



Rapid neural reorganization during retrieval practice predicts subsequent long-term retention and false memory

Liping Zhuang^{1,2,7}, Jingyi Wang^{1,2,7}, Bingsen Xiong^{1,2}, Cheng Bian^{1,2}, Lei Hao^{1,2,3}, Peter J. Bayley^{4,5} and Shaozheng Qin^{1,2,6} ✉

Active retrieval can alter the strength and content of a memory, yielding either enhanced or distorted subsequent recall. However, how consolidation influences these retrieval-induced seemingly contradictory outcomes remains unknown. Here we show that rapid neural reorganization over an eight-run retrieval practice predicted subsequent recall. Retrieval practice boosted memory retention following a 24-hour (long-term) but not 30-minute delay, and increased false memory at both delays. Long-term retention gains were predicted by multi-voxel representation distinctiveness in the posterior parietal cortex (PPC) that increased progressively over retrieval practice. False memory was predicted by unstable representation distinctiveness in the medial temporal lobe (MTL). Retrieval practice enhanced the efficiency of memory-related brain networks, through building up PPC and MTL connections with the ventrolateral and dorsolateral prefrontal cortex that predicted long-term retention gains and false memory, respectively. Our findings indicate that retrieval-induced rapid neural reorganization together with consecutive consolidation fosters long-term retention and false memories via distinct pathways.

For centuries, humans have attempted to improve memory using different mnemonic strategies. Perhaps the simplest strategy, known as retrieval practice, is through the act of retrieval, or actively recalling information as memory for that information becomes strengthened. Such retrieval practice can boost long-term retention, with robust memory gains after intervals of days or months¹, suggesting that memory consolidation is also involved². The value of retrieval practice is well recognized and widely used in education³, and also in the clinic to aid age-associated memory impairment⁴. Yet, active retrieval can also transiently change the memory strength and content⁵, thereby could render the memory trace labile and to be altered by current experience⁶, potentially inducing false memory⁷. However, little is known about how memories are reorganized through retrieval practice and subsequent consolidation to produce two seemingly contradictory effects.

Multiple psychological theories attempt to account for the effects of retrieval practice on memory⁸. The elaboration theory suggests that active retrieval can enrich the links of semantic networks, because additional associations and alternative routes are formed by retrieving the targeted memory^{9,10}. The search-set restriction theory suggests that retrieval strengthens cue–target associations while suppressing irrelevant ones, which shares some similarities to the bifurcation model¹¹. However, these theories appear insufficient to address how false memory is produced by retrieval practice. Recent episodic context accounts offer a new perspective in suggesting that memories can be updated by current contexts at each retrieval, and the current and previous contexts could serve as multiple cues to increase the accessibility of targeted memories¹². This echoes neurocognitive evidence that retrieval practice

facilitates memory updating through refined neural representations to discriminate individual memories^{13–16}. It also offers a possible explanation for retrieval-induced false memory through which retrieved memories become labile during experience^{17,18}, analogous to a kind of new learning. Such retrieval-induced learning is believed to involve a dynamic assembly of memory-related neural ensembles that reconstruct representations to meet the ever-changing environmental needs¹⁹. However, there are still gaps in our understanding regarding the neurocognitive mechanisms on how retrieval practice actively reshapes the original representations to predict subsequent memory outcomes.

Another fundamental question regarding retrieval practice is how retrieval-induced transient changes in memory traces are transformed into stable representations to support subsequent long-term retention, while still being malleable enough for flexible needs. Traditional views on systems consolidation posit that memories initially rely on the MTL (especially the hippocampus), and slowly transform into stable representations through strengthening cortical connections over time¹⁶. But this theory has been challenged by recent evidence showing that newly acquired memories can become enduring engrams in the prefrontal cortex (PFC) in rodents soon after encoding²⁰, or in the human PPC via repeated study–test cycles²¹, supporting a theory viewing retrieval as a fast route to memory consolidation¹³. However, recent studies have demonstrated support for the systems consolidation theory by showing that offline consolidation could still be required for long-lasting memory after retrieval practice^{22,23}. In addition, results are mixed on the question as to whether systems consolidation could foster false memories^{24–26}. Thus, the question of how retrieval-induced online

¹State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China.

²Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing, China. ³College of Teacher Education, Southwest University, Chongqing, China. ⁴War Related Illness and Injury Study Center, US Department of Veterans Affairs, Palo Alto, CA, USA. ⁵Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. ⁶Chinese Institute for Brain Research, Beijing, China. ⁷These authors contributed equally:

Liping Zhuang, Jingyi Wang. ✉e-mail: szqin@bnu.edu.cn

changes work together with offline consolidation to predict both true and false memory outcomes remains open.

The act of episodic retrieval involves multiple brain systems. Previous neuroimaging studies using univariate analyses have identified active engagement of the MTL, the ventrolateral and dorsolateral prefrontal cortex (VLPFC and DLPFC, respectively), and the PPC regions in retrieval practice^{27–29}, but they provided limited insight into the dynamic nature of retrieval-induced reconstructive modifications of the original memory representations. Multivariate pattern analysis, especially the measurement of neural distinctiveness, has been widely used to assess multi-voxel representation patterns linked to individual memories during encoding and retrieval^{15,30}. So far, little is known about how memory-related neural representations in prefrontal, parietal and MTL systems evolve over retrieval practice, and how these online changes contribute to subsequent memory consolidation. Furthermore, memory retrieval involves a complex network of widely distributed brain systems, including the VLPFC and DLPFC for retrieval search³¹ and episodic context updating^{32,33} as well as updating of existing memories³⁴, the MTL for associative memories^{14,35} and the PPC for representing specific retrieved content³⁶. It is thus of interest to evaluate how the dynamic assembly of memory-related brain networks among MTL, PFC and PPC regions support the online reinstatement of memories during retrieval practice. Network neuroscience now offers graph-based network approaches to provide insight into how interregional functional connections are dynamically reconfigured over a course of progressive learning, including motor learning³⁷ and mnemonic training³⁸. There is little information available as to how retrieval practice reconfigures memory-related large-scale brain networks to predict subsequent memory outcomes after consolidation.

Here we address the above questions by integrating event-related functional magnetic resonance imaging (fMRI) with an eight-run retrieval practice and prospective consolidation paradigm across 2 days (Fig. 1a) in combination with advanced analysis of multi-voxel representation patterns and network configurations. During the memory acquisition phase, participants were first trained to acquire 48 face–scene associations. Participants then underwent fMRI scanning while they were performing an eight-run memory practice phase. In this phase, participants either actively retrieved 16 pairs of associations (the RP condition), or not retrieved for another 16 pairs (the NR condition), using each face as a cue. The remaining 16 pairs were not presented during the memory practice phase, serving as the baseline condition. Thereafter, two cued-recall tests were given outside the scanner to assess memory retention for face–scene associations after 30-min (short-term) and 24-h (long-term) intervals. Memory performance was scored for the remembrance of face–scene associations, and false memory if incorrect content was given when describing the associated scene. Multivariate and network analyses of neural activity and connectivity over an eight-run retrieval practice allowed us to track dynamic changes in neural representations and network configurations, and to determine which changes predicted these two memory outcomes. On the basis of the aforementioned empirical observations of retrieval practice and systems consolidation models, we expected retrieval practice to boost long-term retention after consolidation. We further expected that such beneficial effects would be associated with retrieval-induced refinement in neural distinctiveness and network configurations in the PPC and PFC. And if retrieval practice could produce false memories by retrieval-mediated learning and context updating¹², we would expect that the MTL-centric neural representations and network reorganization would predict such false memory outcomes.

Results

Retrieval practice boosts retention with false memories. We first examined the effectiveness of retrieval practice on memory retention in immediate and delayed tests. A repeated-measure 2 × 3 analysis

of variance (ANOVA) for associative memory performance, with time (immediate versus delayed) and condition (RP versus NR versus baseline conditions) as within-participant factors, revealed a main effect of conditions ($F_{(2,112)} = 23.97, P < 0.001, \eta^2_p = 0.30$, 90% confidence interval (CI) = 0.18, 0.39) and condition-by-time interaction ($F_{(2,112)} = 10.38, p < 0.001, \eta^2_p = 0.16$, 90%, CI = 0.06, 0.25; Fig. 1b and see Supplementary Results 2 for post hoc tests). Critically, we observed superior long-term retention (relative to baseline) for face–scene associations in the delayed test for RP ($21.5 \pm 3.1\%$) than NR condition ($7.6 \pm 3.3\%$) ($t_{(56)} = 4.37, P < 0.001$, Cohen's $d = 0.58$, 95% CI = 0.08, 0.20) (Fig. 1c). No statistical difference was observed for short-term retention in the immediate test ($t_{(56)} = -1.38, P = 0.17$, Cohen's $d = -0.18$, 95% CI = $-0.09, 0.02$, $BF_{01} = 2.83$). Moreover, we computed long-term retention gains by subtracting retention performance in the immediate test from the delayed test. Paired t -tests revealed that long-term retention for the RP condition gained significantly higher retention scores ($18.1 \pm 4.6\%$) after consolidation than the NR condition ($0.60 \pm 4.6\%$) ($t_{(56)} = 4.19, P < 0.001$, Cohen's $d = 0.56$, 95% CI = 0.09, 0.26) (Fig. 1d). Such gains resulted from a marked increase of long-term retention in the RP after 24-h, and no change in the NR between immediate and delayed tests (Fig. 1b). Additional analyses for vividness ratings are provided in Supplementary Result 1 and Fig. 1.

Parallel 2 × 3 ANOVA for false memory scores revealed a main effect of conditions ($F_{(2,98)} = 9.46, P < 0.001, \eta^2_p = 0.16$, 90% CI = 0.06, 0.26; Fig. 1e) and detailed statistics are provided in Supplementary Results 2. We then compared participants' memory scores (relative to baseline) for false content recalled in the immediate and delayed tests for RP and NR conditions. We observed significantly higher false memory for the RP than NR condition in the delayed test after 24 h ($t_{(49)} = 2.49, P = 0.016$, Cohen's $d = 0.35$, 95% CI = 0.09, 0.87) as well as a marginally significant effect in the immediate test after 30 min ($t_{(49)} = 1.92, P = 0.06$, Cohen's $d = 0.27$, 95% CI = $-0.02, 0.89$; Fig. 1f). On average, false memory scores across the two delays were significantly higher in the RP (0.65 ± 0.14) than NR condition (0.19 ± 0.16) ($t_{(49)} = 2.87, P = 0.006$, Cohen's $d = 0.41$, 95% CI = 0.14, 0.78; Fig. 1g). Together, these results indicate that retrieval practice boosts long-term retention gains for remembering face–scene associations after consolidation and also induces false memory in general.

Retrieval practice refines representations in PPC and MTL. Next, we examined retrieval-induced dynamic changes in memory-related neural representations over eight-run retrieval practices. We restricted our analyses to retrieval-related brain systems derived from a large-scale meta-analysis on the NeuroSynth platform. To verify the validity of this mask, we conducted a set of condition-specific similarity analyses for multi-voxel activity patterns during the RP (or NR) condition relative to the canonical retrieval-related activation map from the NeuroSynth (Fig. 2a). Such an approach allowed us to simply assess how the average neural activity patterns across all trials within the RP (or NR) condition was similar to the NeuroSynth canonical reference pattern. A paired t -test revealed a higher pattern similarity in RP than NR ($t_{(49)} = 5.51, P < 0.001$, Cohen's $d = 0.78$, 95% CI = 0.02, 0.05; Fig. 2b). Further analyses of changes in this similarity metric over the eight runs revealed smaller variance (variability) of neural similarity in the RP than NR ($t_{(49)} = -3.47, P = 0.001$, Cohen's $d = -0.49$, 95% CI = $-0.01, -0.004$; Fig. 2c). These results indicate a greater and more stable involvement of retrieval-related brain systems in RP than NR (Supplementary Fig. 2).

We then examined dynamic changes in the distinctiveness of memory-related neural representations across individual trials over retrieval practice, by analysing the dissimilarity of intertrial multi-voxel neural activity patterns in the canonical retrieval-related brain mask separately for RP and NR trials in each run (Fig. 2d). The distinctiveness measure reflects how individual memories are refined and become discrete from each other through

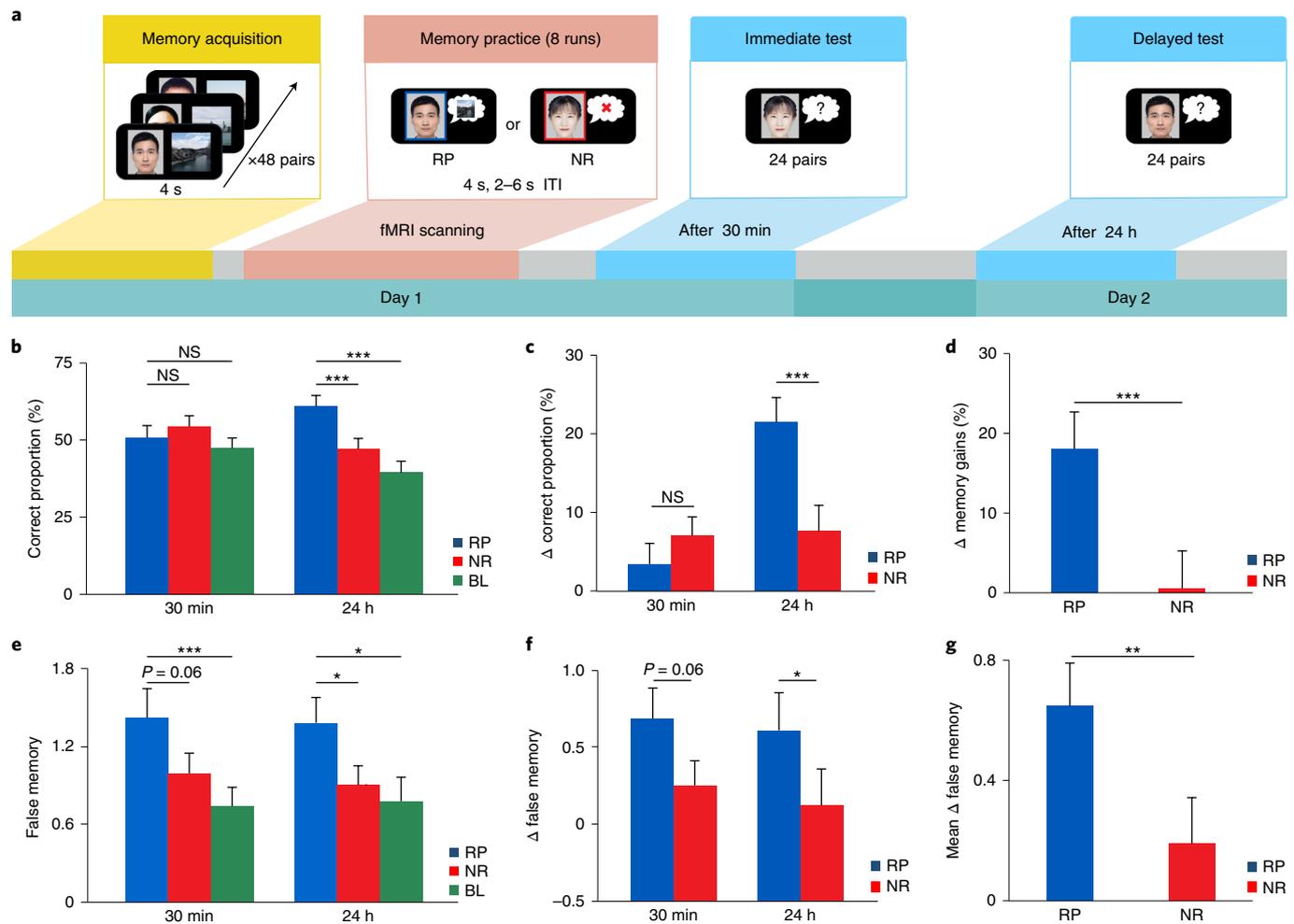


Fig. 1 | Experimental design and memory differences on RP and NR at two tests. **a**, Experimental design. The experiment consisted of three phases including memory acquisition, memory practice and subsequent memory tests. During memory acquisition, participants were trained to acquire 48 face-scene associations. During memory practice in the scanner, participants were instructed to perform either RP (blue) and NR attempts over eight runs. Two cued-recall tests were performed to assess subsequent memory after short-term (30-min) and long-term (24-h) intervals. **b**, Bars depict face-scene associative memories in the RP, NR and baseline conditions at the immediate (30-min) and delayed (24-h) tests. A 2×3 repeated-measures ANOVA revealed a main effect of condition ($F_{(2,112)} = 23.97, P < 0.001, \eta_p^2 = 0.30, 90\% \text{ CI} = 0.18, 0.39$), and condition-by-time interaction ($F_{(2,112)} = 10.38, P < 0.001, \eta_p^2 = 0.16, 90\% \text{ CI} = 0.06, 0.25$). **c**, Bars depict memory performance for face-scene associations under the RP condition (relative to the corresponding baseline) outperformed the NR condition at the delayed test only ($t_{(56)} = 4.37, P < 0.001, d = 0.58, 95\% \text{ CI} = 0.08, 0.20$). **d**, Long-term retention gains that were calculated from the delayed recall performance relative to the immediate recall performance showed significant difference between the RP and NR conditions ($t_{(56)} = 4.19, P < 0.001, d = 0.56, 95\% \text{ CI} = 0.09, 0.26$). **e**, Parallel 2×3 ANOVA for false memory scores revealed a main effect of the conditions ($F_{(2,98)} = 9.46, P < 0.001, \eta_p^2 = 0.16, 90\% \text{ CI} = 0.06, 0.26$) for the RP, NR and baseline trials. **f**, Bar graphs depict the difference between false memory scores of the RP and NR conditions relative to their corresponding baselines on the 30-min ($t_{(49)} = 1.92, P = 0.06, d = 0.27, 95\% \text{ CI} = -0.02, 0.89$) and 24-h recall tests ($t_{(49)} = 2.49, P = 0.016, d = 0.35, 95\% \text{ CI} = 0.09, 0.87$). **g**, Bar graphs depict the difference of the mean false memory scores (after subtracting their corresponding baselines) across two recall tests between the RP and NR conditions ($t_{(49)} = 2.87, P = 0.006, d = 0.41, 95\% \text{ CI} = 0.14, 0.78$). Note: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS, not statistically significant. Error bars represent the standard error of the mean (s.e.m.).

the act of retrieval in each of eight runs. This analysis revealed a linear increase in intertrial neural distinctiveness over eight runs during the RP ($F_{(1,398)} = 10.95, P = 0.008$, after Bonferroni correction, 'corrected' hereafter, $\eta_p^2 = 0.03, 90\% \text{ CI} = 0.01, 0.06$; Fig. 2e), but such an effect was not evident in the NR ($F_{(1,398)} = 5.60, P = 0.15$ corrected, $\eta_p^2 = 0.01, 90\% \text{ CI} = 0, 0.04, \text{BF}_{01} = 1.63$; Fig. 2f). Additional neural pattern fidelity and whole-brain univariate validation analyses are provided in Supplemental Results 3 and 4 and Supplementary Figs. 3–5.

Moreover, we investigated how retrieval practice refined neural representations in core memory-related brain systems. We decomposed the overall NeuroSynth brain mask into the PPC, MTL and

PFC systems (Fig. 3a), and computed their corresponding intertrial neural pattern distinctiveness for RP and NR trials in each run. We observed a significant linear increase in neural distinctiveness over eight runs in the RP trials only in the PPC ($F_{(1,398)} = 12.06, P = 0.005$ corrected, $\eta_p^2 = 0.03, 90\% \text{ CI} = 0.008, 0.06$), and a marginally significant increase in the PFC ($F_{(1,398)} = 6.70, P = 0.08$ corrected, $\eta_p^2 = 0.02, 90\% \text{ CI} = 0.002, 0.04$), but not in the MTL ($F_{(1,398)} = 2.94, P = 0.70$ corrected, $\eta_p^2 = 0.007, 90\% \text{ CI} = 0.001, 0.03, \text{BF}_{01} = 2.19$; regression plots in the Fig. 3b). However, we did not observe any reliable increase in these systems for the NR trials (all $P > 0.16$ corrected, $\text{BF}_{01} = 1.33, 1.52$ and 1.31 for PPC, PFC and MTL, respectively). In the RP condition, the intertrial neural distinctiveness in

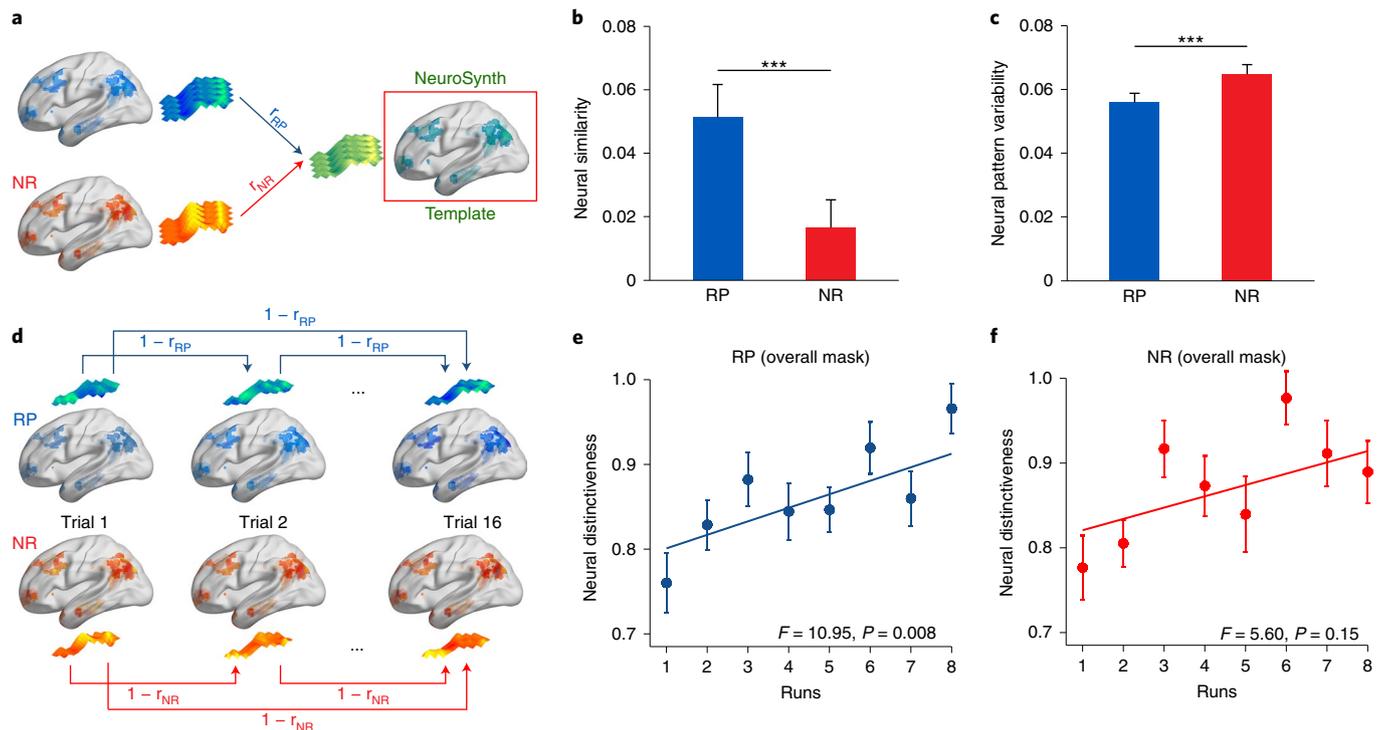


Fig. 2 | Dynamic changes in memory-related multi-voxel patterns over retrieval practice. **a**, An illustration of multi-voxel pattern similarity analysis, by computing Pearson's correlation of multi-voxel neural activity pattern in RP (in blue) and NR (in red) condition with the canonical retrieval-related brain activation map derived from the NeuroSynth platform (in green). **b**, Bars depict the difference of the averaged neural pattern similarity in the canonical retrieval-related brain mask across participants between RP and NR conditions ($t_{(49)} = 5.51, P < 0.001, d = 0.78, 95\% \text{ CI} = 0.02, 0.05$). **c**, Bars depict the difference of the averaged variance of neural pattern similarity over eight runs in RP and NR conditions ($t_{(49)} = -3.47, P = 0.001, d = -0.49, 95\% \text{ CI} = -0.01, -0.004$). **d**, An illustration of intertrial multi-voxel pattern distinctiveness, by computing pairwise pattern dissimilarity (that is, $1 - \text{Pearson's correlation coefficients}$) among RP (or NR) trials in each run. **e, f**, Linear regression plots show a significant linear increase in intertrial neural distinctiveness over eight runs in the RP condition (**e**) ($F_{(1,398)} = 10.95, P = 0.008, \eta^2_p = 0.03, 90\% \text{ CI} = 0.01, 0.06$), but it did not reach statistical significance in the NR condition (**f**) ($F_{(1,398)} = 5.60, P = 0.15, \eta^2_p = 0.01, 90\% \text{ CI} = 0, 0.04, \text{BF}_{01} = 1.63$). Note that $***P < 0.001$. P values in **e** and **f** were adjusted with the Bonferroni correction. Error bars represent s.e.m.

the final run was significantly higher than the first run in the PPC ($t_{(49)} = 3.55, P < 0.001, \text{Cohen's } d = 0.50, 95\% \text{ CI} = 0.07, 0.27$) and PFC ($t_{(49)} = 2.91, P = 0.005, \text{Cohen's } d = 0.41, 95\% \text{ CI} = 0.05, 0.25$; bar graphs in the Fig. 3b), but these effects were not evident for the NR trials (all $t_{(49)} < 1.08$, all $P > 0.28$; all $\text{BF}_{01} > 2.30$). Critically, the distinctiveness of the final run in the PPC was positively correlated with long-term memory gains in the RP ($r = 0.37, P = 0.009, 95\% \text{ CI} = 0.10, 0.58$) but not the NR trials ($r = -0.06, P = 0.67, 95\% \text{ CI} = -0.33, 0.22; \text{BF}_{01} = 5.20$; Fig. 3c). Further, Steiger's test revealed a significant difference in correlations between the RP and NR trials ($z = 2.27, p = 0.01$). The intertrial neural distinctiveness of the final run in the MTL was positively correlated with false memory in the RP ($r = 0.37, P = 0.02, 95\% \text{ CI} = 0.07, 0.60$) but not the NR trials ($r = -0.06, P = 0.69, 95\% \text{ CI} = -0.36, 0.24; \text{BF}_{01} = 4.87$; Fig. 3d). Further, Steiger's test revealed a significant difference in these correlations between conditions ($z = 2.13, P = 0.017$). These results indicate that retrieval practice leads to heterogeneous dynamics of intertrial neural distinctiveness in the PPC, PFC and MTL, with higher distinctiveness in the PPC being related to better long-term retention gains and higher distinctiveness in the MTL being related to more false memories.

Retrieval practice reorganizes memory-related brain networks.

To track the dynamic assembly of large-scale memory-related brain networks including 15 nodes in the PPC, the PFC and the MTL over retrieval practice, we constructed a network consisting of 15×15 pairwise links for each run (Fig. 4a) using a

generalized form of context-dependent psychophysiological interaction (gPPI). Global efficiency was computed to assess how easily information flowed across a network via the shortest path between all pairs of nodes. Overall, we observed a significant increase in global network efficiency in the RP (t -test for slopes: $t_{(49)} = 2.78, P = 0.008, \text{Cohen's } d = 0.39, 95\% \text{ CI} = 0.15, 0.63$), but a decrease in the NR ($t_{(49)} = -2.05, P = 0.046, \text{Cohen's } d = -0.29, 95\% \text{ CI} = -0.53, -0.05$; Fig. 4b). Follow-up tests revealed higher slopes in the RP than the NR ($t_{(49)} = 3.37, P = 0.001, \text{Cohen's } d = 0.48, 95\% \text{ CI} = 0.18, 0.77$; Fig. 4c). Additional analyses also revealed dynamic network configuration changes over retrieval practice (Supplementary Fig. 6).

We then examined how dynamic reconfiguration of memory-related brain networks over retrieval practice contributed to long-term retention gains after consolidation. We implemented a network-behaviour prediction analysis by training a support vector regression (SVR) model with leave-one-out cross-validation, based on network features over eight runs in the RP (or NR) condition as the input data and long-term retention gains as the output variable. A feature selection procedure revealed that information derived from the top 1% links of all eight runs as the input features robustly predicted the accuracy of long-term retention (Supplementary Fig. 7a). The stable links across cross-validation iterations were then visualized in a cumulative manner in Fig. 5a.

We next performed a stepwise prediction analysis to characterize changes in predictive values as a function of eight runs by using selected links described above. The prediction values in the RP

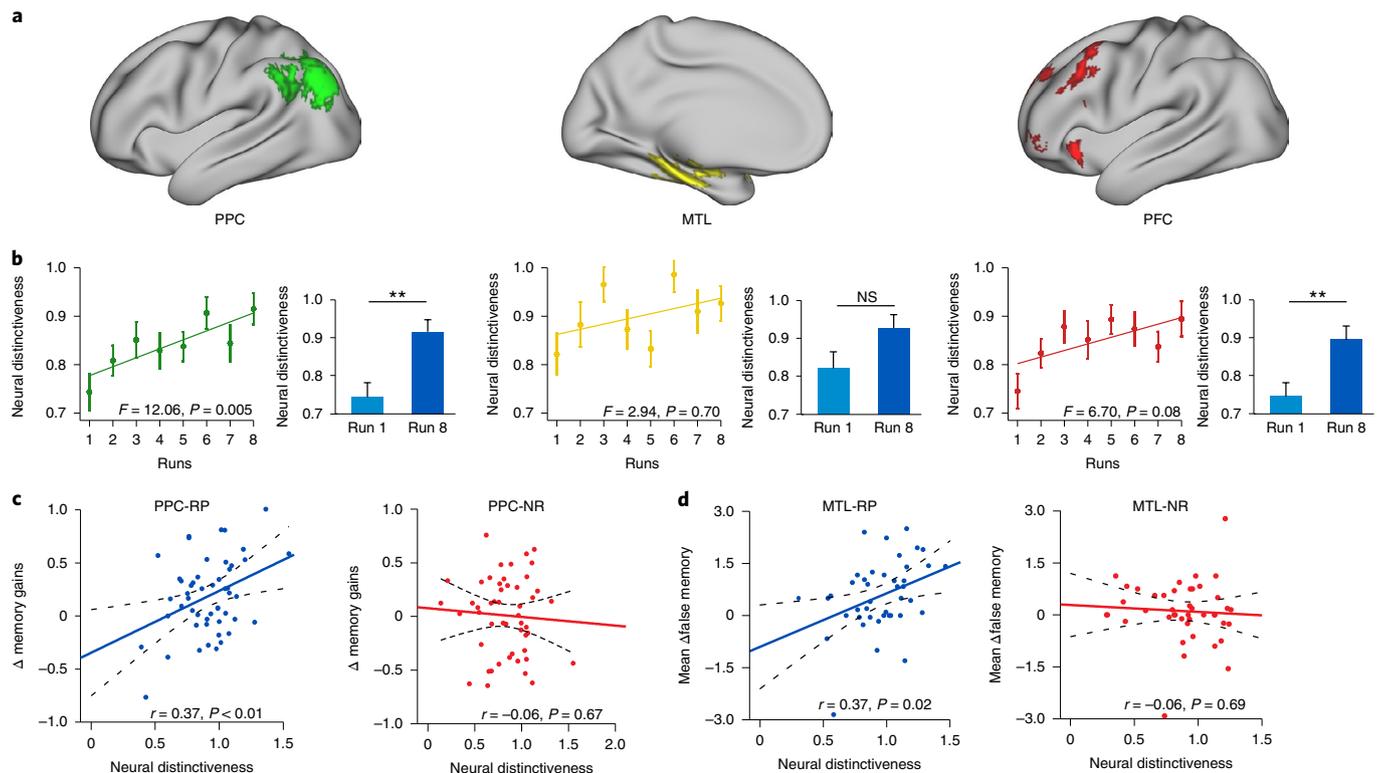


Fig. 3 | Dynamic changes in intertrial neural pattern distinctiveness in the PPC, MTL and PFC systems and their relation to subsequent memory outcomes. **a**, Lateral views of the PPC (left panel), MTL (middle panel) and PFC (right panel) systems. **b**, Linear regression plots show a significant linear increase in intertrial neural pattern distinctiveness over eight runs in the PPC ($F_{(1,398)} = 12.06, P = 0.005, \eta^2_p = 0.03, 90\% \text{ CI} = 0.008, 0.06$), a linear trend in the PFC ($F_{(1,398)} = 6.70, P = 0.08, \eta^2_p = 0.02, 90\% \text{ CI} = 0.002, 0.04$) and a null effect in the MTL ($F_{(1,398)} = 2.94, P = 0.70, \eta^2_p = 0.007, 90\% \text{ CI} = 0.001, 0.03, \text{BF}_{01} = 2.19$). Bar graphs depict corresponding intertrial neural distinctiveness between the first and the final runs in each system. **c**, Scatter plots show positive correlation of long-term retention gains for face-scene associations with intertrial neural distinctiveness in the final run for the RP (left, $r = 0.37, P = 0.009, 95\% \text{ CI} = 0.10, 0.58$), but not for the NR (right, $r = -0.06, P = 0.67, 95\% \text{ CI} = -0.33, 0.22; \text{BF}_{01} = 5.20$). **d**, Scatter plots show positive correlation of false memory scores with intertrial neural distinctiveness in the final run for the RP condition (left, $r = 0.37, P = 0.02, 95\% \text{ CI} = 0.07, 0.60$), but not for the NR condition (right, $r = -0.06, P = 0.69, 95\% \text{ CI} = -0.36, 0.24; \text{BF}_{01} = 4.87$). Note that beta and P values are from least-squares fitting and P values were adjusted with the Bonferroni correction. Error bars represent s.e.m.

were highest over the first five runs and remained at a stable level from then on (Supplementary Fig. 7c, blue line). This prediction in the RP outperformed the NR condition (Supplementary Fig. 7c, grey lines). To illustrate the network properties in the RP, we performed graph theory-based network analyses for the selected links over eight runs. The betweenness centrality, which referred to the fraction of all shortest paths in the network that passed through a given node³⁹, was computed to identify nodes that play a central role in coordinating with others over retrieval practice. These analyses revealed a general increase in betweenness for nodes in the PPC, PFC and MTL networks (t -test for slopes: $t_{(49)} = 10.64, P < 0.001, \text{Cohen's } d = 1.51, 95\% \text{ CI} = 0.78, 1.15; t_{(49)} = 9.98, P < 0.001, \text{Cohen's } d = 1.41, 95\% \text{ CI} = 0.40, 0.60; t_{(49)} = 8.67, P < 0.001, \text{Cohen's } d = 1.23, 95\% \text{ CI} = 0.20, 0.33$, respectively), with a more prominent increase in the PPC network (PPC versus PFC $t_{(49)} = 8.77, P < 0.001, \text{Cohen's } d = 1.24, 95\% \text{ CI} = 0.36, 0.57$; PPC versus MTL $t_{(49)} = 9.15, P < 0.001, \text{Cohen's } d = 1.29, 95\% \text{ CI} = 0.55, 0.86$; Fig. 5c). Of the 15 nodes, the right VLPFC showed the most prominent increase far above the others (compared to the second highest increased region of interest (ROI) $t_{(49)} = 2.53, P = 0.015, \text{Cohen's } d = 0.36, 95\% \text{ CI} = 0.05, 0.45$; Fig. 5c). These results indicate that dynamic reconfiguration of memory-related brain networks over retrieval practice is predictive of long-term retention gains, with the right VLPFC emerging as a hub of information processing that coordinated with other nodes.

PPC-VLPFC network configurations predict long-term retention. Given that the neural distinctiveness in the PPC was predictive of long-term retention gains and the VLPFC emerged as a hub to coordinate with other brain regions over retrieval practice, we thus proposed that functional communication between these two regions would facilitate the refinement of neural distinctiveness in the PPC in the final run. To test this hypothesis, we implemented additional prediction analysis using the links of the right VLPFC, with PPC nodes as input features. This analysis revealed that connectivity strength between these nodes was highly predictive of long-term retention gains ($r_{(\text{predicted}, \text{observed})} = 0.65, P < 0.001, 95\% \text{ CI} = 0.47, 0.79$; Fig. 5b), indicating the communication ability of these regions during retrieval practice is critical for memory retention outcome.

Furthermore, we examined the relationship between VLPFC network connectivity and intertrial neural pattern distinctiveness in the PPC nodes that has been considered to be an ‘output buffer’ for specific representations of the retrieved content^{36,40}. We observed that connectivity strength of the effective links between the VLPFC and PPC nodes was positively predictive of intertrial neural pattern distinctiveness in the PPC in the final run ($r = 0.29, P = 0.04, 95\% \text{ CI} = 0.01, 0.52$). These results indicate that functional connectivity between the right VLPFC and PPC over retrieval practice is critical to predict intertrial neural pattern distinctiveness in the PPC and long-term retention gains.

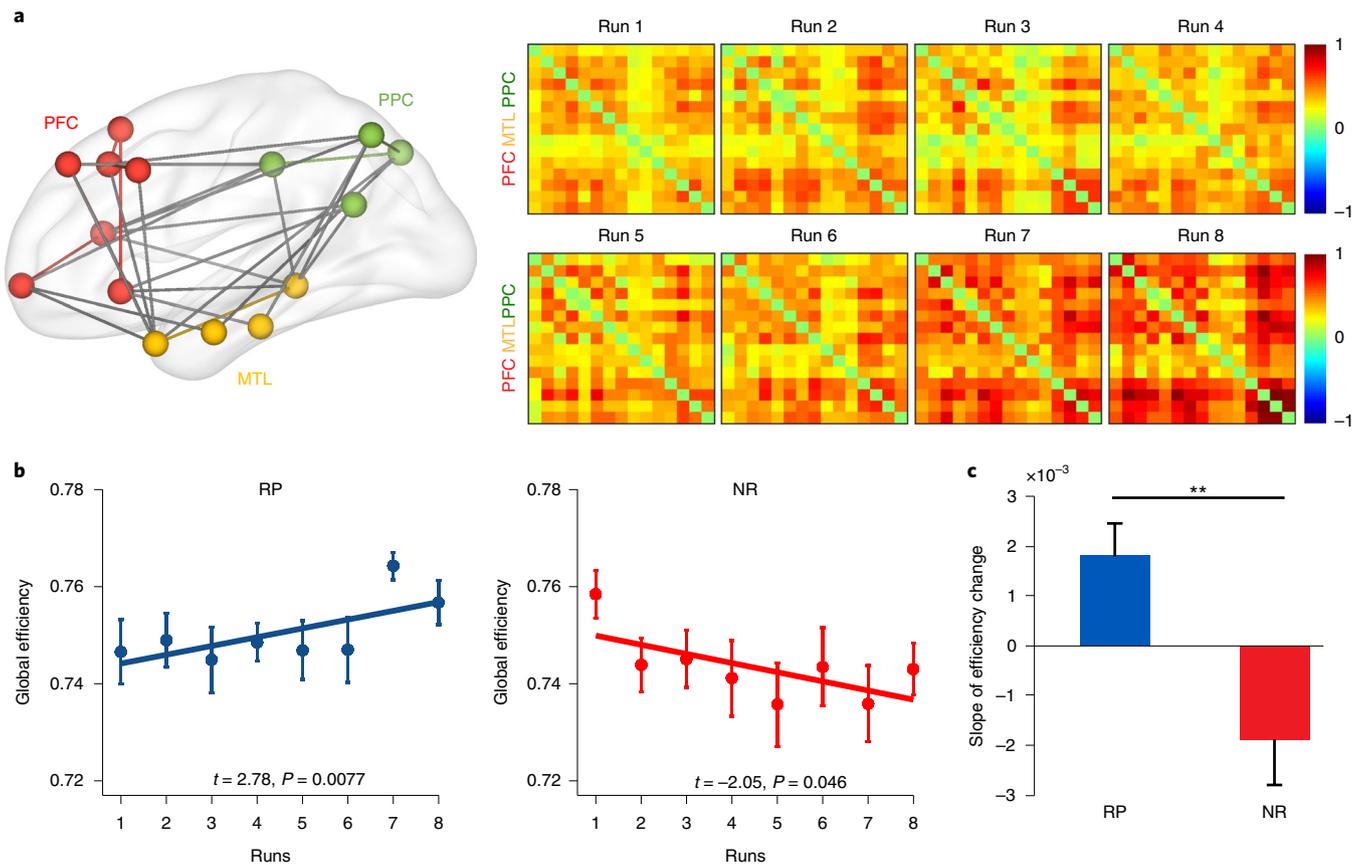


Fig. 4 | Dynamic changes in memory-related neural network configurations over retrieval practice. **a**, A lateral view of memory-related brain networks and network matrices show pairwise interregional connectivity links over eight runs during retrieval practice. Colour bars reflect task-dependent functional connectivity strength based on a gPPI approach. The strength of links shows a gradual increase over the course of retrieval practice. **b**, Linear regression plots show a significant linear increase in network efficiency as a function of eight runs for the RP condition (left, t -test for slopes $t_{(49)} = 2.78, P = 0.008, d = 0.39, 95\% \text{ CI} = 0.15, 0.63$), and a linear decrease over eight runs for the NR condition (right, $t_{(49)} = -2.05, P = 0.046, d = -0.29, 95\% \text{ CI} = -0.53, -0.05$). **c**, Bars depict the slope of dynamic changes in network efficiency, with significantly higher in the RP than NR condition ($t_{(49)} = 3.37, P = 0.001, d = 0.48, 95\% \text{ CI} = 0.18, 0.77$). Note that beta and P values are derived from the least-squares fitting. $**P < 0.01$. Error bars represent s.e.m.

MTL-DLPFC rapid network configurations predict false memory. On the basis of above observed neural distinctiveness in the MTL linked to false memories, we investigated how retrieval-induced changes in the MTL-centric functional networks contribute to subsequent false memories. We conducted a machine learning-based prediction analysis for false memory scores with a focus on the MTL-centric links with prefrontal and parietal nodes over eight runs in the RP condition (Fig. 6a). Given that false memory showed a similar pattern for 30-min and 24-h intervals, we used the general false memory (relative to baseline) scores collapsed across the two intervals for this prediction analysis. This analysis revealed the MTL-based connectivity with the DLPFC emerged as the best predictor for false memories ($r = 0.39, P = 0.02, 95\% \text{ CI} = 0.11, 0.62$; Fig. 6b). The connectivity strength between the hippocampal and DLPFC nodes in the final run was positively predictive of intertrial neural distinctiveness of the final run in the MTL ($r = 0.40, P = 0.008, 95\% \text{ CI} = 0.11, 0.63$; Fig. 6c), but this effect was not evident in the PPC ($r = 0.22, P = 0.15, 95\% \text{ CI} = -0.08, 0.49, BF_{01} = 1.94$).

Moreover, we conducted a prediction analysis for false memories on the basis of the entire 15×15 network data, by training an independent SVR model with leave-one-out cross-validation and feature selection similar to the above analysis for long-term retention. The most stable links that overlapped across iterations to predict false memory outcomes were visualized in Supplementary

Fig. 9a. The prediction values gradually increased over eight-run retrieval practice in the RP condition, which outperformed the NR condition (Supplementary Fig. 7d). Graphic network analyses for selected links revealed a gradual increase in betweenness centrality for the DLPFC, hippocampus, VLPFC and lateral parietal cortex (LPC) (t -test for slopes showed the minimum $t_{(42)} = 5.91, P < 0.001$, Cohen's $d = 0.90, 95\% \text{ CI} = 0.29, 0.59$; Supplementary Fig. 9b–d). Together, these results indicate that the MTL connectivity with the DLPFC during retrieval practice is predictive of false memory outcomes and also associated with intertrial neural distinctiveness in the MTL.

Discussion

By tracking dynamic changes in neural representations and network configurations, we investigated the neurocognitive mechanisms reshaping memories during retrieval practice and subsequent consolidation. Retrieval practice boosted long-term retention after consolidation with a 24-h but not a 30-min delay, and increased false memory in general. Long-term retention gains were correlated with neural distinctiveness in the PPC that increased over retrieval practice, whereas false memory was correlated with unstable neural distinctiveness in the MTL. Retrieval practice gradually enhanced efficiency of memory-related brain networks, and built up PPC and MTL connections with the VLPFC and DLPFC to predict long-term retention gains and false memory. Our findings indicate

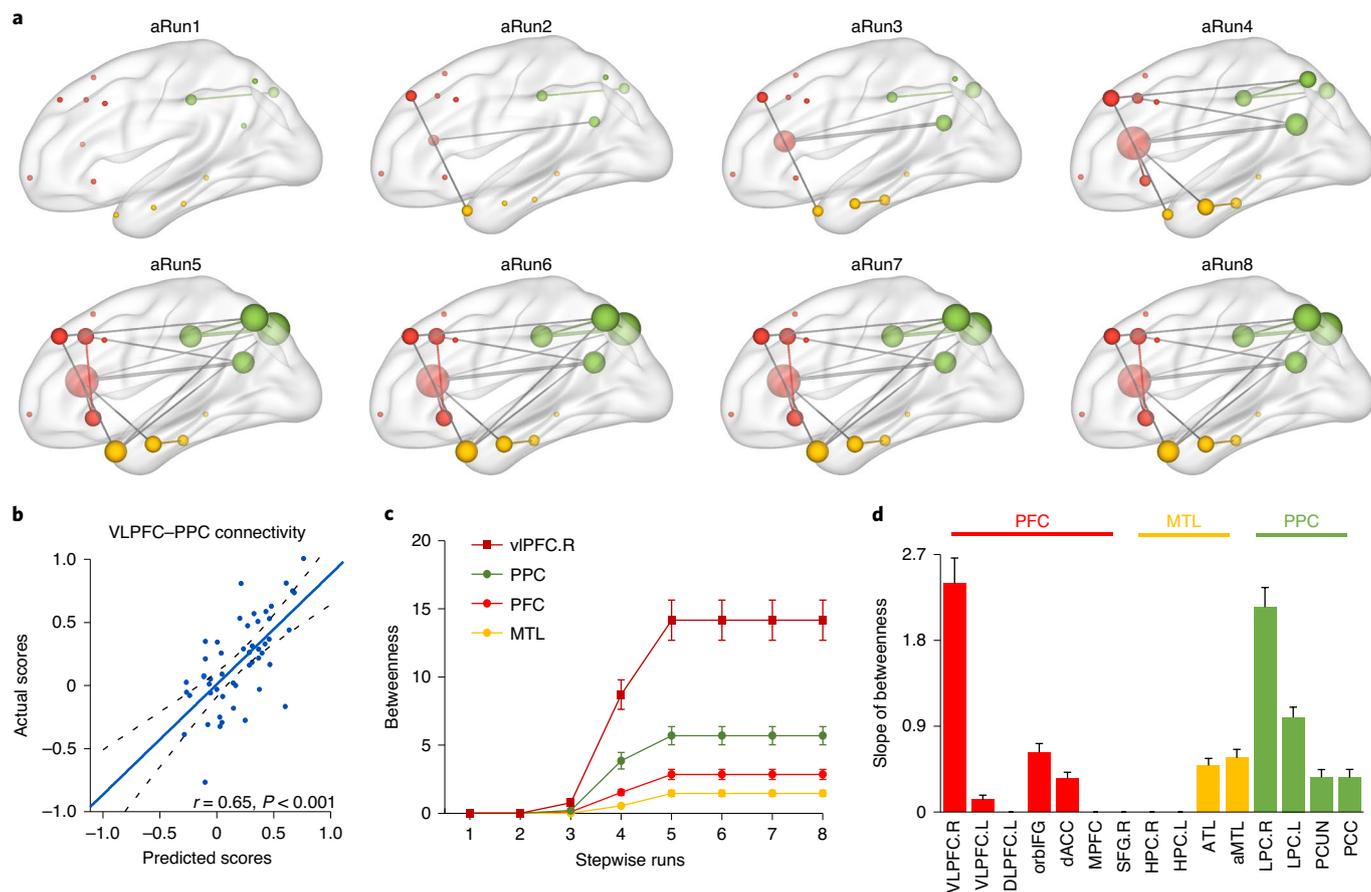


Fig. 5 | Brain network-based prediction of long-term retention gains after consolidation and network reconfiguration over retrieval practice. a, Lateral views of selected links critical for predicting long-retention gains projected onto a glass brain to illustrate dynamic changes in memory-related neural network configurations over eight runs in a cumulative manner (indicated as aRunX) of retrieval practice. Red, green and yellow nodes are ROI in the PFC, the MTL and the PPC, respectively. **b**, Scatter plot showing a robust positive correlation between observed long-term retention gains and predicted outcomes from prediction analysis based on functional connectivity strength of the right VLPFC with PPC nodes during retrieval practice ($r_{(\text{predicted}, \text{observed})} = 0.65$, $P < 0.001$, 95% CI = 0.47, 0.79). **c**, Line plots show betweenness centrality over eight runs for the RP in the VLPFC, MTL, PFC and PPC systems. **d**, Bars depict the slopes of dynamic changes in betweenness centrality over eight runs in 15 nodes for the RP condition. Abbreviations for 15 nodes are listed in the Methods. Error bars represent s.e.m.

that retrieval practice reshapes memory-related neural representations and network configurations to promote long-term retention gains while producing false memory. In the following paragraphs, we will discuss our results on long-term retention gains and false memories, subsequently.

The first aspect is the neurocognitive underpinnings of how retrieval practice promotes long-term retention gains. In line with previous findings of retrieval practice^{1,3}, the long-term retention gains appeared time-dependent in the RP condition only after 24 h. Since such measure was no difference in immediate recall between the RP and NR conditions and no further manipulations took place afterwards, we speculate that subsequent offline systems consolidation is crucial to promote long-term retention gains. This finding does not necessarily contradict recent models showing repeated retrieval as a fast route to memory consolidation²¹. Echoing our observations, recent evidence also suggests that while the formation of memory engrams might be initiated during online practice, consecutive offline consolidation appears necessary to transform them into stable representations for long-term storage^{22,23}. Hence, we speculate that multiple retrieval attempts on the targeted memories might have prioritized or ‘tagged’ these memories as important or future-relevant, turning them into a super-ordinate position for

subsequent consolidation. Being proposed also by Ferreira and colleagues²², this tagging hypothesis could be a possible account (although not excluding other mechanisms) for long-term retention gains after consolidation in the RP, while memories that did not undergo retrieval practice might be tagged as less important and vulnerable to being forgotten later. As we will discuss below, our observed brain-behaviour associations may be accounted for by a process of gradual direct encoding into the neocortical networks over retrieval practice working together with consecutive offline consolidation.

At the neural representation level, retrieval practice led to heterogeneous changes in multi-voxel activity patterns in memory-related brain systems, with the most prominent increase in intertrial neural distinctiveness in the PPC. This measure assessed how fine-tuned neural representations discriminated among individual memories⁴¹, partially analogous to pattern separation^{42,43}. This finding probably indicates that those representations of the retrieved events became increasingly differentiated from each other in the PPC over retrieval practice¹⁴. Critically, the PPC’s neural distinctiveness in the final run in the RP but not NR condition was positively associated with long-term retention gains. This association emerges in the final (rather than the first) run, suggesting the evolution of

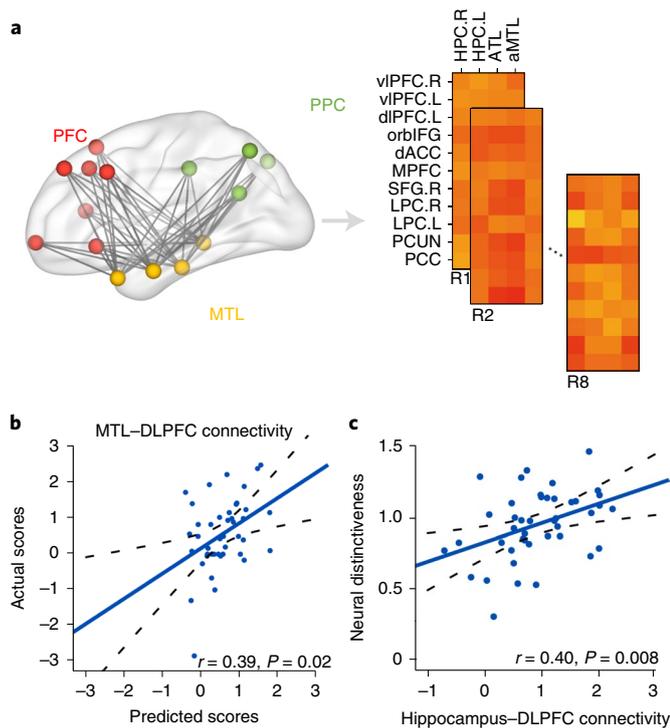


Fig. 6 | MTL-based network prediction of general false memory outcome.

a, A lateral view of the MTL-based functional connectivity with the other prefrontal and parietal nodes of memory-related brain network. Red, green and yellow nodes in the PFC, MTL and PPC, respectively. Network matrices show pairwise interregional connectivity links of the MTL with prefrontal and parietal nodes over eight runs during retrieval practice. Colour bars reflect task-dependent functional connectivity strength. **b**, Scatter plot shows a positive correlation between the actual and predicted false memory outcomes from prediction analysis based on functional connectivity of the MTL and the DLPFC nodes during retrieval practice ($r = 0.39$, $P = 0.02$, 95% CI = 0.11, 0.62). **c**, Scatter plot depicts a positive correlation between hippocampal-DLPFC connectivity and neural distinctiveness in the MTL ($r = 0.40$, $P = 0.008$, 95% CI = 0.11, 0.63). Abbreviations for 15 nodes are listed in the Methods. Error bars represent s.e.m.

fine-tuned neural representations is progressive over retrieval practice. Together with the recent evidence that memory engrams can be detected in the PPC soon after repeated rehearsal²¹, it is thus reasonable to assume that PPC plays a role in tuning memory representations into a differentiated status over retrieval practice. How such online neural tuning process cooperates with offline consolidation, with or without sleep, to promote long-lasting memory needs to be addressed in future studies.

Beyond neural representations, we found that connectivity between large-scale brain networks became strengthened over eight runs, with higher global efficiency in the RP than NR condition. On the basis of the graphic theory of brain networks¹⁴, the strengthening of connectivity with higher efficiency may reflect a gradual build-up of more effective routes over retrieval practice, through which the target information could be reweighted as important for future use. Moreover, our network-based prediction results show that retrieval-induced network reconfigurations are critical to predict long-term retention gains, with rapid growth in betweenness over the initial five runs. The most prominent effect emerged in the right VLPFC. As betweenness reflects the transfer of information flow through a network¹⁵, the VLPFC may thus act as a hub to drive network reconfiguration over retrieval practice. The MTL exhibited a generally lower yet similar trajectory to the PFC and PPC, indicating

constant engagement in retrieval practice. Given long-term retention gains observed after consolidation, the above network reconfigurations may be preparatory for setting up relevant connections to be prioritized for subsequent consolidation into long-term store.

We further found that the right VLPFC connectivity with the PPC during retrieval practice provided the most information to predict long-term retention gains. The VLPFC, a key locus that exerts top-down modulation of episodic retrieval, involves mental elaboration and context updating²⁷. The PPC is considered as an ‘output buffer’ for active representations of retrieved content through converging information from other cortical inputs^{36,40}. Thus, the strengthened VLPFC–PPC connectivity might facilitate reinstatement of the target memory representation anchored in the PPC. With repeated retrieval, the routes between these nodes may have been gradually established, and the target information was thus weighted as a high priority. Indeed, we observed that the VLPFC–PPC connectivity was positively predictive of intertrial neural distinctiveness in the PPC that further predicted better long-term retention gains. Recent rodent models posit that the formation of long-term memory involves early tagging and reweighting of cortical networks that subsequently support the memory^{20,46}. It is thus possible that the VLPFC could drive an evaluation signal to reweight the repeatedly reactivated events as important over retrieval practice, through which representations of the target information in the PPC can be prioritized into a super-ordinate position for subsequent consolidation. Critically, the reweighted memory appeared to mature through subsequent consolidation, as we did not observe memory gains in the immediate recall, probably due to the possibility that they had still not matured into stabilized and discrete engrams²¹. Indeed, a recent study suggests that sleep consolidation is critical to preserve newly formed memory engrams via repeated study²³.

The second aspect relates to how retrieval practice produced false memories at both 30-min and 24-h intervals. Behaviourally, this appears consistent with previous reports that retrieval increases the likelihood of false memory^{7,47}. Based on the episodic context model, the retrieved memory could be updated by new information from the current contexts¹², which could provide additional cues to recall target memories. Integrating new information during retrieval, however, might generate ‘mismatched’ episodic contexts among different memories, thereby leading to false memory⁴⁸. Moreover, multiple retrieval attempts may also involve semantic or gist-based encoding strategies that could also generate false memories^{49–51}. Our findings suggest that in real-world applications such as classrooms and clinical training, relevant strategies should be considered to minimize false memory, for instance, by incorporating feedback after each retrieval^{52,53}.

We found that intertrial neural distinctiveness in the MTL for the RP but not NR trials emerged as the best predictor of false memory. This is reminiscent of the finding that the MTL especially the hippocampus is responsible for false memory⁵⁴. The MTL is crucial for reconstructive processes, contextual binding and recollection in episodic memory⁵⁵. Episodic contexts support recollection, but mismatched contexts may lead to false memory. Thus, higher neural distinctiveness in the MTL might reflect a retrieval-mediated updating of new context information that could cause mismatching during multiple retrieval attempts, thereby producing more false memories in the recall tests. This interpretation accommodates the hippocampus (MTL) as a ‘fast learner’^{56,57}, as false memory emerges in the immediate recall.

At the network level, unlike the VLPFC and PPC that were critical for predicting long-term retention gains, we found the most prominent connectivity links among the MTL and DLPFC nodes predicted subsequent false memories. This suggests that functional coordination between these regions over retrieval practice was responsible for producing false memories. This is in line with

previous findings on the constructive nature of MTL-prefrontal episodic memory systems and the DLPFC involvement in false memories⁵⁴. During retrieval practice, the MTL may be involved in active reconstruction of episodic memories for future use⁵⁸. The DLPFC enables the updating of existing memories with new information³⁴, which could lead to possible interference with the target memories, thereby producing false memories. This interpretation is supported by a positive correlation between hippocampal–DLPFC connectivity and the neural distinctiveness of the MTL in the final run. Hence it appears that retrieval-induced false memory is mainly supported by MTL-centred neural configurations, separating from the networks supporting the long-term retention. One could ask why true and false memories that are not subjectively differentiated during recall, but are associated with distinct neural predictors during retrieval practice. We suggest that the reconstructive nature of episodic memory^{59,60} may provide an explanation for this dissociation. That is, the dissociable neural predictors for long-term retention and false memory likely reflect the involvement of multiple retrieval-induced mnemonic and reconstructive processes that can reshape and alter distinct aspects of memory representations for the multidimensional episodic information of a retrieval event. Together, our findings expand previous research on retrieval practice benefits on memory, by highlighting that active retrieval reshapes episodic memories through rapid reorganization of neural representations and network connectivity to produce false memories.

Nevertheless, our study has some limitations. First, we did not directly assess how well participants engaged in the RP and NR conditions, although the indirect measure of vividness and subsequent memory performance indicated a clear dissociation of our RP and NR manipulations. Second, our design could not dissociate successful retrieval attempts from run-to-run during the practice phase and individual differences in ‘speak-out’ tendency for each trial may influence false memory scoring. Third, the NR might induce potential confounds such as suppression of associated memories^{61,62}. Our data appear to neutralize this possibility, because of no any differences between the NR and baseline conditions. Finally, although sleep is crucial for offline consolidation⁵⁶, and may prioritize memories for retrieval⁶³, our experimental design cannot address this possibility due to a lack of direct comparisons between sleep and wake intervals. Future studies are required to overcome these limitations.

In conclusion, our findings provide new evidence for distinct mechanisms of neural reorganization during retrieval practice that produce two seemingly contradictory outcomes: long-term retention and false memories. Retrieval practice may consist of both a gradual refinement of neural representations and network reconfiguration working together with an offline process of consolidation, leading to subsequent long-term retention and false memories. Future studies may usefully explore this suggestion for the development of interventions and strategies to improve memory in both healthy and clinical populations.

Methods

Participants. Fifty-seven young, healthy college students (32 females, range from 19 to 29 years of age) participated in this study. All participants were right-handed with normal or corrected-to-normal vision, and reported no history of neurological or psychiatric disease. The Institutional Review Board approved the ethical protocol of our study for Human Subjects at Beijing Normal University, and written informed consent was obtained from all participants before the experiment. Due to unexpected data unavailability, the sample size of false memory reduced to 50. Data from seven participants were excluded from further analyses due to excessive head motion during fMRI scanning with root mean squared motion parameters over a voxel's width. This resulted in 56 participants for long-term retention and 50 participants for false memory as the behavioural sample size, with 50 and 43 as their corresponding imaging data sample size, respectively.

Materials. Forty-eight pairs of face–scene associations were used in this study. Faces with neutral expressions were selected from a standardized Chinese face database⁶⁴, which were carefully selected using the criteria previously reported⁶⁴.

We have obtained consents for publication from the owner of the example faces shown in the Fig. 1. Forty-eight complex scenes were selected from the International Affective Picture System with half negative and half neutral in emotional valence (corresponding data are provided in the Supplementary Methods). Faces and scenes had been used in an independent study with a minimal relatedness in content to each other, and matched on luminance⁶⁴. Faces and scenes were randomly paired to create 48 face–scene associations across participants.

Experimental design and procedure. The experiment consisted of three phases: memory acquisition, retrieval practice and two recall tests after 30-min and 24-h intervals. In the acquisition phase outside the scanner, participants were trained to memorize 48 face–scene associations with study-test cycles. For each cycle, each face–scene association was presented for 6 s and participants were instructed to remember face–scene associations for subsequent memory tests. Thereafter, they performed an associative memory recognition test in which a given face in the upper centre of the screen and two scenes on the left and right of the lower screen were displayed. One scene was the correct association while the other one randomly selected from other associations. Participants were asked to select either the left or the right scene as the picture associated with that face, without feedback. After a cycle of associative memory tests was finished, the program scored the participant's performance. The training session would continue until the participant reached at least 90% accuracy.

During the retrieval practice phase inside the scanner, 32 face–scene associations were randomly selected from the acquisition phase and the remaining 16 associations served as the baseline condition and were not presented in this phase. For each trial, a face surrounded by either a blue or red rectangle frame was presented for 4 s. Participants were instructed to engage in either active retrieval practice (that is, RP, cued with a blue rectangle) or passively view the face trying to not retrieve the associated scene (NR, cued with a red rectangle) of the scene associated with the face cue. We chose the NR as a control condition with matched visual stimulation. Thereafter, participants were asked to rate the vividness of each recalled scene on a four-point scale (1 being ‘not at all’ and 4 being ‘extremely’), which could verify the dissociation of RP and NR manipulations (Supplementary Fig. 1). The original scene was not shown, in accordance with many studies on retrieval practice without feedback^{1,3,31}. Trials were jittered with an intertrial interval varying from 2 to 6 s (averaged 4 s with 1 s steps). The entire memory practice phase consisted of eight runs with 2-min breaks between runs, and lasted 36 min in total, with 4.5 min for each run.

In the memory test phase outside the scanner, memory performance for face–scene associations was assessed by two independent cued-recall tests after 30-min (immediate recall) and 24-h (delayed recall) intervals. All 32 faces from the practice phase were randomly split into two halves as cues for the immediate and delayed recall tests respectively. Participants were asked to orally recall and describe the scenes associated of each face cue. It is worth noting that 16 face–scene associations from the memory acquisition phase did not appear during the retrieval practice phase, and were also split into two halves corresponding to immediate and delayed recall tests. The presentation of face cues was randomized at test across participants. For each face cue, participants had a maximum of 30 s to verbally describe the associated scene with audio recording. Two raters who were blind to the purposes of the study, as well as the experimental design, including the RP and NR conditions and the two recall tests. They scored each participant's oral recall independently. When inconsistencies were encountered, a final score of each item was made by further discussion with a final consensus between the raters. All participants reported adequate sleep during the night after retrieval practice, with approximately 8 h of sleep (average 7.68 ± 1.05 h). Both the participants and the experimenter were blind to the purpose and hypotheses of this study.

Memory performance and behavioural analyses. To assess episodic memory for face–scene associations, the raters scored participants' memory performance according to their oral recall of the information that was enough to identify the associated scenes (scored the correct or not), as well as false information recalled. Each participant's memory accuracy (correct proportion) was quantified as a proportion of face–scene associations later remembered from all associations within each condition (RP, NR, baseline) at each testing (30-min, 24-h). Each participant's false memory was quantified as the sum of false information recalled for the complex scenes associated with each cue, and only false information of the correctly recalled associations was counted. We further computed the memory retention and false memory scores for the immediate and delayed tests by subtracting the memory accuracy and their false memory scores from the corresponding baseline condition. Such subtraction could provide a measure of the putative effects for RP and NR conditions, since the baseline trials did not undergo the retrieval practice and hence represented a baseline level of memory. Each participant's long-term retention gain after consolidation was computed by subtracting retention scores in the immediate recall test from that of the delayed recall test. The mean of false memory across the two recall tests was also computed for the RP and NR conditions. Pearson's correlations and prediction analyses were conducted to assess the relationship of retrieval-induced changes in neural representations and network reconfiguration with long-term retention gains and false memory outcomes.

Imaging acquisition. Whole-brain images were acquired on a Siemens Trio 3.0 Tesla magnetic resonance scanner. Functional images were acquired using an echo-planar imaging sequence (37 slices; repetition time (TR), 2,000 ms; time to echo (TE), 30 ms; flip angle, 90°; voxel size, 3.5 × 3.5 × 3.5 mm; field of view, 224 × 224 mm). High-resolution T1-weighted anatomical images were acquired by using a magnetization-prepared-rapid acquisition gradient echo sequence (144 slices; TR, 2,530 ms; TE, 3.39 ms; voxel size, 1.3 × 1.0 × 1.3 mm; flip angle, 7°; field of view, 256 × 256 mm).

Imaging preprocessing. Brain images were preprocessed using Statistical Parametric Mapping toolbox (SPM8, <http://www.fil.ion.ucl.ac.uk/spm>). The first four volumes of functional images were discarded to allow for signal equilibrium. Remaining images were realigned to the mean image of each run and corrected for slice acquisition timing. Subsequently, functional images were coregistered to each participant's grey matter image segmented from corresponding T1-weighted image and spatially normalized into the stereotaxic template of the Montreal Neurological Institute. Finally, images were smoothed using a 6-mm full-width at half-maximum Gaussian kernel.

Univariate general linear model analysis. To assess task-related brain responses during retrieval practice, separate regressors were modelled for trials in the RP and NR conditions, and convolved with the canonical hemodynamic response function (HRF) in SPM8. Each participant's motion parameters were included to regress out potential effects of head movement. We included high-pass filtering using a cutoff of 1/128 Hz to remove high frequency noise and corrections for serial correlations using a first-order autoregressive model (AR(1)) in the general linear model (GLM) framework.

Contrast parameter estimate images for task-related brain responses in RP (or NR) condition generated at the individual-participant level were submitted to subsequent analyses for multi-voxel pattern similarity and fidelity analyses over the course of retrieval practice.

To further assess trial-wise brain responses for the RP and NR trials during the practice phase, each trial was modelled as a separate regressor and convolved with the HRF implemented in SPM8. This resulted in a total of 32 regressors for each run, with 16 RP trials and 16 NR trials. The other parameter settings were the same as the univariate GLM for task-related estimation of brain responses. Contrast parameter estimate images for each RP/NR trial, initially generated at the individual-participant level, were submitted to subsequent analyses for intertrial multi-voxel pattern distinctiveness over eight runs in RP and NR conditions separately.

ROI selection. To define a memory-related brain mask, we used the NeuroSynth platform for large-scale, automated synthesis of fMRI data (<http://neurosynth.org>) with 'memory retrieval' as a search term and generated a reverse inference mask. We then refined the mask using a criterion of a height threshold of $P < 0.001$ ($z > 3.0$) and a spatial extent cluster size of more than 30 voxels. Then, the whole-brain mask was segmented into 15 ROI based on spatially contiguous voxels. These ROI included the left VLPFC, right VLPFC, left DLPFC, orbito-inferior frontal gyrus, dorsal anterior cingulate cortex, medial PFC, right superior frontal cortex, right hippocampus, left hippocampus, left anterior temporal lobe, temporal pole or anterior MTL, right LPC, left LPC, precuneus and posterior cingulate cortex.

Neural pattern similarity. Three multi-voxel neural pattern metrics were computed to characterize changes in retrieval-induced multi-voxel pattern similarity over the course of retrieval practice. The first one referred to neural pattern similarity over eight runs during retrieval practice. The z score map of retrieval-induced neural activation pattern was first obtained from the NeuroSynth platform as a canonical reference by using the overall brain mask defined above. Multi-voxel activity patterns for RP and NR conditions were then separately extracted from the same mask in each run. And then we computed Pearson's correlation coefficients for the respective multi-voxel patterns of RP and NR with the canonical reference map for each run. Thereafter, correlation coefficients were Fisher's z transformed and submitted to compute neural pattern similarity across eight runs for RP and NR separately.

$$S_{\text{similarity}} = \text{Corr}(X_i, Y) \quad i \in [1 : 8]$$

where X_i is condition-related activity patterns of run i for each condition, and Y is multi-voxel activity pattern from the NeuroSynth.

The other metrics, reflecting condition-related neural fidelity and trial-specific neural similarity, are described in the Supplemental Methods.

Intertrial neural pattern distinctiveness. A neural pattern distinctiveness metric was computed to characterize changes in intertrial multi-voxel activity pattern dissimilarity among RP (or NR) trials over the course of eight-run retrieval practice. The multi-voxel activity patterns for each trial in the RP (or NR) condition was extracted from the mask in each run. We then computed their corresponding Pearson's correlation coefficients with activity patterns of other remaining trials from the same condition of the same run. These coefficients

were transformed into Fisher's z scores and averaged, and the average was then subtracted from one to yield a distinctiveness metric for each run. The formula as follows:

$$S_{\text{dist}} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n [1 - \text{Corr}(T_i, T_j)]}{n(n-1)/2}$$

where n is the total number of trials in each condition, T_i is trial-related activity pattern associated with trial i and T_j is trial-related activity pattern for trial j from the same condition of the run. We then fitted the linear function ($y = a \times x + b$, where x denotes the run number from one to eight and y denotes the intertrial neural distinctiveness) to the dynamic neural pattern distinctiveness. Due to the last run (the eighth run in this study) representing the final brain state after memory practice, we used the eighth run for further correlational analyses with memory performance as implemented by previous studies²⁷.

Network construction. Network nodes consisted of the 15 ROI above. For each participant, a 15 × 15 connectivity matrix was created for the RP (or NR) condition in each run by using gPPI analysis. The gPPI approach was widely used to assess task-dependent functional connectivity of a specific seed or ROI with the rest of the brain, after removing potential confounds of overall task activation and common driving inputs⁶⁵. Specifically, mean time series from each seed ROI were extracted and then deconvolved so as to uncover neuronal activity (that is, physiological variable) and multiplied with the task design vector contrasting the RP condition versus the fixation condition (that is, a binary psychological variable) to form a psychophysiological interaction (PPI) vector. This interaction vector was convolved with a canonical HRF to form the PPI regressor of interest. The psychological variable representing task design (RP versus fixation) as well as mean-corrected time series of each seed ROI were also included in the GLM to remove overall task-related activation and the effects of common driving inputs on brain connectivity. To ensure normality, connectivity values of each task condition were Fisher's z transformed. Note that only the voxels within the 15 ROI were included in this analysis to save computational resources. Separate gPPI analyses were conducted for each seed ROI to assess its task-dependent functional connectivity with the remaining ROI.

Network-based brain-behaviour prediction analysis. Network matrices derived from the gPPI analyses were submitted to subsequent brain-behaviour prediction analyses based on machine learning algorithms using the LIBSVM toolbox (<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>). The matrix connectivity data across eight runs were treated as the input variable, and long-term retention gains (or false memories) as the output variable. Separate SVR models were trained using a classic leave-one-out cross-validation approach to predict individual's long-term retention gains or false memory outcomes independently (Supplementary Fig. 7). All feature ranking and selection (Supplementary Fig. 7a,b) were carried out on the training data only, without examining the test data. This approach was iterated to compute the predicted scores for each individual. We then calculated Pearson's correlation coefficients between predicted and observed scores as the prediction accuracy to quantify the strength of the brain-behaviour relationship.

Feature ranking and selection were subsequently performed. For each model, we obtained the predictive weight for each link, which was sorted in descending order. Only links whose weights were above certain thresholds were selected as input features in the model for the follow-up stepwise prediction procedure. Following the convention in the field, we implemented a set of different thresholds and obtained the top 1% of links to achieve the best prediction accuracy (Supplementary Fig. 7a). The statistical significance of prediction accuracy was assessed by a permutation test procedure. For each permutation, we randomly shuffled behavioural scores, and then computed the brain-behaviour prediction accuracy. This procedure was iterated 1,000 times for each run. Thereafter, 1,000 permuted prediction values were sorted in descending order, and the significance P values of 0.01 and 0.05 were respectively computed by dividing the position number (that is, the prediction accuracy located) by 1,000 for each run. Thereafter, two-step prediction procedures were used to characterize the dynamic changes in brain-behaviour prediction values and how the selected links evolved over eight-run retrieval practice. Since the selected features differed slightly from iteration to iteration due to the leave-one-out approach, we first identified the consensus functional connectivity⁶⁶ with stable features (>50%) across iterations^{67,68}. Then, we projected those links back to each run to track their evolutionary trajectories over the progression of eight runs in a cumulative way, and further trained a stepwise prediction model to predict each individual's long-term retention gains (or false memory scores). It is worth noting that these stepwise analyses were conducted for confirmatory purposes to characterize how retrieval-induced network reconfigurations contributed to subsequent memory outcomes, rather than to determine the generalizability for the other samples⁶⁹. Finally, the selected links were submitted for graph theoretical analyses and network visualization by superimposing onto a glass brain template.

Global efficiency. Global efficiency represents the average inverse shortest path length in a network, and is inversely related to the characteristic path length. Global efficiency is defined as

$$E_{\text{global}} = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}$$

where N denotes the total nodes in a network, and G is the network graph, i and j are ROI, d_{ij} denotes the length of the shortest path between node i and j . A brain network with higher global efficiency reflects more efficient information communication across different nodes.

Betweenness. As one of the most frequently used metrics in network analysis, betweenness centrality represents the degree to which information passes through a node. The more information that passes through a node, the higher the influence this node has on a network⁴². It was computed according to the networks with selected features from prediction analysis. Specifically, links that were predictive of an individual's long-term retention gains after consolidation were maintained and the remaining ones were set to 0. Thereafter, the resultant networks were entered into GREYNA for network analysis (<https://www.nitrc.org/projects/gretna/>). The betweenness centrality of node i was defined as follows:

$$b_i = \frac{1}{(n-1)(n-2)} \sum_{\substack{h, j \in N \\ h \neq j, h \neq i, j \neq i}} \frac{\rho_{hj}^{(i)}}{\rho_{hj}}$$

where N is the set of all nodes in the network and n is the number of nodes. ρ_{hj} is the number of shortest paths between h and j , and $\rho_{hj}^{(i)}$ is the number of shortest paths between h and j that pass through i .

Statistical analysis. Statistical testing for behavioural and imaging data was performed using R (v.3.4.1) and MATLAB (v.R2016a), respectively. Values are presented as mean \pm s.e.m., unless indicated otherwise. Repeated ANOVAs and Student's t -tests were used to assess differences between conditions of interest for equal variances, when normality was assumed. The effect size for repeated-measures ANOVA is partial η^2 (indicated as η_p^2). The effect size for paired samples t -tests was calculated using Cohen's d . We reported a 95% confidence interval for t -tests, and 90% confidence interval for F -tests, which two have been considered as equal⁷⁰. Pearson's correlation r was used to assess the relationship between variables. The data met the assumptions of the statistical tests used, that is, normality and equal variances were formally tested. The two-tailed P values are reported for statistical testing, except where otherwise specified. The permutation test was implemented for the prediction analyses. To further quantify the level of evidence for the null hypotheses, we calculated the Bayes factor using the JASP software (v.0.14.1, <https://jasp-stats.org/>). The Bayes factor analyses were based on the default priors for ANOVA and paired t -test design (scale r on an effect size of 0.707).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All of the necessary behavioural and brain imaging data are available from https://github.com/QinBrainLab/2017_RetrievalPractice. Source data are provided with this paper.

Code availability

All of the necessary behavioural and brain imaging codes are available from https://github.com/QinBrainLab/2017_RetrievalPractice.

Received: 18 April 2020; Accepted: 22 July 2021;
Published online: 07 October 2021

References

- Karpicke, J. D. & Roediger, H. L. 3rd The critical importance of retrieval for learning. *Science* **319**, 966–968 (2008).
- Racsmay, M., Conway, M. A. & Demeter, G. Consolidation of episodic memories during sleep: long-term effects of retrieval practice. *Psychological Sci.* **21**, 80–85 (2010).
- Roediger, H. L. 3rd & Butler, A. C. The critical role of retrieval practice in long-term retention. *Trends Cogn. Sci.* **15**, 20–27 (2011).
- Pastotter, B. & Bauml, K. T. Testing enhances subsequent learning in older adults. *Psychol. Aging* **34**, 242–250 (2019).
- Gershman, S. J., Monfils, M. H., Norman, K. A. & Niv, Y. The computational nature of memory modification. *eLife* **6**, e23763 (2017).
- Nadel, L. & Land, C. Memory traces revisited. *Nat. Rev. Neurosci.* **1**, 209–212 (2000).
- Roediger, H. L. III., Jacoby, J. D. & McDermott, K. B. Misinformation effects in recall: creating false memories through repeated retrieval. *J. Mem. Lang.* **35**, 300–318 (1996).
- van den Broek, G. et al. Neurocognitive mechanisms of the 'testing effect': a review. *Trends Neurosci. Educ.* **5**, 52–66 (2016).
- Carpenter, S. K. & DeLosh, E. L. Impoverished cue support enhances subsequent retention: support for the elaborative retrieval explanation of the testing effect. *Mem. Cognition* **34**, 268–276 (2006).
- Carpenter, S. K. Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *J. Exp. Psychol.: Learn., Mem., Cognition* **35**, 1563–1569 (2009).
- Kornell, N., Bjork, R. A. & Garcia, M. A. Why tests appear to prevent forgetting: a distribution-based bifurcation model. *J. Mem. Lang.* **65**, 85–97 (2011).
- Karpicke, J. D., Lehman, M. & Aue, W. R. in *Psychology of Learning and Motivation* Vol. 61 (ed. Ross, B. H.) Ch. 7, 237–284 (Academic Press, 2014).
- Antony, J. W., Ferreira, C. S., Norman, K. A. & Wimber, M. Retrieval as a fast route to memory consolidation. *Trends Cogn. Sci.* **21**, 573–576 (2017).
- Wimber, M., Alink, A., Charest, I., Kriegeskorte, N. & Anderson, M. C. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. *Nat. Neurosci.* **18**, 582–589 (2015).
- Hulbert, J. C. & Norman, K. A. Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cereb. Cortex* **25**, 3994–4008 (2015).
- Ye, Z., Shi, L., Li, A., Chen, C. & Xue, G. Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. *eLife* **9**, 57023 (2020).
- McDaniel, M. A. & Masson, M. E. Altering memory representations through retrieval. *J. Exp. Psychol.: Learn., Mem., Cognition* **11**, 371–385 (1985).
- Yassa, M. A. & Reagh, Z. M. Competitive trace theory: a role for the hippocampus in contextual interference during retrieval. *Front. Behav. Neurosci.* **7**, 107 (2013).
- Epp, J. R., Silva Mera, R., Kohler, S., Josselyn, S. A. & Frankland, P. W. Neurogenesis-mediated forgetting minimizes proactive interference. *Nat. Commun.* **7**, 10838 (2016).
- Tonegawa, S., Morrissey, M. D. & Kitamura, T. The role of engram cells in the systems consolidation of memory. *Nat. Rev. Neurosci.* **19**, 485–498 (2018).
- Brod, S. et al. Fast track to the neocortex: a memory engram in the posterior parietal cortex. *Science* **362**, 1045 (2018).
- Ferreira, C. S., Charest, I. & Wimber, M. Retrieval aids the creation of a generalised memory trace and strengthens episode-unique information. *NeuroImage* **201**, 115996 (2019).
- Himmer, L., Schönauer, M., Heib, D. P. J., Schabus, M. & Gais, S. Rehearsal initiates systems memory consolidation, sleep makes it last. *Sci. Adv.* **5**, eaav1695 (2019).
- Diekelmann, S., Born, J. & Wagner, U. Sleep enhances false memories depending on general memory performance. *Behavioural Brain Res.* **208**, 425–429 (2010).
- Pardilla-Delgado, E. & Payne, J. D. The impact of sleep on true and false memory across long delays. *Neurobiol. Learn. Mem.* **137**, 123–133 (2017).
- Lo, J. C., Chong, P. L. H., Ganesan, S., Leong, R. L. F. & Chee, M. W. L. Sleep deprivation increases formation of false memory. *J. Sleep. Res.* **25**, 673–682 (2016).
- Kuhl, B. A., Dudukovic, N. M., Kahn, I. & Wagner, A. D. Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nat. Neurosci.* **10**, 908–914 (2007).
- Hutchinson, J. B., Uncapher, M. R. & Wagner, A. D. Posterior parietal cortex and episodic retrieval: convergent and divergent effects of attention and memory. *Learn. Mem.* **16**, 343–356 (2009).
- Johnson, M. K., Raye, C. L., Mitchell, K. J. & Ankudowich, E. The cognitive neuroscience of true and false memories. *Neb. Symp. Motiv. Neb. Symp. Motiv.* **58**, 15–52 (2012).
- Kriegeskorte, N. & Kievit, R. A. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn. Sci.* **17**, 401–412 (2013).
- Kuhl, B. A., Shah, A. T., DuBrow, S. & Wagner, A. D. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nat. Neurosci.* **13**, 501–506 (2010).
- Chapados, C. & Petrides, M. Ventrolateral and dorsomedial frontal cortex lesions impair mnemonic context retrieval. *Proc. R. Soc. B: Biol. Sci.* **282**, 20142555 (2015).
- Jiang, J., Wang, S.-F., Guo, W., Fernandez, C. & Wagner, A. D. Prefrontal reinstatement of contextual task demand is predicted by separable hippocampal patterns. *Nat. Commun.* **11**, 2053 (2020).
- Kluehn, L. M., Dandolo, L. C., Jocham, G. & Schwabe, L. Dorsolateral prefrontal cortex enables updating of established memories. *Cereb. Cortex* **29**, 4154–4168 (2019).
- Ranganath, C. et al. Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia* **42**, 2–13 (2004).

36. Wagner, A. D., Shannon, B. J., Kahn, I. & Buckner, R. L. Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* **9**, 445–453 (2005).
37. Bassett, D. S., Yang, M., Wymbs, N. F. & Grafton, S. T. Learning-induced autonomy of sensorimotor systems. *Nat. Neurosci.* **18**, 744–751 (2015).
38. Dresler, M. et al. Mnemonic training reshapes brain networks to support superior memory. *Neuron* **93**, 1227–1235 e1226 (2017).
39. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *NeuroImage* **52**, 1059–1069 (2010).
40. Sestieri, C., Shulman, G. L. & Corbetta, M. The contribution of the human posterior parietal cortex to episodic memory. *Nat. Rev. Neurosci.* **18**, 183–192 (2017).
41. Karlsson Wirebring, L. et al. Lesser neural pattern similarity across repeated tests is associated with better long-term memory retention. *J. Neurosci.* **35**, 9595 (2015).
42. Sahay, A., Wilson, D. A. & Hen, R. Pattern separation: a common function for new neurons in hippocampus and olfactory bulb. *Neuron* **70**, 582–588 (2011).
43. Colgin, L. L., Moser, E. I. & Moser, M.-B. Understanding memory through hippocampal remapping. *Trends Neurosci.* **31**, 469–477 (2008).
44. Tanimizu, T. et al. Functional connectivity of multiple brain regions required for the consolidation of social recognition memory. *J. Neurosci.: Off. J. Soc. Neurosci.* **37**, 4103–4116 (2017).
45. Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35–41 (1977).
46. Lesburguères, E. et al. Early tagging of cortical networks is required for the formation of enduring associative memory. *Science* **331**, 924 (2011).
47. McDermott, K. B. Paradoxical effects of testing: repeated retrieval attempts enhance the likelihood of later accurate and false recall. *Mem. Cognition* **34**, 261–267 (2006).
48. Loftus, E. F. Planting misinformation in the human mind: a 30-year investigation of the malleability of memory. *Learn Mem.* **12**, 361–366 (2005).
49. Qin, S., Hermans, E. J., van Marle, H. J. F. & Fernández, G. Understanding low reliability of memories for neutral information encoded under stress: alterations in memory-related activation in the hippocampus and midbrain. *J. Neurosci.* **32**, 4032 (2012).
50. Zhu, B. et al. Multiple interactive memory representations underlie the induction of false memory. *Proc. Natl Acad. Sci. USA* **116**, 3466 (2019).
51. Chadwick, M. J. et al. Semantic representations in the temporal pole predict false memories. *Proc. Natl Acad. Sci. USA* **113**, 10180–10185 (2016).
52. Jou, J. & Foreman, J. Transfer of learning in avoiding false memory: the roles of warning, immediate feedback, and incentive. *Q. J. Exp. Psychol.* **60**, 877–896 (2007).
53. McConnell, M. D. & Reed Hunt, R. Can false memories be corrected by feedback in the DRM paradigm? *Mem. Cognition* **35**, 999–1006 (2007).
54. Schacter, D. L. & Slotnick, S. D. The cognitive neuroscience of memory distortion. *Neuron* **44**, 149–160 (2004).
55. Jeong, W., Chung, C. K. & Kim, J. S. Episodic memory in aspects of large-scale brain networks. *Front. Hum. Neurosci.* **9**, 454 (2015).
56. Diekelmann, S. & Born, J. The memory function of sleep. *Nat. Rev. Neurosci.* **11**, 114–126 (2010).
57. McClelland, J. L., McNaughton, B. L. & O'Reilly, R. C. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Rev.* **102**, 419–457 (1995).
58. Wimber, M., Rutschmann, R. M., Greenlee, M. W. & Bäuml, K.-M. Retrieval from episodic memory: Neural mechanisms of interference resolution. *J. Cogn. Neurosci.* **21**, 538–549 (2009).
59. Schacter, D. L., Addis, D. R. & Buckner, R. L. Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.* **8**, 657–661 (2007).
60. Moscovitch, M., Cabeza, R., Winocur, G. & Nadel, L. Episodic memory and beyond: the hippocampus and neocortex in transformation. *Annu. Rev. Psychol.* **67**, 105–10 (2016). Vol 67.
61. Liu, Y. et al. Memory consolidation reconfigures neural pathways involved in the suppression of emotional memories. *Nat. Commun.* **7**, 13375 (2016).
62. Anderson, M. C. et al. Neural systems underlying the suppression of unwanted memories. *Science* **303**, 232–235 (2004).
63. Stickgold, R. & Walker, M. P. Sleep-dependent memory triage: evolving generalization through selective processing. *Nat. Neurosci.* **16**, 139–145 (2013).
64. Qin, S. et al. Probing the transformation of discontinuous associations into episodic memory: an event-related fMRI study. *NeuroImage* **38**, 212–222 (2007).
65. McLaren, D. G., Ries, M. L., Xu, G. & Johnson, S. C. A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *NeuroImage* **61**, 1277–1286 (2012).
66. Dosenbach, N. U. et al. Prediction of individual brain maturity using fMRI. *Science* **329**, 1358–1361 (2010).
67. Dai, Z. et al. Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3). *NeuroImage* **59**, 2187–2195 (2012).
68. Zeng, L.-L. et al. Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain* **135**, 1498–1507 (2012).
69. Kohoutová, L. et al. Toward a unified framework for interpreting machine-learning models in neuroimaging. *Nat. Protoc.* **15**, 1399–1435 (2020).
70. Steiger, J. H. Beyond the *F* test: effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psych. Methods* **9**, 164–182 (2004).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant nos. 32130045, 31522028, 81571056 and 82021004), the Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning (grant no. CNLZD1503), the Major Project of National Social Science Foundation (grant nos. 19ZDA363 and 20&ZD153) and the Fundamental Research Funds for the Central Universities. J.W. was supported by the International Postdoc Exchange Program in China, and the Special fund (grant no. 2018T110060) from China Postdoctoral Science Foundation Grant. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. We thank Y. Liu and W. Lin for their assistance in conducting the experiment and data collection, and we thank Z. Cui for his advice on data analysis. We also thank F.D. Weber for his valuable comments for the paper.

Author contributions

S.Q. conceived the experiment. B.X. performed data collection and analysis. L.Z., J.W., C.B. and L.H. performed data analysis. L.Z., J.W., P.J.B. and S.Q. wrote the paper. All authors contributed to data discussion and interpretation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41562-021-01188-4>.

Correspondence and requests for materials should be addressed to Shaozheng Qin.

Peer review information *Nature Human Behaviour* thanks Isabella Wagner and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection SIEMENS MAGNETOM 3.0 Tesla TrioTim syngo; E-prime 2.0.10

Data analysis Matlab2015b; SPM8 for fMRI data analysis; LIBSVM toolbox for prediction analysis; GRETNA for network analysis; Workbench, BrainNet Viewer and ggplot2 for visualization; the template was obtained from the NeuroSynth website; JAPS for Bayes factor analyses

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data and codes from this study are available from https://github.com/QinBrainLab/2017_RetrievalPractice, and brain imaging data are available from the corresponding author on request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	In total, there were 57 subjects participated this study. Due to some unexpected data unavailability, in the end, here were 56 data for long-term retention and 50 data for false memory as the behavioral sample size, with 50 and 43 as their corresponding imaging data sample size, respectively. It is a relatively large sample size in neuroimaging studies based on prior experience of the investigators with similar experiments previously published.
Data exclusions	Data from seven participants were excluded from further analyses due to excessive head motion during fMRI scanning with root mean squared motion parameters over a voxel's width.
Replication	Our study undertook several steps to improve reproducibility. First, we had a relatively large sample size (N=57) to gain the robustness and the stability of our behavioral and neuroimaging findings. Second, we opted experimental designs to test the robust benefits of retrieval practice that have been shown by previous behavioral studies. Third, we employed both conventional and innovative analytic approaches to test the significance of our behavioral and neuroimaging data, including support vector regression, machine learning based leave-one-out cross validation, non-parametric permutation test.
Randomization	It's a within-subject design without participants allocation. Participants were recruited randomly.
Blinding	It's a double-blind experiment.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	Fifty-seven young, healthy college students (32 females, range from 19 to 29 years old) participated in this study. All participants were right-handed with normal or corrected-to-normal vision, and reported no history of neurological or psychiatric disease.
Recruitment	Participants were recruited by advertisement and flyers.
Ethics oversight	Beijing Normal University

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	Event-related design
Design specifications	8 runs; 32 trials per run; the length of each trial: 1s fixation + 4s conditional operation + 1s rating+ 2-6s ITI
Behavioral performance measures	Whether subjects were performing the task as instructed was confirmed by the difference of vividness rating. Retrieval practice yielded significantly higher vividness rating (mean \pm S.D., 3.21 ± 0.40) as compared to control condition (1.73 ± 0.26) ($t(56) = 26.55$, $p < 0.001$).

Acquisition

Imaging type(s)	functional and structural
Field strength	3.0 Tesla
Sequence & imaging parameters	Functional images were acquired by using an echo-planar imaging sequence (37 slices; TR, 2000ms; TE, 30ms; flip angle, 90°; voxel size, $3.5 \times 3.5 \times 3.5$ mm; FOV, 224×224 mm). High-resolution T1-weighted anatomical images were acquired by using a magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence (144 slices; TR, 2530ms; TE, 3.39ms; voxel size, $1.3 \times 1.0 \times 1.3$ mm; flip angle, 7°; FOV, 256×256 mm).
Area of acquisition	Whole brain
Diffusion MRI	<input type="checkbox"/> Used <input checked="" type="checkbox"/> Not used

Preprocessing

Preprocessing software	SPM8
Normalization	Functional images were co-registered linearly to each participant's gray matter image segmented from corresponding high-resolution T1-weighted image and spatially non-linearly normalized into the stereotactic template of the Montreal Neurological Institute (MNI)
Normalization template	Montreal Neurological Institute (MNI152)
Noise and artifact removal	Each participant's motion parameters from the realignment procedure were included to regress out potential effects of head movement on brain response. We included high-pass filtering using a cutoff of 1/128hz to remove high frequency noise and corrections for serial correlations using a first-order autoregressive model (AR(1)) in the GLM framework.
Volume censoring	The first 4 volumes of functional images were discarded for signal equilibrium.

Statistical modeling & inference

Model type and settings	Univariate general linear model (GLM) :separate regressors of interest were modeled for RP and NR conditions, and convolved with the canonical hemodynamic response function (HRF) implemented in SPM8. In addition, each participant's motion parameters from the realignment procedure were included to regress out potential effects of head movement on brain response. Trial-wise estimation of brain response: each cue was modeled as a separate regressor and convolved with the HRF implemented in SPM8. This resulted in a total of 32 regressors for each run. The other parameter settings were the same as above univariate GLM for task-related estimation of brain responses. Neural representation stability: The z-score map of neural activation pattern was first obtained from the Neurosynth platform as a canonical reference. The multivoxel activity patterns on each run were separately extracted from the same mask, and then computed their corresponding Pearson's correlation coefficients with the canonical reference map respectively. Thereafter, correlation coefficients were Fisher's z-transformed and submitted to compute neural representation stability. Neural representation distinctiveness: The multivoxel activity pattern for each trial during retrieval practice was extracted from the mask in each run. We then computed their corresponding Pearson's correlation coefficients with neural activity patterns of other remaining trials from the same condition of the same run. These coefficients were transformed into Fisher's z-scores and averaged, then the average was subtracted from 1 to yield the distinctiveness metric for each run.
Effect(s) tested	Standard two-tailed t-tests are used to determine the most of significant differences. Prediction accuracy was examined by permutation test.
Specify type of analysis:	<input type="checkbox"/> Whole brain <input type="checkbox"/> ROI-based <input checked="" type="checkbox"/> Both

Anatomical location(s) A reverse inference mask at the whole brain level was generated from the Neurosynth platform by using search term "memory retrieval", this memory-related mask included the left ventral lateral prefrontal cortex (l_VLPFC), right ventral lateral prefrontal cortex (r_VLPFC), left dorsal lateral prefrontal cortex (l_DLPFC), orbito-inferior frontal gyrus (orbIFG), dorsal anterior cingulate cortex (dACC), ventral medial prefrontal cortex (vmPFC), right superior frontal cortex (r_SFG), right hippocampus (r_HPC), left hippocampus (l_HPC), anterior temporal lobe (ATL), anterior medial temporal lobe (aMTL), right lateral parietal cortex (r_LPC), left lateral parietal cortex (l_LPC), precuneus (PrC) and posterior cingulate cortex (PCC).

Statistic type for inference
(See [Eklund et al. 2016](#))

Voxel-wise

Correction

Permutation test

Models & analysis

- | n/a | Involvement in the study |
|--------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Functional and/or effective connectivity |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Graph analysis |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Multivariate modeling or predictive analysis |

Functional and/or effective connectivity

Task-related functional connectivity using full name of (gPPI)

Graph analysis

weighted graph; subject-level; node summaries; participation coefficient; betweenness

Multivariate modeling and predictive analysis

RSA in conditional and trial-wise level; brain-behavior prediction analyses using the LIBSVM toolbox (Support Vector Regression, SVR); features extraction based weights; leave-one-out approach; permutation test that prediction procedure using random data was iterated 1000 times