of knowledge representations in the human brain. *Neuron* **107**, 383–393.e5 (2020).
46. R. B. Yaffe, M. S. D. Kerr, S. Damera, S. V. Sarma, S. K. Inati, K. A. Zaghloul, Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18727–18732 (2014).
47. D. Pacheco Estefan, M. Sánchez-Fibla, A. Duff, A. Principe, R. Rocamora, H. Zhang, N. Axmacher, P. F. M. J. Verschure, Coordinated representational reinstatement in the human hippocampus and lateral temporal cortex during episodic memory retrieval. *Nat. Commun.* **10**, 2255 (2019).
48. K. F. LaRocque, M. E. Smith, V. A. Carr, N. Witthoft, K. Grill-Spector, A. D. Wagner, Global similarity and pattern separation in the human medial temporal lobe predict subsequent memory. *J. Neurosci.* **33**, 5466–5474 (2013).
49. Y. Xu, Reevaluating the sensory account of visual working memory storage. *Trends Cogn. Sci.* **21**, 794–815 (2017).
50. H. Zhang, J. Fell, N. Axmacher, Electrophysiological mechanisms of human memory consolidation. *Nat. Commun.* **9**, 4103 (2018).
51. T. B. Christophel, P. C. Klink, B. Spitzer, P. R. Roelfsema, J.-D. Haynes, The distributed nature of working memory. *Trends Cogn. Sci.* **21**, 111–124 (2017).
52. R. M. Cichy, D. Pantazis, A. Oliva, Resolving human object recognition in space and time. *Nat. Neurosci.* **17**, 455–462 (2014).
53. A. S. Ghuman, N. M. Brunet, Y. Li, R. O. Konecky, J. A. Pyles, S. A. Walls, V. Destefino, W. Wang, R. M. Richardson, Dynamic encoding of face information in the human fusiform gyrus. *Nat. Commun.* **5**, 5672 (2014).
54. R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, A. Oliva, Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Sci. Rep.* **6**, 27755 (2016).
55. M. J. Chadwick, R. S. Anjum, D. Kumaran, D. L. Schacter, H. J. Spiers, D. Hassabis, Semantic representations in the temporal pole predict false memories. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10180–10185 (2016).
56. Z. Ye, B. Zhu, L. Zhuang, Z. Lu, C. Chen, G. Xue, Neural global pattern similarity underlies true and false memories. *J. Neurosci.* **36**, 6792–6802 (2016).
57. S. Kapur, F. I. Craik, E. Tulving, A. A. Wilson, S. Houle, G. M. Brown, Neuroanatomical correlates of encoding in episodic memory: Levels of processing effect. *Proc. Natl. Acad. Sci.* **91**, 2008–2011 (1994).
58. P. A. Packard, A. Rodríguez-Fornells, N. Bunzeck, B. Nicolás, R. de Diego-Balaguer, L. Fuentemilla, Semantic congruence accelerates the onset of the neural signals of successful memory encoding. *J. Neurosci.* **37**, 291–301 (2017).
59. S. W. Davis, B. R. Geib, E. A. Wing, W.-C. Wang, M. Hovhannisyan, Z. A. Monge, R. Cabeza, Visual and semantic representations predict subsequent memory in perceptual and conceptual memory tests. *Cereb. Cortex* **31**, 974–992 (2021).
60. A. Gilboa, H. Marlatte, Neurobiology of schemas and schema-mediated memory. *Trends Cogn. Sci.* **21**, 618–631 (2017).
61. D. Tse, R. F. Langston, M. Kakeyama, I. Bethus, P. A. Spooner, E. R. Wood, M. P. Witter, R. G. M. Morris, Schemas and memory consolidation. *Science* **316**, 76–82 (2007).
62. T. Sommer, The emergence of knowledge and how it supports the memory for novel related information. *Cereb. Cortex* **27**, 1906–1921 (2017).
63. O. Bein, N. Reggev, A. Maril, Prior knowledge promotes hippocampal separation but cortical assimilation in the left inferior frontal gyrus. *Nat. Commun.* **11**, 4590 (2020).
64. U. Hasson, J. Chen, C. J. Honey, Hierarchical process memory: Memory as an integral component of information processing. *Trends Cogn. Sci.* **19**, 304–313 (2015).
65. Z.-X. Liu, C. Grady, M. Moscovitch, Effects of prior-knowledge on brain activation and connectivity during associative memory encoding. *Cereb. Cortex* **27**, 1991–2009 (2017).
66. S. K. Jeong, Y. Xu, Behaviorally relevant abstract object identity representation in the human parietal cortex. *J. Neurosci.* **36**, 1607–1619 (2016).
67. S. M. Polyn, V. S. Natu, J. D. Cohen, K. A. Norman, Category-specific cortical activity precedes retrieval during memory search. *Science* **310**, 1963–1966 (2005).
68. A. Jafarpour, L. Fuentemilla, A. J. Horner, W. Penny, E. Duzel, Replay of very early encoding representations during recollection. *J. Neurosci.* **34**, 242–248 (2014).
69. T. Staudigl, C. Vollmar, S. Noachtar, S. Hanslmayr, Temporal-pattern similarity analysis reveals the beneficial and detrimental effects of context reinstatement on human memory. *J. Neurosci.* **35**, 5373–5384 (2015).
70. D. Zeithamova, A. L. Dominick, A. R. Preston, Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron* **75**, 168–179 (2012).
71. A. M. Albers, P. Kok, I. Toni, H. C. Dijkerman, F. P. de Lange, Shared representations for working memory and mental imagery in early visual cortex. *Curr. Biol.* **23**, 1427–1431 (2013).
72. S. Xie, D. Kaiser, R. M. Cichy, Visual imagery and perception share neural representations in the alpha frequency band. *Curr. Biol.* **30**, 2621–2627.e5 (2020).
73. J. A. Lewis-Peacock, B. R. Postle, Temporary activation of long-term memory supports working memory. *J. Neurosci.* **28**, 8765–8771 (2008).
74. T. C. Kietzmann, C. J. Spoerer, L. K. A. Sörensen, R. M. Cichy, O. Hauk, N. Kriegeskorte, Recurrence is required to capture the representational dynamics of the human visual system. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 21854–21863 (2019).
75. K. Fukuda, G. F. Woodman, Visual working memory buffers information retrieved from visual long-term memory. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5306–5311 (2017).
76. M. F. Carr, S. P. Jadhav, L. M. Frank, Hippocampal replay in the awake state: A potential substrate for memory consolidation and retrieval. *Nat. Neurosci.* **14**, 147–153 (2011).
77. A. Tambini, L. Davachi, Awake reactivation of prior experiences consolidates memories and biases cognition. *Trends Cogn. Sci.* **23**, 876–890 (2019).
78. N. C. Tronson, J. R. Taylor, Molecular mechanisms of memory reconsolidation. *Nat. Rev. Neurosci.* **8**, 262–275 (2007).
79. Y. Liu, R. J. Dolan, Z. Kurth-Nelson, T. E. J. Behrens, Human replay spontaneously reorganizes experience. *Cell* **178**, 640–652.e14 (2019).
80. Z. Ye, L. Shi, A. Li, C. Chen, G. Xue, Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. *eLife* **9**, e57023 (2020).
81. S. Mirjalili, P. Powell, J. Strunk, T. James, A. Duarte, Context memory encoding and retrieval temporal dynamics are modulated by attention across the adult lifespan. *eNeuro* **8**, ENEURO.0387-20.2020 (2021).
82. J. D. Karpicke, H. L. Roediger, The critical importance of retrieval for learning. *Science* **319**, 966–968 (2008).
83. J. C. Hulbert, K. A. Norman, Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cereb. Cortex* **25**, 3994–4008 (2015).
84. A. S. Ghuman, Dynamic neural representations: An inferential challenge for fMRI. *Trends Cogn. Sci.* **23**, 534–536 (2019).
85. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, F.-F. Li, ImageNet: A large-scale hierarchical image database, in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 248–255.

86. P. Bao, L. She, M. McGill, D. Y. Tsao, A map of object space in primate inferotemporal cortex. *Nature* **583**, 103–108 (2020).
87. D. Tingley, T. Yamamoto, K. Hirose, L. Keele, K. Imai, Mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59**, 10.18637/jss.v059.i05, (2014).
88. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).

Acknowledgments
Funding: G.X. was supported by the National Science Foundation of China (31730038), the China-Israel collaborative research grant (NSFC 31861143040), and the Guangdong Pearl River Talents Plan Innovative and Entrepreneurial Team grant #2016ZT06S220. N.A. received funding from the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation)—Projektnummer 316803389—SFB 1280, via Projektnummer 122679504—SFB 874, and via DFG grant AX 82/3. **Author contributions:** Conceptualization: J.L. and G.X. Methodology: J.L., T.Y., L.R., D.N., Q.Y., and B.L. Investigation: J.L., H.Z., L.Z., N.A., and G.X. Supervision: N.A. and G.X. Writing—original draft: J.L. and G.X. Writing—review and editing: J.L., H.Z., N.A., and G.X. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 7 February 2021
Accepted 17 August 2021
Published 8 October 2021
10.1126/sciadv.abg9715

Citation: J. Liu, H. Zhang, T. Yu, L. Ren, D. Ni, Q. Yang, B. Lu, L. Zhang, N. Axmacher, G. Xue, Transformative neural representations support long-term episodic memory. *Sci. Adv.* **7**, eabg9715 (2021).

Transformative neural representations support long-term episodic memory

Jing LiuHui ZhangTao YuLiankun RenDuanyu NiQinhao YangBaoqing LuLiang ZhangNikolai AxmacherGui Xue

Sci. Adv., 7 (41), eabg9715.