Neural Traces of Forgotten Memories Persist in Humans and are Behaviorally Relevant

Tom Willems,[1,6] * Konstantinos Zervas,[1] Finn Rabe,[1,4] Andrea Federspiel,[2,3] Katharina Henke[1,5]

[1]University of Bern, Institute of Psychology, Bern, Switzerland

[2]Department of Diagnostic and Interventional Neuroradiology, Inselspital, University Hospital Bern, University of Bern, Bern, Switzerland

[3]Translational Imaging Center Bern, Swiss Institute for Translational and Entrepreneurial Medicine, University of Bern, Bern, Switzerland

[4]Department of Adult Psychiatry and Psychotherapy, University of Zurich, Zurich, Switzerland

[5]Senior author

[6]Lead contact

*Correspondence: tom.willems@unibe.ch

SUMMARY

For a long time, forgetting has been taken as the dissipation of the neural memory traces (engrams). However, recent engram research in mice suggests that the engrams of forgotten memories do persist. This raises the question whether engrams underlying human episodic memories also persist despite forgetting? And do forgotten memories influence human behavior implicitly? To address this question, 40 men and women learned 96 face-object pairs. Using high-resolution functional magnetic resonance imaging at 7 Tesla we mapped the fate of the 96 memories at the systems level from learning to a 30-minute and onward to a 24-hour memory test. Upon each retrieval attempt, participants indicated whether they remembered or forgot the memory. Univariate and multivariate analyses of the functional brain data revealed that the engrams of forgotten memories remain implemented in the episodic memory network and continue to influence the accuracy of guessing responses at test. Overnight, the engrams of forgotten memories became implemented more deeply within bilateral hippocampus, while consciously accessible memories were neocorticalized overnight. The engrams of both consciously accessible and inaccessible (forgotten) memories were shifted from the 30-minute to the 24-hour retrieval within the right hippocampus and anterior cingulate gyrus such that pattern dissimilarities supported the correct retrieval responses at 24 hours. This is evidence that forgotten human episodic memories remain implemented in the episodic memory network and continue to influence human behavior implicitly.

INTRODUCTION

Most neuroscientific research focusses on learning and remembering, while the nature of forgetting has remained overlooked. Learning involves the strengthening of connections between coactivated neurons that are part of a neuronal ensemble (memory trace or engram), which is associated with a particular memory.[1] Memories are encoded in sparse neural ensembles distributed across the entire brain.[2] Following learning, engram cells that were active during learning are reactivated, which supports the strengthening (consolidation) of the memory. Forgetting is the flip side of memory consolidation. A

common view is that forgetting is due to the dissipation of the entire engram due to neurogenesis, interference, and due to the brain's tendency to degrade molecular and cellular memory traces leading to the disconnection of engram cells from the engram circuit.[3] Recent engram research in rodents, however, suggests that it is unlikely that the engram of forgotten memories vanishes completely.[2] Instead, forgetting may occur due to a reduced accessibility of engrams that silently persist. Engram accessibility may depend on the intactness of the engram itself, on the systemic state of the individual, and on the quality of the retrieval cues.[4] For example, Guskjolen and colleagues[5] found in young mice with natural infantile amnesia that optogenetic stimulation of tagged hippocampal neurons during contextual fear learning could recover forgotten memories for up to 3 months. Also, Yates et al.[6] found that human infantile amnesia is not a failure of encoding, but rather a failure of conscious retrieval of the memories that turn inaccessible due to post-encoding processes. Bolsius et al[7] et al. found that sleep-deprivation-induced amnesia in mice could be reversed by both optogenetic stimulation and the drug roflumilast, which acts by preventing the breakdown of enzymes involved in long-term potentiation. In adult mice, Autore and colleagues[8] showed that although natural retroactive interference during learning resulted in forgetting and decreased engram cell reactivation in the retrieval situation, optogenetic stimulation of previously labeled engram cells enabled memory expression. These discoveries beg the question whether hippocampal and neocortical engram components underlying adult, human episodic memories would also persist despite forgetting and – if so – whether the forgotten memories associated with a persistent engram would continue to influence human behavior implicitly. Here, we tested these hypotheses making use of anterograde and retrograde interference during learning to increase forgetting rates in healthy young men and women. Indeed, much forgetting is likely due to interference – i.e., multiple memories encoded closely in time creating competition during consolidation.[3,9,10] But why should the traces of forgotten memories remain implemented in the brain? Perhaps because the forgotten memories may regain importance for behavioral adjustment and survival when anticipating or being confronted with a similar situation later in life.

On a cellular level, the degree of reinstatement is dependent on the strength of connectivity between engram cells.[11] Synaptic consolidation provides for an active strengthening of the connectivity between engram neurons within an engram component via cellular mechanisms like long-term-potentiation.[12,13] Systems consolidation describes a time-dependent change of episodic memory representation. These changes are characterized by a strengthening of neocortical, mostly medial prefrontal, engram components.[14,15] Over time, retrieval tends to (but not always does) become independent from the hippocampal engram component.[16–18] Because engrams undergo changes at multiple levels of neural organization from synapses to cells to neuronal ensembles, it is possible to study engrams at the systems level in humans with high-resolution functional magnetic resonance imaging (fMRI) and multivariate analyses.[19] The degree of engram reinstatement within human neocortex and hippocampus has been found to predict retrieval success and the amount of information retrieved.[20–22] Besides reinstatement-

related similarity, memory accuracy might also be linked to a transformation of the trace reflected by distinct activation patterns within the hippocampus and dissimilarity between the learning and retrieval events.[23–25] This dissimilarity of learning- and retrieval-related activation patterns may be caused by the rapid transformation of cell-ensemble dynamics in the hippocampus,[26–28] by the reconsolidation of memories following a reactivation,[29] and by pattern separation processes that provide for an orthogonal representation of similar memories to prevent interference.[25,30]

Learning- and retrieval-related activation patterns that underlie forgotten rather than remembered adult human episodic memories have not yet been examined to our knowledge. The reason for this neglect might be conceptual and definitional in nature: If forgetting from episodic memory were associated with persisting engrams in the same neural circuits underlying forgotten memories, which continue to influence behavior implicitly, this would entail that episodic retrieval is not all (conscious) or none (absent). Yet, unconscious episodic memory retrieval does not exist in the framework of traditional models of human memory because these models link episodic memory to consciousness.[31–33] Accordingly, forgetting/remembering is all or none. Newer views of episodic memory allow for unconscious episodic memory formation and retrieval.[34–36] These views open the field conceptually to studying all shades of forgetting from losing conscious access to finally also losing unconscious access to an episodic memory. The finding of persistent neural engrams of consciously acquired episodic memories which subsequently lost conscious accessibility would therefore inform models of human episodic memory.

There is some animal and human evidence indicating that memory traces of forgotten episodic memories might indeed influence behavior implicitly. For example, mice that had learned a Morris water maze as infants forget the location of the platform as they grow up but are significantly quicker at relearning the location of the platform than naive adult mice.[37] The memories of previously learned and then forgotten images may still be recognized implicitly at 24 hours following learning and yield fMRI-measured brain activation changes in regions partly overlapping with regions activated during conscious image recognition.[38] Importantly, there is substantial evidence for the feasibility of an unconscious retrieval of subliminally (unconsciously) encoded or sleep-encoded associative memories in humans,[39–43] bolstering the claim that episodic learning and retrieval can proceed outside conscious awareness.

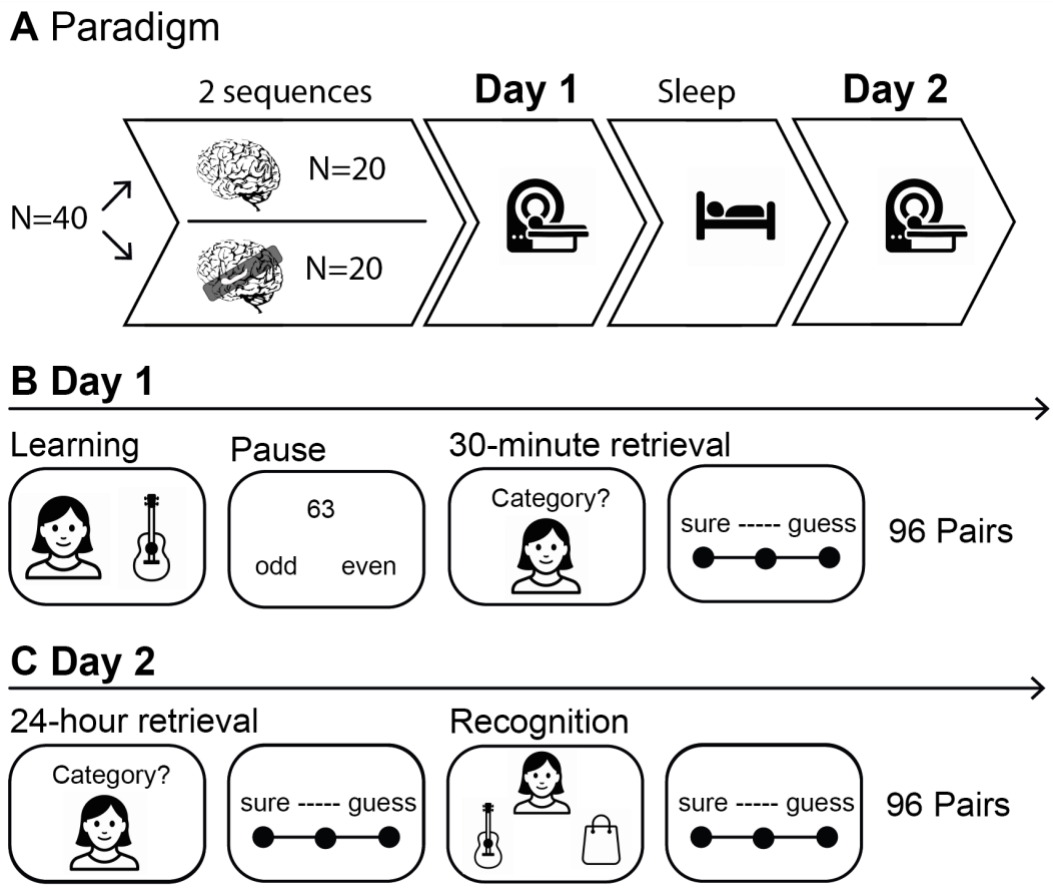


Figure 1. Experimental design.
A) *Overall procedure. Forty participants were randomly assigned to one of the two functional magnetic resonance imaging (fMRI) sequences, one fMRI sequence covering the whole brain and the other covering only the medial temporal lobe. Participants took part in two (f)MRI sessions, separated by 24 hours including one night of sleep.*
B) Day 1 procedure. Participants learned naturalistic face-object pairs (shown here as pictograms for illustration) and immediately retrieved the pairs with the category retrieval task (immediate retrieval not shown in the figure) to assess whether successful learning occurred. Next, participants had a short pause of 5 minutes, during which they remained in the scanner and performed an odd-even task preventing them from actively rehearsing the face-object pairs. Following the short pause, participants took the 30-minute category retrieval. They chose the face-associated object category (organic or inorganic) and rated the confidence of their response (sure, unsure, guess). Due to the short pause and the pseudo-random shuffling of the face-object pairs, the interval between the one-trial learning and the 30-minute category retrieval was constant at 30 minutes for each pair.
C) Day 2 procedure. Repetition of the category retrieval at 24 hours including the confidence rating, followed by a 2-alternative forced-choice object recognition task and another confidence rating.

In this translational endeavor, we sought to provide an account of the engrams of forgotten memories providing evidence that they remain implemented in the human hippocampus and neocortex and continue to guide behavior through their reactivation. To this end, the neural activity of 40 healthy young adults was imaged under high-resolution fMRI at 7T. The increased signal-to-noise ratio of 7T fMRI improves the spatial resolution, which enhances the measurement of the (dis)similarity of activity patterns for individual memory traces over time.[19] The large number of associations rapidly learned was intended to maximize forgetting by anterograde and retrograde interference.[9,10,44,45] Participants were imaged once more after 24 hours, a delay during which most of human forgetting occurs.[46–48] Half of participants (N = 20) were

examined with the small FOV fMRI sequence, focusing on the bilateral medial temporal lobe (MTL; 0.8 mm isotropic voxel size) to target the hippocampal engram components, and half (N = 20) with the whole-brain fMRI sequence (1.1 mm isotropic voxel size) to target neocortical engram components. Engram components were traced using univariate and multivariate (representational similarity) fMRI analyses at the enhanced signal-to-noise ratio of 7T fMRI. Remembered and forgotten memories were probed separately and compared to uncover their underlying brain activations.

During learning, we presented a series of 96 face-object combinations, each combination being shown only once for rapid learning (Figure 1). The presented objects were sampled from two categories, organic (e.g., kiwi) and inorganic (e.g., stapler). Memories were prompted with two retrieval tasks: Cued by the image of a previously presented face, participants indicated whether the face-associated object was organic or inorganic (category retrieval) and cued by the same face plus the original object and a foil object, participants chose the object that was originally presented with the face (2-alternative forced-choice recognition). Category retrieval followed the learning of a face-object combination immediately to ensure that learning was successful, and it was repeated 30 minutes and 24 hours following learning to track forgetting. The recognition task was only given at the 24 hours test following category retrieval. After each retrieval/recognition response, participants indicated whether their just given response was (1) a pure guess, (2) an unsure response, or (3) a sure response. All responses that were neither guess nor sure qualified for an unsure response. We evaluated only the extremes, i.e., the guess (forgotten) and the sure (consciously remembered) responses comparing their underlying brain activations.

RESULTS

Behavioral results

Because the sample consisted of young and healthy men and women, there were almost no false memories in the sense of incorrect sure responses. On an immediate category retrieval, which was done to ensure correct memory formation, participants retrieved the object categories with a mean accuracy of 98% (SD = 2%). All responses were rated ‘sure’ (consciously retrieved). Those few trials that yielded an incorrect response were excluded from the subsequent analyses.

Paired-sample, two-tailed t-tests revealed a significant drop in the frequency of sure responses from 30-minute (M = 21.7%, SD = 14.2%) to 24 -hour (M = 21.6%, SD = 15.3%) retrieval ($t(39) = 5.41$, $p < 0.001$). In contrast, the frequency of guess responses rose from 30-minute (M = 39.8%, SD = 15.6%) to 24-hour (M = 50.8%, SD = 20.07%) retrieval (paired t-test, two-tailed, $t(39) = -5.36$, $p < 0.001$). The frequency of unsure responses decreased between these retrieval time-points (paired t-test, two-tailed, M_{30min} = 33.1%, SD_{30min} = 13.4%, M_{24h} = 27.6% SD_{24h} = 15.8%, $t(39) = 3.07$, $p = 0.003$). On the recognition task given at 24 hours following the category retrieval, paired t-tests revealed that the frequency of sure responses almost doubled from the category retrieval at 24-hours to the recognition task (M_{recog} = 44.9%, SD_{recog} = 22.9%, M_{24h} = 21.6%, SD_{24h} = 15.3%, paired t-test, two-tailed, $t(39) = -10.778$, $p < 0.001$). This suggests

that many of the face-object combinations that were no longer consciously accessible on the category retrieval task regained conscious access with an additional cue. Conversely, the number of guess responses decreased by half (paired t-test, two-tailed, M_{24h} = 50.8% SD_{24h} = 20.07%, M_{recog} = 27.1%, SD_{recog} = 17.3%, $t(39) = 9.63$, $p < 0.001$), while the number of unsure responses remained unchanged (paired t-test, two-tailed, M_{24h} = 27.6% SD_{24h} = 15.8%, M_{recog} = 28.0%, SD_{recog} = 15.2% $t(39) = -0.15$, p = 0.88). Behavioral data for all 40 participants are summarized in Figure 2 and Table S1.

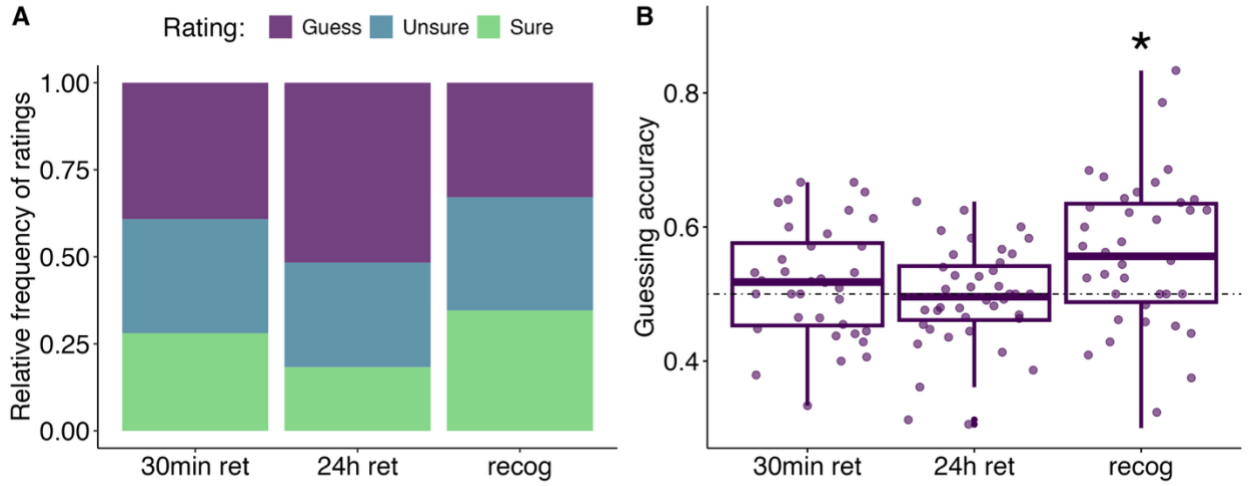


Figure 2. Behavioral results.
A) *Relative frequencies (bar sections) of confidence judgements across retrieval tasks. Frequency of sure responses drops at 24-hours retrieval and rises again during recognition. Frequency of guess responses shows the opposite pattern of change. Frequency of unsure responses drops after 24 hours and remains constant between the 24-hour category retrieval and recognition.*
B) Guessing accuracy (percentage correct of all guess responses) in the retrieval tasks and the recognition task. Above chance accuracy was observed only for guess responses given on the recognition task (p = 0.005). These results include all 40 participants.
Abbreviations: ret = retrieval; recog = recognition. Note: dotted line represents chance level of 0.5

The retrieval accuracy on the 30-minute and 24-hour category retrieval was almost at ceiling level for the sure responses (M_{30min} = 87.8%, SD_{30min} = 10.1%; M_{24h} = 89.6%, SD_{24h} = 13.5%) but not better than chance for the guess responses (M_{30min} = 50.1 %, SD_{30min} = 10%; μ = 50%, one-sample, two tailed t-test, $t(39) = 0.05$, p = 0.95; M_{24h} = 49.4%, SD_{24h} = 7.1%; μ = 50%, , one-sample, two tailed t-test, $t(39) = -0.4$, p = 0.64). The unsure responses yielded an accuracy above chance level at the 30-minute category retrieval (M_{30min} = 56.8%, SD_{30min} = 10.3%; μ = 50%, one-sample, two tailed t-test, $t(39) = 4.13$, p = 0.0002) and at the 24-hour category retrieval (M_{24h} = 54.1%, SD_{24h} = 9.5%; μ = 50%, one-sample, two tailed t-test, $t(39) = 2.7$, p = 0.01). The recognition accuracy for sure responses was at ceiling level (M_{recog} = 94%, SD_{recog} = 12.9%), significantly better than chance for the unsure responses (M_{recog} = 77.2%, SD_{recog} = 12.5%; one-sample, two tailed t-test, μ = 50%, $t(39) = 14.1$, $p < 0.0001$), and significantly better than chance for the guess responses (M_{recog} = 57.5%, SD_{recog} = 14.1%; one-sample, two tailed t-test, μ = 50%, $t(39) = 3.35$, p = 0.005). The Bayes Factor provided moderate evidence in favor of the alternative

model (BF_{10} = 7.4). Hence, even when participants were only guessing which object was presented along with the face, their intuition guided them with a better-than-chance accuracy. These analyses included all 40 participants.

fMRI results from univariate analyses

The height threshold was set to p = 0.001, uncorrected for multiple comparisons, and the cluster extent threshold was set to at least 10 voxels. Note that for the purpose of preserving transparency about the sub-threshold results[49], some figures are presented at a height threshold of p = 0.005.

For the fMRI analyses, the correct guess responses were contrasted to the incorrect guess responses based on the hypothesis that a reactivation of the putative memory trace underlying a forgotten memory (therefore a guess response) would more often occur in trials that yield a correct than incorrect response. Because sure responses were almost always correct, remembered trials were also contrasted with incorrect guess responses. For brain-behavior correlations, we used the z-scored percentage correct of all guess responses (guessing accuracy) for contrasts of guess responses and the z-scored absolute number of correct sure responses (sure accuracy) for contrasts of sure responses.

Common hippocampal areas mediated correct sure and correct guess responses at the 30-minute and 24-hour category retrieval

We asked whether forgotten memories, i.e., guess responses, were accompanied by activity increases within the episodic memory network like consciously accessible memories, i.e., sure responses.[50–54] This was indeed the case, as revealed by the following analyses. We isolated activity increases underlying associative retrieval from activity increases underlying mere face recognition and retrieval effort by computing the contrast correct sure responses versus incorrect guess responses and the contrast correct guess responses versus incorrect guess responses. These contrasts were then statistically conjugated to find common brain activation underlying correct sure and correct guess responses.

For the 30-minute category retrieval, the conjunction analysis using the data from the small FOV fMRI sequence yielded results in the right hippocampus (right hippocampus: $p_{uncor} < 0.001$, $T(285) = 4.05$, peak at coordinates based on the Montreal Neurological Institute (MNI) [30,-32,-9], see Figure 3A and Table S2). Further significant conjunctive results were located in the left posterior cingulate gyrus. These results were not replicated in the whole-brain fMRI sequence that had a slightly worse spatial resolution: the same conjunction analysis with data from the whole-brain fMRI sequence yielded no significant result at p < 0.001 uncorrected and k = 10 voxel extent.

For the 24-hour category retrieval, the conjunction analysis using the data from the small FOV fMRI sequence yielded also a result in the right hippocampus, albeit slightly below our chosen threshold of p < 0.001 uncorrected (right hippocampus: $p_{uncor} = 0.0012$, $T(285) = 3.07$, peak at MNI [26,-18,-17]). For illustrative purposes, see Figure 3B at p < 0.005 uncorrected. Significant results (at p < 0.001) were

located in the left lingual gyrus and bilateral posterior cingulate gyrus (see Table S3). The same conjunction analysis computed with data from the whole-brain fMRI sequence yielded no significant result. The lower power of conjunctive activity in the right hippocampus head at 24 hours might be due to a distinct overnight consolidation for consciously accessible versus consciously inaccessible memories, see below.

Hence, the hippocampus and other relevant structures of the episodic memory retrieval network were involved in the reactivation of both consciously accessible and consciously inaccessible associative memories at the 30-minute and at the 24-hour category retrieval.

Hippocampal activity underlying correct guess responses correlated with the guessing accuracy on the category retrieval task at 30 minutes and at 24 hours

Given that correct guess responses and correct sure responses were accompanied by activations in the same right hippocampal region, we then examined whether right hippocampal activity underlying correct guess responses versus incorrect guess responses would correlate with guessing accuracy across participants. Using the small FOV fMRI sequence for this analysis, a cluster in the right hippocampus head correlated with the guessing accuracy and even survived family-wise error correction (peak at MNI [27,-8,-25], cluster-level $p_{FWE} = 0.019$, $T(18) = 6.31$, $R = 0.72$, see Figure 3C, upper panel and Table S4 for all correlation results). Besides the right hippocampus, significant correlations were located in multiple bilateral inferior and middle temporal regions putatively supporting the reactivation of semantic object-related information and the precuneus and the posterior cingulate gyrus related to visual imagery.

The same correlation computed for the 24-hour retrieval using the small FOV fMRI sequence yielded a significant cluster in the same area within the right hippocampus head (peak at MNI [28, -10,-28], $p_{uncor} > 0.001$, $T(18) = 4.70$, $R = 0.70$, see Figure 3C, lower panel) besides multiple further clusters located in the right and left hippocampus (see Table S5). This result suggests that the bilateral hippocampus is involved in the retrieval of inaccessible memories after 24 hours, probably due to over-night memory consolidation, as detailed in a chapter below. Besides bilateral hippocampal areas, significant correlations were located in several bilateral middle temporal regions supporting the reactivation of semantic object-related information.

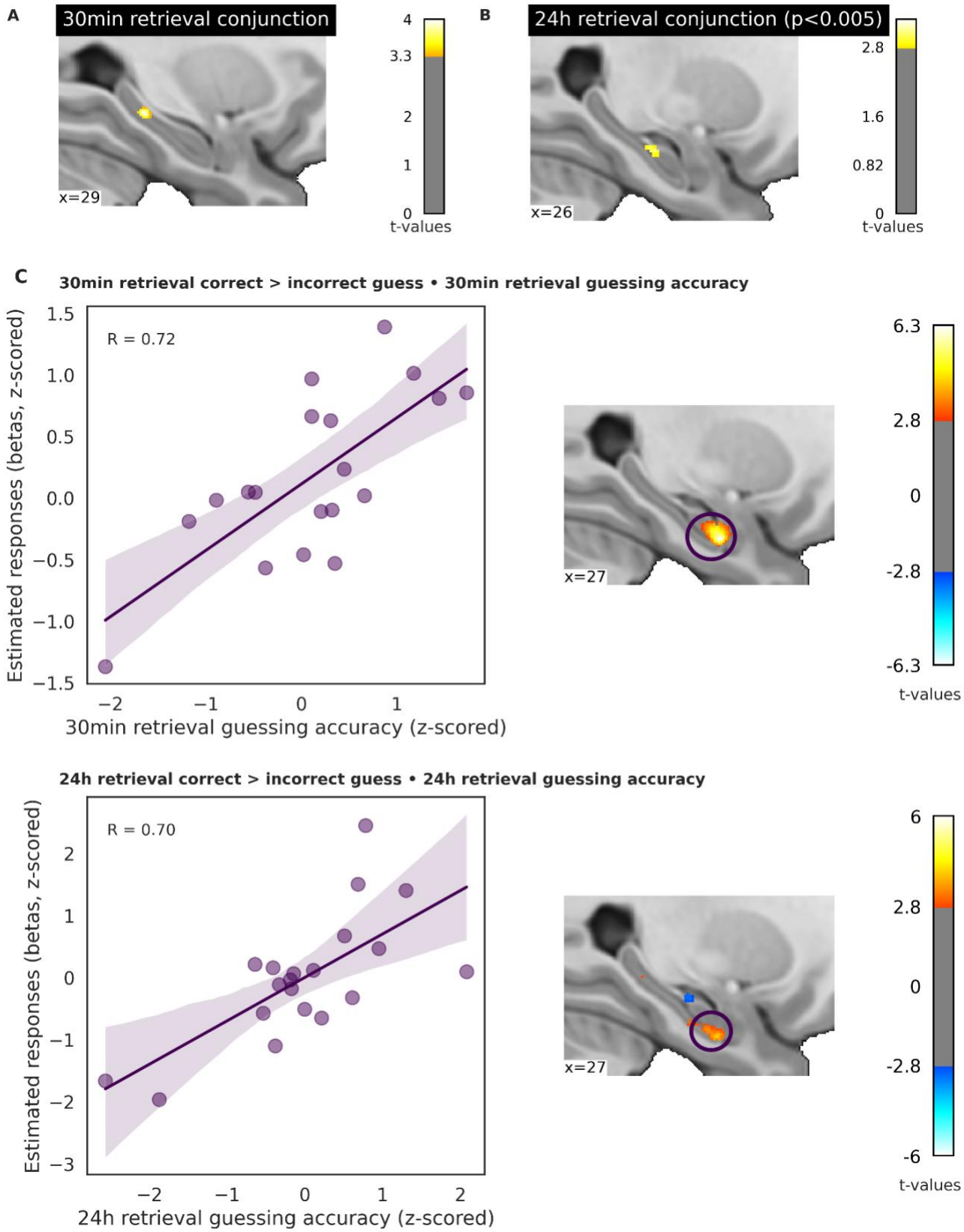

A
30min retrieval conjunction
x=29
4
3.3
2
1
0
t-values
B
24h retrieval conjunction (p<0.005)
x=26
2.8
1.6
0.82
0
t-values
C
30min retrieval correct > incorrect guess • 30min retrieval guessing accuracy
R = 0.72
Estimated responses (betas, z-scored)
30min retrieval guessing accuracy (z-scored)
x=27
6.3
2.8
0
-2.8
-6.3
t-values
24h retrieval correct > incorrect guess • 24h retrieval guessing accuracy
R = 0.70
Estimated responses (betas, z-scored)
24h retrieval guessing accuracy (z-scored)
x=27
6
2.8
0
-2.8
-6
t-values

Figure 3. Right hippocampal activity increased for both correct sure and correct guess responses at the 30-minute and the 24-hour category retrieval and correlated for guess responses with guessing accuracy at both testing time points.

A) Common right hippocampal activity increases underlying correct sure and correct guess responses at the 30-minute category retrieval. We isolated activity increases underlying associative retrieval from activity increases underlying mere face recognition, task characteristics, and retrieval effort by computing the contrast correct sure responses > incorrect guess responses and the contrast correct guess responses > incorrect guess responses. These contrasts were then statistically conjugated to find common brain activation underlying correct sure and correct guess responses. For a table of cluster sizes and t-values, see Table S2. Results presented in this panel stem from the small FOV fMRI sequence.

B) Common right hippocampal activity rises underlying correct sure and correct guess responses at the 24-hour category retrieval. iStatistical analysis analogous to 3A. For a table of cluster sizes and t-values, see Table S3. Results presented in this panel stem from the small FOV fMRI sequence.

C) Right hippocampal activity underlying correct guess responses on the category retrieval task correlated with the guessing accuracy on the category retrieval task at 30 minutes and at 24 hours. Upper panel: positive brain-behavior correlation in the right hippocampus. We show the contrast of correct guess > incorrect guess responses correlated with the guessing accuracy on the category retrieval task at 30 minutes (for cluster sizes and t-values, see Table S4). Lower panel: positive brain-behavior correlation in the right hippocampus. We show the contrast of correct guess > incorrect guess responses correlated with the guessing accuracy on the category retrieval task at 24-hours (for cluster sizes and t-values, see Table S5). Outliers (above absolute z-value of 3) were removed. Results presented in this panel stem from the small FOV fMRI sequence.

Hippocampal functional connectivity during learning and 30-minute retrieval correlated with the 30-minute guessing accuracy for forgotten associations

We next asked which neural circuits would interact with the right hippocampus when forming new associative memories that would later be forgotten. To this end, we computed a general psycho-physiological interaction analysis (gPPI) with the right hippocampus head as the seed region (based on the correlative results above) using the data of the whole-brain fMRI sequence. The comparison of learning trials that would yield correct versus incorrect guess responses at the 30-minute category retrieval (subsequent memory analysis) revealed that seed-based connectivity estimates correlated between-subjects with guessing accuracy. This was true for the right hippocampus head-to-bilateral anterior cingulate/medial superior frontal gyrus connectivity (peak at MNI [12,42,-4], peak-level $p_{FWE} < 0.05$, $T(18) = 7.8$, R = 0.92, see Figure 4A) and right hippocampus head-to-left hippocampus, right hippocampus head-to-bilateral lingual gyrus, and right hippocampus head-to-bilateral cuneus/precuneus connectivity (all results listed in Table S6). When computing this same correlation analysis with data collected during the 30-minute category retrieval, the right hippocampus-to-medial superior frontal gyrus connectivity again correlated with guessing accuracy (peak at MNI [12,46,28], cluster-level $p_{FWE} < 0.001$, $T(18) = 6.25$, R = 0.85, see Figure 4B). Connectivity from right hippocampus to bilateral lingual gyrus, right cuneus and bilateral precuneus no longer correlated at the 30-minute retrieval. However, right hippocampus-to-various prefrontal regions and right hippocampus-to-left inferior temporal gyrus connectivity correlated between-subjects with retrieval accuracy at the 30-minute retrieval (all results listed in Table S7). No significant correlations were found between seed-based connectivity and guessing accuracy at 24-hour retrieval. The reason for this lack of significant results at 24 hours might be the overnight consolidation of forgotten memories that provided for a deeper rooting of forgotten memories in the hippocampus rather than the neocortex (this will be detailed below).

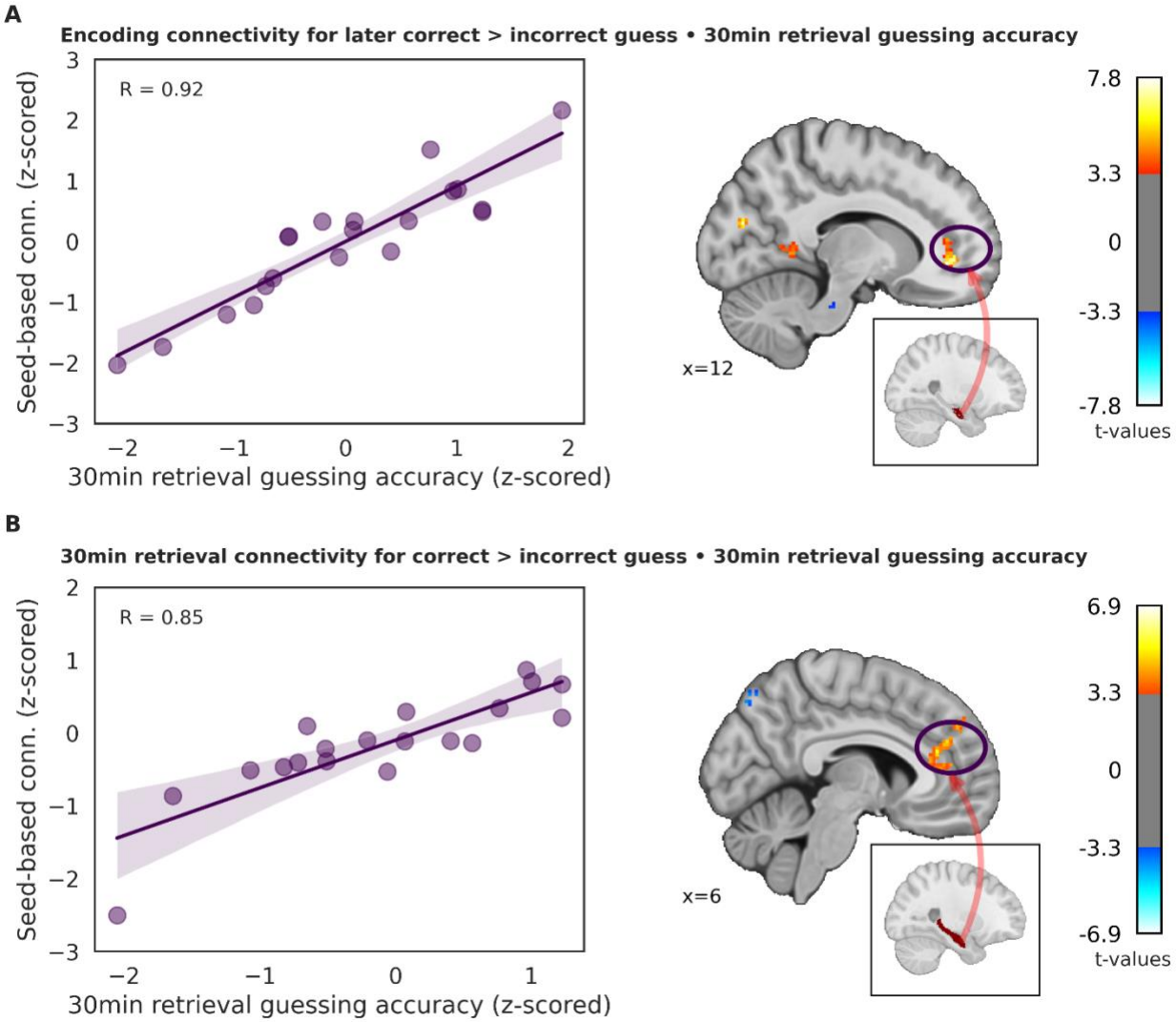


Figure 4: Hippocampal functional connectivity measured during learning and during the 30-minute category retrieval correlated with the 30-minute category retrieval guessing accuracy.
A) Hippocampal-medial prefrontal functional connectivity during learning correlated with the later recorded 30-minute category retrieval guessing accuracy. We computed the functional connectivity between the seed region in the right hippocampus head (see box inlet) with neocortical regions during learning using a gPPI. In participants with a higher guessing accuracy on the 30-minute category retrieval task, the functional connectivity was stronger between the right hippocampus head and the right anterior cingulate gyrus and the right medial prefrontal cortex. Outliers (above absolute z-value of 3) were removed. For a table of cluster sizes and t values see Table S6. Results presented in this panel stem from the whole-brain fMRI sequence.
B) Hippocampal-medial prefrontal functional connectivity during the 30-minute category retrieval correlates with the 30-minute category retrieval guessing accuracy. We computed the functional connectivity between the seed region in the right whole hippocampus (see box inlet) with neocortical regions during the 30-minute category retrieval using a gPPI. Participants with a higher guessing accuracy on the 30-minute category retrieval task, exhibited a stronger functional connectivity between the right hippocampus and the right medial prefrontal cortex. Outliers (above absolute z-value of 3) removed. For cluster sizes and t-values see Table S7. Results presented in this panel stem from the whole-brain fMRI sequence.

Overnight consolidation brought a neocorticalization for sure responses and deeper/broader hippocampal implementation for guess responses

Following the first visit to the MR Center, the learned face-object associations underwent overnight consolidation before participants reactivated the associations again during the category retrieval task on day 2. Following overnight consolidation, memory traces underlying sure versus guess responses became implemented more strongly in the neocortex as the standard model of memory consolidation predicts. This was the result of the stacked contrast (24-hour category retrieval: correct sure > correct guess) > (30-minute category retrieval: correct sure > correct guess responses) computed with the whole-brain fMRI data. It revealed strong overnight increases in activity in many bilateral fronto-parietal, temporal and occipital regions (all results in Table S8). The same computation with data from the small FOV fMRI sequence did not yield significant results.

In contrast, memory traces underlying correct guess responses on the category retrieval task were implemented more broadly and more strongly within bilateral hippocampus on the second versus the first day. This was the result of the comparison of correct guess responses given at the 24-hour category retrieval versus correct guess responses given at the 30-minute category retrieval using the data of the small FOV fMRI sequence. The result revealed overnight activity increases in bilateral hippocampus head (left hippocampus: peak at MNI [-27, -16, -14], $p_{uncor} < 0.001$, $T(18) = 4.62$, right hippocampus: peak at MNI [39,-22,-12], $p_{uncor} < 0.001$, $T(18) = 3.91$, see Table S9 for all results). The reverse contrast did not yield significant results in the hippocampus but results in the left-hemisphere: entorhinal area and middle temporal gyrus (see Table S10). Repeating this analysis using the data from the whole-brain fMRI sequence yielded merely one cluster in the inferior occipital gyrus for correct guess responses at 24 hours versus correct guess responses at 30 minutes (peak at MNI [-30 -90 -12], $p_{uncor} < 0.001$, $T(18) = 3.81$).

Next, we computed a gPPI analysis with the seed region in the right hippocampus head using the small FOV fMRI sequence to examine overnight changes in the functional connectivity within the medial temporal lobe underlying correct guess responses. Contrasting the functional connectivity underlying the correct guess responses given at the 24-hour category retrieval with the functional connectivity underlying correct guess responses given at the 30-minute category retrieval and correlating the results with the 24-hours guessing accuracy, we see an overnight increase in the behaviorally relevant functional connectivity of the right hippocampus head with left hippocampal regions (peak at MNI [-26,-30,-4], cluster-level $p_{FWE} = 0.011$, $T(18) = 5.37$, R = 0.61), with right entorhinal cortex, and with several regions within the right lateral temporal lobe (Figure 5; for all results, see Table S11). Hence, the overnight consolidation left the memory traces underlying guess responses implemented more broadly and deeply in bilateral hippocampus, which benefitted the accuracy of the guess responses given at 24 hours.

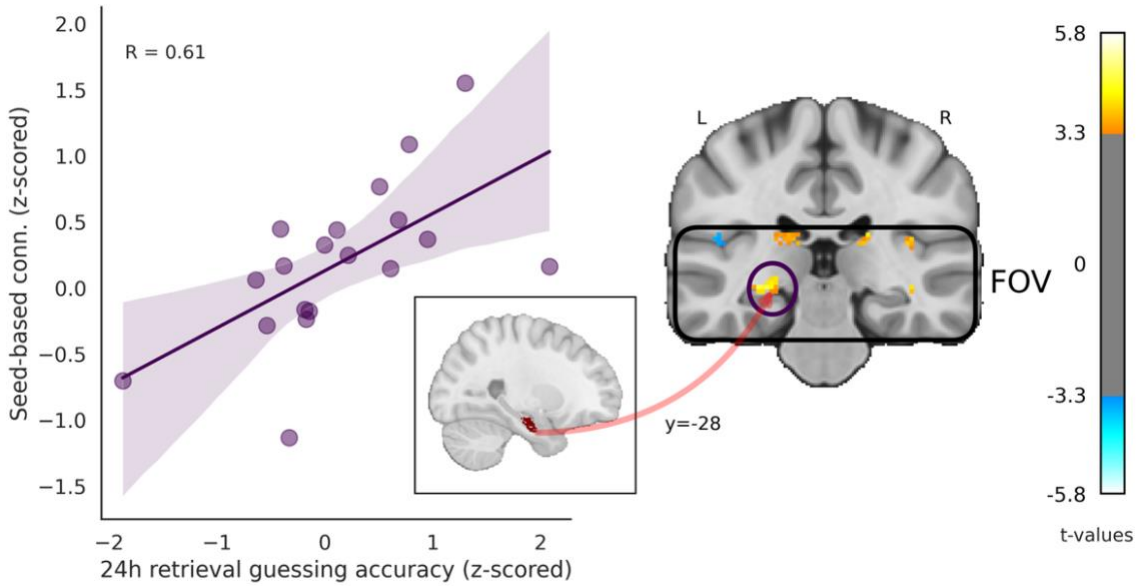


Figure 5. Overnight functional connectivity increases improved the guessing accuracy at the 24-hour category retrieval task. *We computed a gPPI analysis with the seed region in the right hippocampus head (see box inlet) using the small FOV fMRI sequence to examine overnight changes in the functional connectivity within the medial temporal lobe underlying correct guess responses on the category retrieval task. We contrasted the functional connectivity underlying the correct guess responses given on the 24-hour category retrieval task with the functional connectivity underlying correct guess responses given on the 30-minute category retrieval task and then correlated the result with the 24-hour guessing accuracy on the category retrieval task.*
Connectivity estimates and guessing accuracy were z-scored. Outliers (above absolute z-value of 3) removed. See Table S11 for all results.

Medial prefrontal cortex activation during guess responses given at the 24-hour category retrieval predicted the subsequent regain of conscious access to the memories

As mentioned in the behavioral results section, the mean guessing accuracy on the 24-hour category retrieval was at chance level but then rose significantly above chance level on the immediately following recognition task, where the face plus two objects (original object and foil object) were presented along with the face cue as additional retrieval cues. Hence, the addition of a valid object retrieval cue to the face retrieval cue (which was also present for category retrieval) improved the guessing accuracy and sometimes even allowed participants to regain conscious access to the previously inaccessible memories. Because the recognition task immediately followed the 24-hour category retrieval task, we were wondering which brain regions elevated activation levels during the 24-hour category retrieval task and whether these activation elevations would predict the subsequent regain of conscious access during the recognition task. To this end, we contrasted the guess responses on the 24-hour category retrieval task that would subsequently yield correct sure responses (i.e., conscious access) on the recognition task with those other guess responses given on the 24-hour category retrieval task that would yield correct guess responses (i.e., no conscious access) on the recognition task. We used the data from the whole-brain fMRI sequence. The medial prefrontal cortex was strongly activated for later correct sure versus correct guess responses given on the recognition task (peak at MNI [0 30 -16], $p_{uncor} < 0.001$, $T(18) = 4.19$, see Figure S1A and Table S12). This suggests that consciously inaccessible memories (guesses)

385 on the 24-hour category retrieval that possess (versus not) a medial prefrontal engram component, tend
386 to become consciously accessible again when given an additional retrieval cue.

387 **Additional object cues given on the recognition task permitted the reactivation of engram**
388 **components in the episodic memory network underlying sure responses and guess responses**

389 Because the right hippocampus head was involved in both guess and sure responses during the category
390 retrieval task, we asked which neural circuits the right hippocampus head would engage in for successful
391 face-object recognition. We computed a gPPI analysis with the seed region in the right hippocampus
392 head using the whole-brain fMRI sequence to examine the functional connectivity underlying guess and
393 sure responses given during recognition. Like in earlier analyses, we subtracted the functional
394 connectivity for incorrect guess responses from the functional connectivity for correct sure and for correct
395 guess responses given during recognition. The results were correlated with the sure and guess
396 recognition accuracy, respectively. Then, we computed a conjunction analysis on the two gPPI results to
397 determine commonly activated and behaviorally relevant neural circuits underlying recognition. The
398 functional connectivity of the right hippocampus head with bilateral lingual gyrus and bilateral calcarine
399 cortex was crucial for recognition success (peak at MNI [-4 -82 -2], $p_{uncor} < 0.001$, $T(18) = 3.53$, note that
400 this was the only cluster found for this conjunction analysis). Hence, the functional connectivity of the right
401 hippocampus with areas of the visual cortex was relevant for the explicit (remembered associations) and
402 implicit (forgotten associations) recognition of the previously learned face-object combinations.

403 Zooming in on the forgotten associations, we were wondering about the difference in the behaviorally
404 relevant functional connectivity of the right hippocampus head for those forgotten memories that still
405 yielded correct versus incorrect guess responses on the recognition task. We computed a gPPI analysis
406 with data from the whole-brain fMRI sequence for the seed region in the right hippocampus head
407 comparing correct versus incorrect guess responses on the recognition task and correlated the results
408 between-subjects with the individual accuracy of the guess responses. The accuracy of the guessed
409 recognition responses increased with a connectivity increase between the right hippocampus head and
410 the left hippocampus tail, between the right hippocampus head and bilateral lingual gyrus, between the
411 right hippocampus head and the left middle temporal gyrus as well as between the right hippocampus
412 head and a large bilateral fronto-parietal network (all results in Table S13).

413 Zooming in on the remembered associations, we computed the same gPPI analysis for correct sure
414 versus incorrect guess responses given on the recognition task and correlated the result between-
415 subjects with the individual accuracy of the sure responses. The result was a similar connectivity of the
416 right hippocampus head with bilateral lingual gyrus, with the right middle temporal gyrus as well as a large
417 bilateral fronto-parietal network, but no connectivity with the left hippocampus tail (see Table S14 for all
418 results). Hence, response accuracy during guessing was mediated by bilateral hippocampal connectivity
419 and for both guess and sure responses was mediated by the connectivity of the right hippocampus head
420 with visual cortices and a large bilateral fronto-parietal network. It thus appears that the addition of the

421 relevant object cue during recognition allowed for a better reactivation of engram components in the
422 episodic memory network underlying both sure and guess responses. Note that the right hippocampus
423 head-to-left hippocampus tail connectivity result for guess responses was not replicated when using the
424 small FOV fMRI sequence. The cause might be participant differences (between-subjects design) in
425 implicit retrieval or poor signal in this very posterior left hippocampal region.

426 **fMRI results from representational similarity analyses**

427 We asked whether more fine-grained engram components, namely individual voxel patterns, would
428 reactivate during distinct memory stages and reflect the persistence or shift regarding certain engram
429 components over time. We expected the repetition of voxel patterns underlying memories in those brain
430 areas that yielded significant results in the above reported univariate analyses. Therefore, we chose a
431 region-of-interest approach deriving the regions-of-interest from the above reported univariate results and
432 from the priori hypothesis that the hippocampal engram components would persist when a memory
433 becomes consciously inaccessible. Of particular interest regarding the trajectory of episodic memories
434 were recurring or shifting voxel patterns in the hippocampus. We computed representational similarity
435 analyses (RSA), i.e., pair-wise comparisons of voxel patterns, comparing encoding – retrieval similarity
436 (ERS) for the 30-minute and 24-hour category retrieval, encoding – recognition (at 24 hours) similarity,
437 and retrieval (30 minutes) – retrieval (24 hours) similarity (RRS) for the category retrieval. We separated
438 trials into guess and sure responses to compare the trajectory of engram components for consciously
439 accessible and consciously inaccessible memories. By subtracting RSA values underlying incorrect guess
440 responses from RSA values underlying correct guess and correct sure responses, we isolated estimates
441 due to associative retrieval success by eliminating estimates that originate from the mere repetition of
442 face cues (yielding face recognition) and by eliminating estimates that originate from shared task
443 requirements (STAR Methods). Thus, the below reported similarity estimates underly successful
444 associative encoding and associative retrieval operations. The height threshold in all RSA analyses was
445 set to $p < 0.05$, uncorrected for multiple comparisons. We report a Bayes factor for each t-test and
446 Pearson correlation.

447 **Similarity of voxel patterns between encoding and the 30-minute category retrieval were relevant**
448 **for retrieval accuracy**

449 Using data from the small FOV fMRI sequence, voxel patterns tended to be similar in the left hippocampal
450 body between encoding and the 30-minute category retrieval for correct sure responses with the degree
451 of this similarity showing a trend toward a positive relationship with the number of correct sure responses
452 across subjects ($p = 0.086$, $R = 0.39$, $BF_{10} = 1.1$) The same computation yielded neither a tendency nor a
453 significant result in the medial temporal lobe for correct guess responses.

454 Within neocortex (using data from the whole-brain fMRI sequence), voxel patterns were significantly
455 similar in the bilateral inferior temporal gyrus for correct sure responses ($t(16) = 2.3$, $p = 0.035$, $BF_{10} = 1.9$,
456 see Figure S2), with no significant between-subjects correlation of the similarity values with the number of

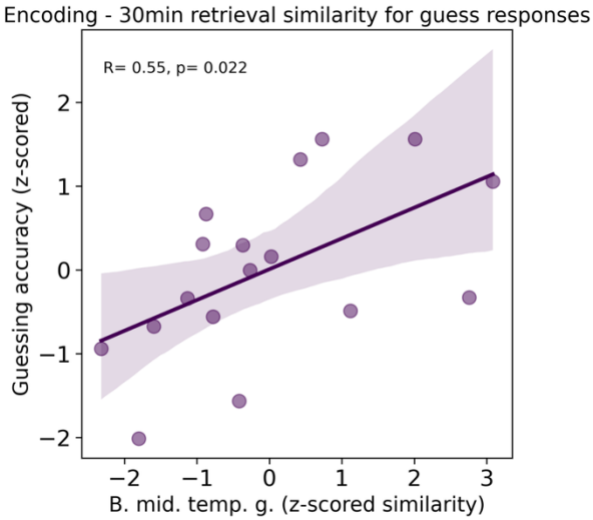
correct sure responses. For correct guess responses, this computation revealed non-significant mean value comparisons but a significant between-subjects correlation between the degree of similarity of voxel patterns in the bilateral middle temporal gyrus and guessing accuracy ($p = 0.022$, $R = 0.55$, $BF_{10} = 3.3$, see Figure 6A). It thus appears that the reactivation of engram components in the inferior and middle temporal gyrus contributed to the conscious and unconscious reactivation of memories at 30 minutes following encoding.

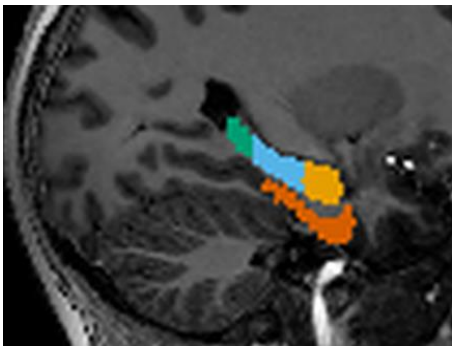
Similarity of voxel patterns between encoding and the 24-hour category retrieval were relevant for retrieval success

Following the first visit at the MR Center, the encoded face-object associations underwent overnight consolidation before participants reactivated the associations again during their second visit at the MR Center on day 2. Overnight memory consolidation may have stabilized the activated voxel patterns from their inception to their 24-hour reactivation during the category retrieval. For both correct guess ($p = 0.029$, $t(19) = 2.35$, $BF_{10} = 2.1$) and correct sure ($p = 0.015$, $t(19) = 2.65$, $BF_{10} = 3.5$) responses, voxel patterns were significantly similar between encoding and the 24-hour category retrieval within the left parahippocampal gyrus (see Figure 6B). Moreover, the degree of this left parahippocampal pattern similarity for sure responses tended to correlate between-subjects with the number of correct sure responses on the 24-hour category retrieval task ($p = 0.058$, $R = 0.43$, $BF_{10} = 1.5$); the same was true for the left hippocampal body ($p = 0.059$, $R = 0.44$, $BF_{10} = 1.5$). Note that the pattern similarity in the left hippocampal body had already correlated between subjects with the number of correct sure responses for the encoding – 30-minute comparison. Hence, medial temporal pattern reinstatements were behaviorally relevant for the conscious retrieval at both time-points. For correct guess responses, however, correlations between medial temporal encoding-retrieval pattern reinstatements and retrieval success did not reach significance. A potential reason is the initially weak dependence on the medial temporal lobe of later forgotten memories. The time after encoding, i.e. consolidation, appeared to have strengthened the medial temporal engram components of forgotten memories. Results from this paragraph are based on data from the small FOV fMRI sequence.

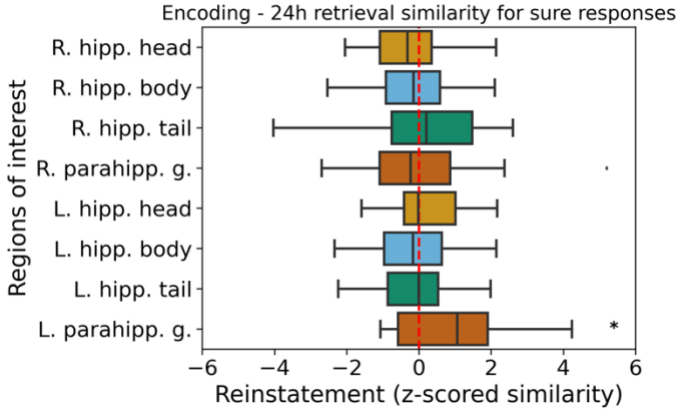
Within neocortex, the similarity of voxel patterns between encoding and the 24-hour category retrieval underlying correct sure responses revealed non-significant mean value comparisons but a significant between-subjects correlation of the similarity values with the number of correct sure responses within the bilateral middle temporal gyrus ($p = 0.006$, $R = 0.64$, $BF_{10} = 10.4$, see Figure S3) as previously seen for the correctly guessed trials for the encoding – 30-minute retrieval comparison. The similarity of voxel patterns between encoding and 24-hour retrieval also revealed non-significant mean value comparisons for correct guess responses but a significant between-subjects correlation with retrieval success in bilateral cuneus ($p = 0.049$, $R = 0.48$, $BF_{10} = 1.8$) plus non-significant correlations within bilateral inferior temporal sulcus ($p = 0.099$, $R = -0.41$, $BF_{10} = 1.0$) and bilateral anterior cingulate gyrus ($p = 0.100$, $R = -0.41$, $BF_{10} = 1.0$).

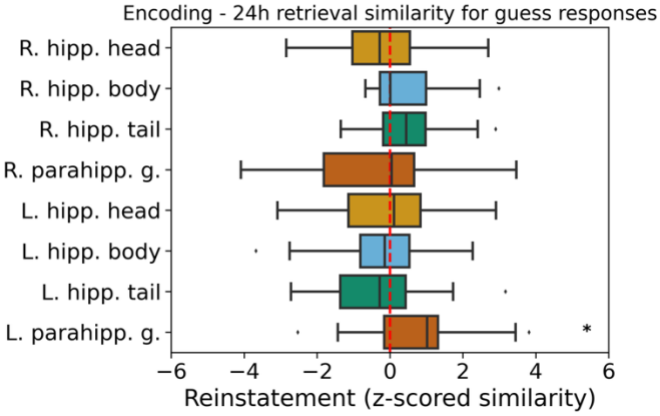
A


Encoding - 30min retrieval similarity for guess responses
R= 0.55, p= 0.022
Guessing accuracy (z-scored)
B. mid. temp. g. (z-scored similarity)

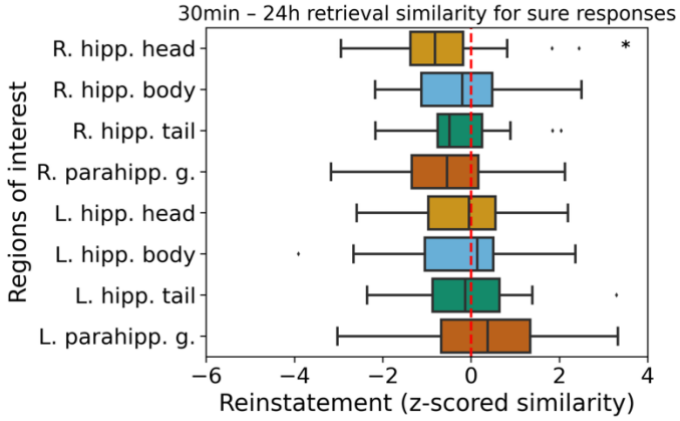


B


Encoding - 24h retrieval similarity for sure responses
R. hipp. head
R. hipp. body
R. hipp. tail
R. parahipp. g.
L. hipp. head
L. hipp. body
L. hipp. tail
L. parahipp. g.
Regions of interest
Reinstatement (z-scored similarity)


Encoding - 24h retrieval similarity for guess responses
R. hipp. head
R. hipp. body
R. hipp. tail
R. parahipp. g.
L. hipp. head
L. hipp. body
L. hipp. tail
L. parahipp. g.
Reinstatement (z-scored similarity)

C


30min – 24h retrieval similarity for sure responses
R. hipp. head
R. hipp. body
R. hipp. tail
R. parahipp. g.
L. hipp. head
L. hipp. body
L. hipp. tail
L. parahipp. g.
Regions of interest
Reinstatement (z-scored similarity)

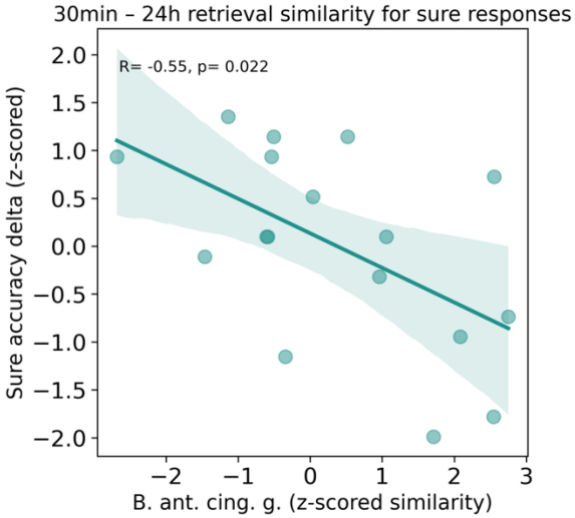

30min – 24h retrieval similarity for sure responses
R= -0.55, p= 0.022
Sure accuracy delta (z-scored)
B. ant. cing. g. (z-scored similarity)

Figure 6. Encoding – retrieval and retrieval – retrieval similarity.
A) *Left panel: Similarity of voxel patterns between encoding and the 30-minute category retrieval were relevant for retrieval success. Significant between-subjects correlation of encoding-retrieval similarity values with guessing accuracy on the 30-minute category retrieval in bilateral middle temporal gyrus. Results presented in this panel were acquired with the whole-brain fMRI sequence. Right panel: small FOV ROI overview for B) and C)*
B) Similarity of voxel patterns between encoding and the 24-hour category retrieval were relevant for retrieval success. Encoding-retrieval similarity values were significant in left parahippocampal gyrus for both correct guess responses (left) and correct sure responses (right) ($t(19) = 2.65$, $p = 0.015$, $BF_{10} = 3.5$; $t(19) = 2.35$, $p = 0.029$, $BF_{10} = 2.1$, respectively). Results presented in this panel were acquired with the small FOV fMRI sequence.
*C) Dissimilarity of voxel patterns between the 30-minute and the 24-hour category retrieval were relevant for retrieval success. Left panel: 30-minute – 24-hours category retrieval dissimilarity in the right hippocampus head for correct sure responses. Results presented in this panel were acquired with the small FOV fMRI sequence. Right panel: 30-minute – 24-hours retrieval dissimilarity values correlated between-subjects with the delta of correct sure responses (24-hours sure accuracy minus 30-minute sure accuracy) in bilateral anterior cingulate gyrus (B. ant. cing. g.). Results presented in this panel were acquired with the whole-brain fMRI sequence. *$p < 0.05$, **$p < 0.01$ by Student's t test. Abbreviations: ROI, region of interest; R., right; L., left; hipp., hippocampus; parahipp., parahippocampus; g., gyrus.*

(Dis)similarity of voxel patterns between the 30-minute and the 24-hour category retrieval were relevant for retrieval success

The retrieval-retrieval comparison of voxel patterns, i.e., the comparison between the 30-minute and the 24-hour category retrieval, for correct sure responses yielded a significant dissimilarity for the right hippocampus head ($p = 0.035$, $t(19) = -2.26$, $BF_{10} = 1.8$, see Figure 6C, left panel). The same analysis for correct guess responses revealed also a significant dissimilarity in the right hippocampus, however in its tail region ($p = 0.046$, $t(19) = -2.12$, $BF_{10} = 1.5$, not displayed). On the other hand, there was a retrieval-retrieval similarity (not dissimilarity) of voxel patterns in the right hippocampus head underlying correct guess responses that correlated significantly ($p = 0.018$, $R = 0.52$, $BF_{10} = 3.7$, see Figure S4) between-subjects with the delta of correct guess responses (i.e., number of correct guess responses given at 24 hours minus the number of correct guess responses at 30 minutes). Hence, participants exhibiting a similar right hippocampal retrieval-retrieval voxel pattern yielded a better retrieval accuracy at 24 hours versus 30 minutes. Although a right hippocampal pattern dissimilarity hallmarked correct retrieval responses at 24 hours versus 30 minutes for both sure and guess responses, a pattern similarity within the right hippocampus head still aided a few participants, who lacked conscious access to memories at both retrieval time-points. Results from this paragraph are based on data from the small FOV fMRI sequence.

Within neocortex, the (dis)similarity of voxel patterns between the 30-minute and the 24-hour category retrieval underlying correct sure responses revealed non-significant mean value comparisons but a significant between-subjects correlation of pattern dissimilarity with the delta of correct sure responses within bilateral anterior cingulate gyrus ($p = 0.021$, $R = -0.55$, $BF_{10} = 3.4$, see Figure 6C, right panel). Thus, the overnight memory consolidation may have also changed the anterior cingulate engram components – and not just the right hippocampal engram components – such that a pattern dissimilarity supported a successful retrieval on the next day. For correct guess responses, the (dis)similarity of voxel pattern between the 30-minute and the 24-hour category retrieval revealed also non-significant mean

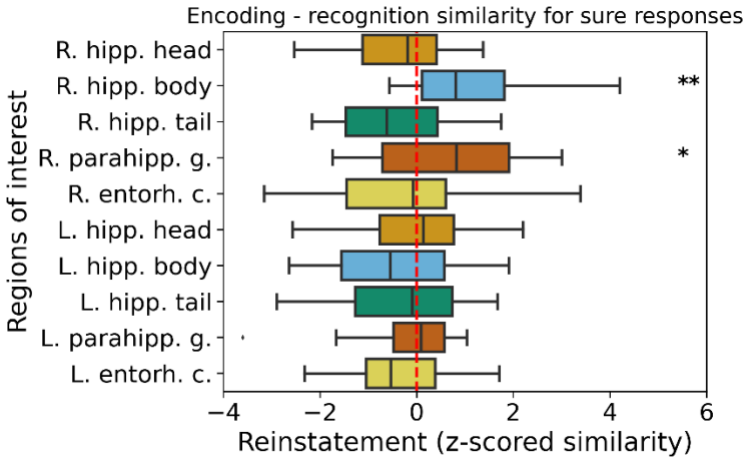
538 value comparisons, yet a significantly positive between-subjects correlation of pattern similarity with the
539 delta of correct guess responses within bilateral cuneus ($p = 0.005$, $R = 0.64$, $BF_{10} = 10.8$, not displayed).
540 Hence, participants with a similar cuneal retrieval-retrieval voxel pattern yielded a better retrieval accuracy
541 at 24 hours versus 30 minutes. Therefore, also the cuneal pattern similarity appears to have aided a few
542 participants, who lacked conscious access to memories at both retrieval time-points. Overall, these
543 results show that, especially for dynamic engram components such as the hippocampus and anterior
544 cingulate gyrus, behaviorally relevant dissimilarity can occur.

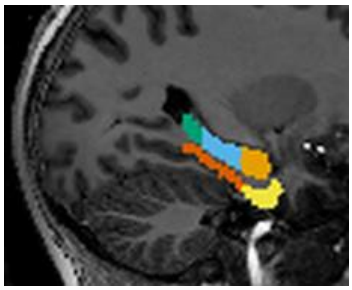
545 **Encoding – recognition pattern similarity**

546 Unlike the retrieval-retrieval dissimilarity results presented above, there was not a single pattern
547 dissimilarity result for the encoding-recognition comparison. For correct sure responses on the recognition
548 task, voxel patterns were significantly similar between encoding and recognition within the right
549 hippocampal body ($p = 0.001$, $t(19) = 3.9$, $BF_{10} = 34.0$) and the right parahippocampal gyrus ($p = 0.048$,
550 $t(19) = 2.12$, $BF_{10} = 1.5$, Figure 7A), with no significant between-subjects correlations of pattern similarity
551 with recognition accuracy. For correct guess responses on the recognition task, neither mean value
552 comparisons nor correlations between pattern reinstatements and recognition accuracy reached
553 significance. Results from this paragraph are based on data from the small FOV fMRI sequence.

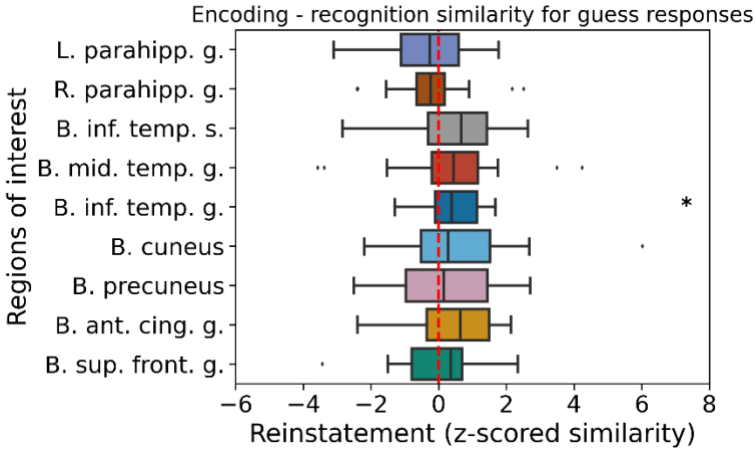
554 Within neocortex, the similarity of voxel patterns between encoding and recognition underlying correct
555 sure responses on the recognition task revealed a significant mean value comparison in the bilateral
556 inferior temporal sulcus ($p = 0.020$, $t(17) = 2.55$, $BF_{10} = 2.9$, see Figure S5). The degree of this inferior
557 temporal pattern similarity did not correlate with recognition success. Moreover, a pattern similarity in
558 bilateral anterior cingulate gyrus correlated significantly between-subjects with recognition accuracy ($p =$
559 0.042, $R = 0.48$, $BF_{10} = 2.0$, see Figure S6) but was insignificant in terms of mean value comparison.
560 Also, the similarity of voxel patterns between encoding and recognition for correct guess responses
561 revealed a significant mean value comparison in the bilateral inferior temporal gyrus ($p = 0.038$, $t(17) =$
562 2.25, $BF_{10} = 1.8$, see Figure 7B). The degree of this inferior temporal pattern similarity did not correlate
563 with accuracy of guess responses given on the recognition task. Overall, these results show that correct
564 sure responses were accompanied by encoding – recognition voxel pattern similarity in the medial
565 temporal lobe, consistent with an episodic reinstatement, while for guess trials, the encoding-recognition
566 reinstatement of neocortical voxel patterns seems to be relevant for retrieval. A potential reason for the
567 lack of encoding-recognition medial temporal similarity results for correct guess responses is the initially
568 weak dependence of later forgotten memories on the medial temporal lobe. The time after encoding, i.e.
569 consolidation, appeared to have strengthened the medial temporal engram components of forgotten
570 memories.

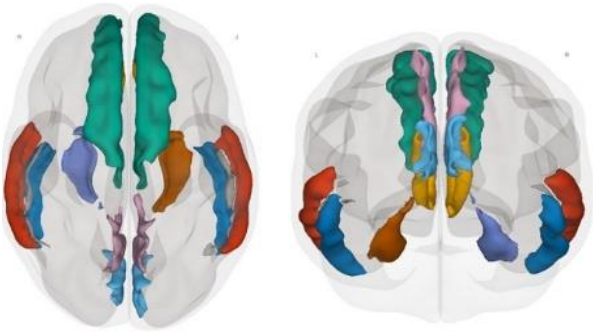
A


Encoding - recognition similarity for sure responses
R. hipp. head
R. hipp. body
R. hipp. tail
R. parahipp. g.
R. entorh. c.
L. hipp. head
L. hipp. body
L. hipp. tail
L. parahipp. g.
L. entorh. c.
Regions of interest
**
*
−4
−2
0
2
4
6
Reinstatement (z-scored similarity)



B


Encoding - recognition similarity for guess responses
L. parahipp. g.
R. parahipp. g.
B. inf. temp. s.
B. mid. temp. g.
B. inf. temp. g.
B. cuneus
B. precuneus
B. ant. cing. g.
B. sup. front. g.
Regions of interest
*
−6
−4
−2
0
2
4
6
8
Reinstatement (z-scored similarity)



573 ***Figure 7. Encoding – recognition pattern similarity.***
574 *A) Encoding-recognition pattern similarity in the right hippocampal body and right parahippocampal gyrus for correct*
575 *sure responses on the recognition task ($t(19) = 3.91$, $p = 0.001$, $BF_{10} = 34.0$; $t(19) = 2.11$, $p = 0.048$, $BF_{10} = 1.5$,*
576 *respectively). Results presented in this panel were acquired with the small FOV fMRI sequence.*
577 *B) Encoding-recognition pattern similarity in bilateral inferior temporal gyrus for correct guess responses on the*
578 *recognition task ($p = 0.038$, $t(17) = 2.25$, $BF_{10} = 1.8$). Results presented in this panel were acquired with the whole-*
579 *brain fMRI sequence.*
580 *Right panels: ROI overview, *p < 0.05, **p < 0.01 by Student's t test. Abbreviations: R., right; L., left; B., bilateral;*
581 *parahipp., parahippocampus; hipp., hippocampus; entorh., entorhinal; inf., inferior; mid., middle; ant., anterior; cing.,*
582 *cingulate; sup., superior; front., frontal; s., sulcus, g., gyrus c., cortex.*
583

584 ## DISCUSSION

585 Inspired by recent engram research in mice suggesting that the engrams of forgotten memories remain
586 implemented in the brain, we asked whether engrams underlying human episodic memories would also
587 persist and whether they influence human behavior implicitly? We mapped the fate of 96 new engrams at
588 the systems level using high-resolution fMRI at 7 Tesla during learning, an immediate, a 30-minute, and a
589 24-hour retrieval test. Participants indicated on each retrieval trial, whether they remembered or forgot the
590 memory. Almost all 96 associative memories could be formed during the one-trial learning event, and
591 they were consciously retrieved at the immediately given category retrieval test. Hence, encoding was
592 successful. Univariate and multivariate analyses of the fMRI data revealed that the engrams of
593 subsequently forgotten memories remain implemented in the episodic memory network and continue to
594 influence the accuracy of guess responses at later stages of category retrieval. The engrams of forgotten
595 memories became deeper implemented within bilateral hippocampus overnight, while consciously
596 accessible memories were neocorticalized overnight. For consciously accessible and inaccessible
597 (forgotten) memories, engram components within right hippocampus and anterior cingulate cortex shifted
598 from the 30-minute to the 24-hour category retrieval. Hence, subjectively forgotten episodic memories
599 appear to remain implemented in the episodic memory network and continue to influence human behavior
600 implicitly.

601 **Traces of consciously inaccessible memories remained in the hippocampus and continued to**
602 **guide retrieval choices**

603 Although the 30-minute and the 24-hour guessing accuracies on the category retrieval task were not
604 above chance level on average, brain activity in the right hippocampus head – an essential brain region
605 for episodic memory[33,34] – underlying correct versus incorrect guess responses correlated with guessing
606 accuracy across participants at both time-points. This strongly suggests that the episodic memory traces
607 persisted in the brain despite loss of conscious access to the memories and that the traces continued to
608 influence the participants' choice behavior at test. Memory traces underlying consciously accessible (sure
609 responses) and consciously inaccessible (guess responses) memories overlapped in the episodic
610 memory retrieval network. The engram-related activity was generally a bit weaker and more
611 circumscribed for forgotten versus remembered memories, although these differences were not significant

statistically. The rather weak activity underlying forgotten memories nevertheless influenced the participants' choice behavior at retrieval testing.[42,43,55]

Distinct consolidation trajectories for consciously accessible and inaccessible memories

Interestingly, consciously accessible and consciously inaccessible memories underwent distinct consolidation trajectories overnight. Consciously accessible memories relied on neocortical structures immediately following encoding and became even more broadly neocorticalized following overnight consolidation, which is typical for systems consolidation of episodic memories. Goto and colleagues[18] had previously demonstrated that a first wave of long-term potentiation acts locally in the hippocampus during the minutes following learning, a second wave of hippocampal long-term potentiation-dependent consolidation occurs during the first sleep phase following learning; and a third wave of long-term potentiation occurs in the anterior cingulate gyrus on the second day following learning and supports the strengthening of neocortical engram components. It appears that the strong, consciously accessible memories soon underwent the third wave of memory consolidation because they became quickly represented in anterior cingulate gyrus and medial prefrontal cortex. At the behavioral level, the consciously accessible memories remained relatively stable overnight: virtually all consciously accessible memories at the 24-hour category retrieval were already consciously accessible at the 30-minute category retrieval.

The consciously inaccessible (forgotten) memories recruited fewer neocortical areas immediately following encoding and following overnight consolidation but they became more broadly implemented in bilateral hippocampus: the left and right hippocampus was more strongly activated during guess responses given at the 24-hour relative to the 30-minute category retrieval. Moreover, the inter-hippocampal connectivity became stronger overnight and correlated with the 24-hour category guessing accuracy. Therefore, overnight memory consolidation may have compensated for a weaker representational status of forgotten memories with the purpose to rescue these memories from degradation by implementing them deeper in the medial temporal lobe and only later into the neocortex. This compensation appears to have occurred during the first two waves of memory consolidation according to Goto[18] and colleagues, which act within the hippocampus. Hence, the consolidation of forgotten memories proceeded slower than described by Goto et al. and was much slower than the consolidation of the consciously accessible memories in this study. Interestingly though, the stronger memories among the forgotten memories also activated the anterior cingulate gyrus at the 24-hour category retrieval, like the consciously accessible memories did, and this subgroup of forgotten memories regained conscious access during the recognition task. The recognition task was easier than the category retrieval task because both the face and the associated object were presented as retrieval cues.

645 **Transformation of engram representations revealed by pattern dissimilarity**

646 A common notion in neuroscience is that neurons that encode information also retain the information and
647 are reactivated in the retrieval process.[2,11,21,22,56,57] Accordingly, we expected to observe retrieval-related
648 activation within the same voxels that had been activated during encoding. Indeed, voxel patterns
649 repeated from encoding to 30-minute category retrieval within medial temporal and neocortical areas. Yet,
650 the voxel patterns in the hippocampus and anterior cingulate gyrus became dissimilar from the 30-minute
651 to the 24-hour category retrieval for both consciously accessible and inaccessible memories. Hence the
652 retrieval and ensuing re-encoding process at the 30-minute retrieval may have shifted activated voxels
653 within these two key areas of episodic memory, as was observed in the animal model: Neurophysiological
654 studies in animals have shown that hippocampal memory traces can change significantly across time as
655 well as depending on the retrieval demands.[26–29,58,59] Lei and colleagues[29] recently reported that
656 hippocampal engrams underlying contextual fear memories in mice shift with repeated retrieval. Memory
657 recall can trigger a change in the hippocampal engram leading to the formation of a new, updated
658 hippocampal engram. Ko and colleagues[60] also found that the time-dependent transformation of
659 contextual fear memories involves a reorganization of hippocampal engram circuits. For our study, we
660 assume that re-encoding and reconsolidation processes have caused the shift in voxel patterns from the
661 30-minute to the 24-hour retrieval. There was no systematic reinstatement or shift of hippocampal engram
662 components from encoding to the 30-minute category retrieval or the 24-hour category retrieval.
663 However, from encoding to the recognition test, where the same face and object displayed during
664 encoding was re-presented, a right hippocampal engram component reinstated for sure responses
665 possibly due to visual recognition.

666 A pattern shift from the 30-minute to the 24-hour category retrieval was not just seen in the right
667 hippocampus but also within bilateral anterior cingulate gyrus. The anterior cingulate gyrus is strongly
668 connected to the hippocampus, heavily involved in early consolidation processes[18], shares cell dynamics
669 with the hippocampus[61] and even harbors place cells.[62] The anterior cingulate gyrus is also involved in
670 storing novel information into already established knowledge.[63] This may explain why a voxel dissimilarity
671 from the 30-minute to the 24-hour category retrieval occurred not only in the hippocampus but also in the
672 anterior cingulate gyrus.

673 **Theoretical implications**

674 This research was inspired by experiments in rodents, which suggest that interference-based forgetting is
675 an adaptive form of forgetting, which is reversible and updateable and thus in favor of an inaccessible
676 rather than unavailable engram.[8] For example, Autore and colleagues[8] showed that although natural
677 retroactive interference during learning resulted in forgetting and decreased engram cell reactivation in
678 the retrieval situation, optogenetic stimulation of previously labeled engram cells enabled memory
679 expression. Hence, engram components remained implemented in the brain despite forgetting. In the
680 consciously inaccessible state, as we have shown here, episodic memories can still be reactivated, they

can still guide behavior, and they are still subject to active consolidation processes. Hence, they are far from gone. This informs traditional models of memory that distinguish between memory systems based on whether learning and retrieval occur consciously or unconsciously.[31–33] These models posit that episodic memories are strictly dependent on consciousness, i.e., one needs to have a conscious experience to relive an episode from one's life. An alternative model suggests that a separation between memory systems based on processing modes is more plausible and is better supported by memory data.[34] According to this newer model, consciousness modulates memories within each memory system in a quantitative way (i.e., memory strength) but does not qualitatively define memory systems. The current findings provide evidence in favor of this processing-based memory model[34] because initially consciously accessible episodic memories can lose conscious accessibility and may regain it when given sufficient retrieval cues. Even those memories that never re-emerge to conscious access remain implemented in the episodic memory network including the hippocampus. These unconscious memories are also subject to active consolidation processes and continue to influence the choice behavior. A future line of research might examine the effects of magnetic or electrical brain stimulation on engrams underlying consciously inaccessible memories. Can brain stimulation reinforce engram components underlying consciously inaccessible memories to bring them back to conscious access?

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Tom Willems (tom.willems@unibe.ch).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The magnetic resonance imaging data have been deposited at "OpenNeuro" and are publicly available.

- Small FOV fMRI group: https://doi.org/10.18112/openneuro.ds006265.v1.0.0
- Whole-brain fMRI group: https://doi.org/10.18112/openneuro.ds006266.v1.0.0

Custom MATLAB, R, and python analysis scripts as well as the behavioral data are available at https://github.com/TomW92/TOAM-fMRI.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by Sitem-Insel Support Funds SISF 2019 to K. Henke and by the SNFS Advanced Grant TMAG-1_209374 to K. Henke. The authors thank Mirco Bristle and Luzius Brogli for their

support during data analysis. Calculations were performed on UBELIX (https://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern.

AUTHOR CONTRIBUTIONS

Conceptualization: T.W., K.H.; Methodology: T.W., A.F., K.H.; Analysis: T.W., K.Z., F.R., A.F.; Writing - original draft: T.W.; Writing - review & editing: T.W., K.Z., K.H.; Visualization: T.W.; Supervision: K.H.; Project administration: K.H.; Funding acquisition: K.H.

DECLARATION OF INTERESTS

The authors declare no competing interest.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES

During the preparation of this work, the author(s) used AI-assisted technologies for the reformatting of tables, to create pictograms for Figure 1, and to generate abbreviations from lists of brain regions and for checking grammar. Thereafter, the author(s) reviewed and edited the content and take(s) full responsibility for the content of the publication.

SUPPLEMENTAL INFORMATION

Supplemental information can be found in the pdf accompanying this manuscript.

REFERENCES

1. Josselyn, S.A., Köhler, S., and Frankland, P.W. (2015). Finding the engram. Nat. Rev. Neurosci. *16*, 521–534. https://doi.org/10.1038/nrn4000.

2. Josselyn, S.A., and Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. Science *367*, eaaw4325. https://doi.org/10.1126/science.aaw4325.

3. Davis, R.L., and Zhong, Y. (2017). The Biology of Forgetting—A Perspective. Neuron *95*, 490–503. https://doi.org/10.1016/j.neuron.2017.05.039.

4. De Snoo, M.L., and Frankland, P.W. (2025). Neurobiological mechanisms of forgetting across timescales. Curr. Opin. Neurobiol. *90*, 102972. https://doi.org/10.1016/j.conb.2025.102972.

5. Guskjolen, A., Kenney, J.W., de la Parra, J., Yeung, B. ru A., Josselyn, S.A., and Frankland, P.W. (2018). Recovery of “Lost” Infant Memories in Mice. Curr. Biol. *28*, 2283-2290.e3. https://doi.org/10.1016/j.cub.2018.05.059.

6. Yates, T.S., Fel, J., Choi, D., Trach, J.E., Behm, L., Ellis, C.T., and Turk-Browne, N.B. (2025). Hippocampal encoding of memories in human infants. Science *387*, 1316–1320. https://doi.org/10.1126/science.adt7570.

7. Bolsius, Y.G., Heckman, P.R.A., Paraciani, C., Wilhelm, S., Raven, F., Meijer, E.L., Kas, M.J.H., Ramirez, S., Meerlo, P., and Havekes, R. (2023). Recovering object-location memories after sleep deprivation-induced amnesia. Curr. Biol. *33*, 298-308.e5. https://doi.org/10.1016/j.cub.2022.12.006.

8. Autore, L., O’Leary, J.D., Ortega-de San Luis, C., and Ryan, T.J. (2023). Adaptive expression of engrams by retroactive interference. Cell Rep. *42*, 112999. https://doi.org/10.1016/j.celrep.2023.112999.

9. Wixted, J.T. (2004). The Psychology and Neuroscience of Forgetting. Annu. Rev. Psychol. *55*, 235–269. https://doi.org/10.1146/annurev.psych.55.090902.141555.

10. Anderson, M.C., and Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. Trends Cogn. Sci. *18*, 279–292. https://doi.org/10.1016/j.tics.2014.03.002.

11. Roy, D.S., Arons, A., Mitchell, T.I., Pignatelli, M., Ryan, T.J., and Tonegawa, S. (2016). Memory retrieval by activating engram cells in mouse models of early Alzheimer’s disease. Nature *531*, 508–512. https://doi.org/10.1038/nature17172.

12. Abraham, W.C., Jones, O.D., and Glanzman, D.L. (2019). Is plasticity of synapses the mechanism of long-term memory storage? Npj Sci. Learn. *4*, 9. https://doi.org/10.1038/s41539-019-0048-y.

13. Clopath, C. (2012). Synaptic consolidation: an approach to long-term learning. Cogn. Neurodyn. *6*, 251–257. https://doi.org/10.1007/s11571-011-9177-6.

14. Aleman-Zapata, A., Morris, R.G.M., and Genzel, L. (2022). Sleep deprivation and hippocampal ripple disruption after one-session learning eliminate memory expression the next day. Proc. Natl. Acad. Sci. *119*, e2123424119. https://doi.org/10.1073/pnas.2123424119.

15. Refaeli, R., Kreisel, T., Groysman, M., Adamsky, A., and Goshen, I. (2023). Engram stability and maturation during systems consolidation. Curr. Biol., S0960982223009818. https://doi.org/10.1016/j.cub.2023.07.042.

16. Brodt, S., Gais, S., Beck, J., Erb, M., Scheffler, K., and Schönauer, M. (2018). Fast track to the neocortex: A memory engram in the posterior parietal cortex. Science *362*, 1045–1048. https://doi.org/10.1126/science.aau2528.

17. Folvik, L., Sneve, M.H., Ness, H.T., Vidal-Piñeiro, D., Raud, L., Geier, O.M., Walhovd, K.B., and Fjell, A.M. (2023). Sustained upregulation of widespread hippocampal–neocortical coupling following memory encoding. Cereb. Cortex *33*, 4844–4858. https://doi.org/10.1093/cercor/bhac384.

18. Goto, A., Bota, A., Miya, K., Wang, J., Tsukamoto, S., Jiang, X., Hirai, D., Murayama, M., Matsuda, T., McHugh, T.J., et al. (2021). Stepwise synaptic plasticity events drive the early phase of memory consolidation. Science *374*, 857–863. https://doi.org/10.1126/science.abj9195.

19. Willems, T., and Henke, K. (2021). Imaging human engrams using 7 Tesla magnetic resonance imaging. Hippocampus *31*, 1257–1270. https://doi.org/10.1002/hipo.23391.

20. Leiker, E.K., and Johnson, J.D. (2014). Neural reinstatement and the amount of information recollected. Brain Res. *1582*, 125–138. https://doi.org/10.1016/j.brainres.2014.07.026.

21. Staresina, B.P., Henson, R.N.A., Kriegeskorte, N., and Alink, A. (2012). Episodic Reinstatement in the Medial Temporal Lobe. J. Neurosci. *32*, 18150–18156. https://doi.org/10.1523/JNEUROSCI.4156-12.2012.

22. Wing, E.A., Ritchey, M., and Cabeza, R. (2015). Reinstatement of Individual Past Events Revealed by the Similarity of Distributed Activation Patterns during Encoding and Retrieval. J. Cogn. Neurosci. *27*, 679–691. https://doi.org/10.1162/jocn_a_00740.

23. Xue, G. (2018). The Neural Representations Underlying Human Episodic Memory. Trends Cogn. Sci. *22*, 544–561. https://doi.org/10.1016/j.tics.2018.03.004.

24. Lim, Y.-L., Lang, D.J., and Diana, R.A. (2023). Cognitive tasks affect the relationship between representational pattern similarity and subsequent item memory in the hippocampus. NeuroImage *277*, 120241. https://doi.org/10.1016/j.neuroimage.2023.120241.

25. LaRocque, K.F., Smith, M.E., Carr, V.A., Witthoft, N., Grill-Spector, K., and Wagner, A.D. (2013). Global Similarity and Pattern Separation in the Human Medial Temporal Lobe Predict Subsequent Memory. J. Neurosci. *33*, 5466–5474. https://doi.org/10.1523/jneurosci.4293-12.2013.

26. Lee, S.-H., Kravitz, D.J., and Baker, C.I. (2019). Differential Representations of Perceived and Retrieved Visual Information in Hippocampus and Cortex. Cereb. Cortex *29*, 4452–4461. https://doi.org/10.1093/cercor/bhy325.

27. Liu, J., Zhang, H., Yu, T., Ren, L., Ni, D., Yang, Q., Lu, B., Zhang, L., Axmacher, N., and Xue, G. (2021). Transformative neural representations support long-term episodic memory. Sci. Adv. *7*, eabg9715. https://doi.org/10.1126/sciadv.abg9715.

28. Xiao, X., Zhou, Y., Liu, J., Ye, Z., Yao, L., Zhang, J., Chen, C., and Xue, G. (2020). Individual-specific and shared representations during episodic memory encoding and retrieval. NeuroImage *217*, 116909. https://doi.org/10.1016/j.neuroimage.2020.116909.

29. Lei, B., Kang, B., Hao, Y., Yang, H., Zhong, Z., Zhai, Z., and Zhong, Y. (2025). Reconstructing a new hippocampal engram for systems reconsolidation and remote memory updating. Neuron *113*, 471-485.e6. https://doi.org/10.1016/j.neuron.2024.11.010.

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

31. Gabrieli, J.D.E. (1998). COGNITIVE NEUROSCIENCE OF HUMAN MEMORY. Annu. Rev. Psychol. *49*, 87–115. https://doi.org/10.1146/annurev.psych.49.1.87.

32. Tulving, E. (2002). Episodic Memory: From Mind to Brain. Annu. Rev. Psychol. *53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114.

33. Squire, L.R., and Dede, A.J.O. (2015). Conscious and unconscious memory systems. Cold Spring Harb. Perspect. Med. *5*, a021667. https://doi.org/10.1101/cshperspect.a021667.

34. Henke, K. (2010). A model for memory systems based on processing modes rather than consciousness. Nat. Rev. Neurosci. *11*, 523–532. https://doi.org/10.1038/nrn2850.

35. Hannula, D.E. (2010). Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. Front. Hum. Neurosci. *4*. https://doi.org/10.3389/fnhum.2010.00166.

36. Reder, L.M., Park, H., and Kieffaber, P.D. (2009). Memory Systems Do Not Divide on Consciousness: Reinterpreting Memory in Terms of Activation and Binding. Psychol. Bull. *135*, 23–49. https://doi.org/10.1037/a0013974.

37. Guskjolen, A., Josselyn, S.A., and Frankland, P.W. (2017). Age-dependent changes in spatial memory retention and flexibility in mice. Neurobiol. Learn. Mem. *143*, 59–66. https://doi.org/10.1016/j.nlm.2016.12.006.

38. Kark, S.M., Slotnick, S.D., and Kensinger, E.A. (2020). Forgotten but not gone: FMRI evidence of implicit memory for negative stimuli 24 hours after the initial study episode. Neuropsychologia *136*, 107277. https://doi.org/10.1016/j.neuropsychologia.2019.107277.

39. Reber, T.P., Luechinger, R., Boesiger, P., and Henke, K. (2012). Unconscious Relational Inference Recruits the Hippocampus. J. Neurosci. *32*, 6138–6148. https://doi.org/10.1523/JNEUROSCI.5639-11.2012.

40. Duss, S.B., Reber, T.P., Hänggi, J., Schwab, S., Wiest, R., Müri, R.M., Brugger, P., Gutbrod, K., and Henke, K. (2014). Unconscious relational encoding depends on hippocampus. Brain *137*, 3355–3370. https://doi.org/10.1093/brain/awu270.

41. Pacozzi, L., Knüsel, L., Ruch, S., and Henke, K. (2022). Inverse forgetting in unconscious episodic memory. Sci. Rep. *12*, 20595. https://doi.org/10.1038/s41598-022-25100-w.

42. Schneider, E., Züst, M.A., Wuethrich, S., Schmidig, F., Klöppel, S., Wiest, R., Ruch, S., and Henke, K. (2021). Larger capacity for unconscious versus conscious episodic memory. Curr. Biol. *31*, 3551-3563.e9. https://doi.org/10.1016/j.cub.2021.06.012.

43. Schmidig, F., Ruch, S., and Henke, K. (2023). Episodic long-term memory formation during slow-wave sleep. https://doi.org/10.7554/elife.89601.1.

44. Anderson, M. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. J. Mem. Lang. *49*, 415–445. https://doi.org/10.1016/j.jml.2003.08.006.

45. Baddeley, A., Eysenck, M.W., and Anderson, M.C. (2015). Memory (Taylor & Francis).

46. Ebbinghaus, H. (1885). Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie (Duncker & Humblot).

47. Murre, J.M.J., and Chessa, A.G. (2023). Why Ebbinghaus' savings method from 1885 is a very 'pure' measure of memory performance. Psychon. Bull. Rev. *30*, 303–307. https://doi.org/10.3758/s13423-022-02172-3.

48. Murre, J.M.J., and Dros, J. (2015). Replication and Analysis of Ebbinghaus' Forgetting Curve. PLOS ONE *10*, e0120644. https://doi.org/10.1371/journal.pone.0120644.

49. Taylor, P.A., Aggarwal, H., Bandettini, P., Barilari, M., Bright, M., Caballero-Gaudes, C., Calhoun, V., Chakravarty, M., Devenyi, G., Evans, J., et al. (2025). Go Figure: Transparency in neuroscience images preserves context and clarifies interpretation.

50. Lu, F., Yang, W., and Qiu, J. (2023). Neural bases of motivated forgetting of autobiographical memories. Cogn. Neurosci. *14*, 15–24. https://doi.org/10.1080/17588928.2022.2136150.

51. Danieli, K., Guyon, A., and Bethus, I. (2023). Episodic Memory formation: A review of complex Hippocampus input pathways. Prog. Neuropsychopharmacol. Biol. Psychiatry *126*, 110757. https://doi.org/10.1016/j.pnpbp.2023.110757.

52. Bogler, C., Zangrossi, A., Miller, C., Sartori, G., and Haynes, J. (2024). Have you been there before? Decoding recognition of spatial scenes from FMRI signals in precuneus. Hum. Brain Mapp. *45*, e26690. https://doi.org/10.1002/hbm.26690.

53. Spaniol, J., Davidson, P.S.R., Kim, A.S.N., Han, H., Moscovitch, M., and Grady, C.L. (2009). Event-related fMRI studies of episodic encoding and retrieval: Meta-analyses using activation likelihood estimation. Neuropsychologia *47*, 1765–1779. https://doi.org/10.1016/j.neuropsychologia.2009.02.028.

54. Cabeza, R., Prince, S.E., Daselaar, S.M., Greenberg, D.L., Budde, M., Dolcos, F., LaBar, K.S., and Rubin, D.C. (2004). Brain Activity during Episodic Retrieval of Autobiographical and Laboratory Events: An fMRI Study using a Novel Photo Paradigm. J. Cogn. Neurosci. *16*, 1583–1594. https://doi.org/10.1162/0898929042568578.

55. Wuethrich, S., Hannula, D.E., Mast, F.W., and Henke, K. (2018). Subliminal encoding and flexible retrieval of objects in scenes. Hippocampus *28*, 633–643. https://doi.org/10.1002/hipo.22957.

56. Ryan, T.J., Ortega-de San Luis, C., Pezzoli, M., and Sen, S. (2021). Engram cell connectivity: an evolving substrate for information storage. Curr. Opin. Neurobiol. *67*, 215–225. https://doi.org/10.1016/j.conb.2021.01.006.

57. Hebscher, M. Enhanced reinstatement of naturalistic event memories due to hippocampal-network-targeted stimulation. 16.

58. Brunec, I.K., Robin, J., Olsen, R.K., Moscovitch, M., and Barense, M.D. (2020). Integration and differentiation of hippocampal memory traces. Neurosci. Biobehav. Rev. *118*, 196–208. https://doi.org/10.1016/j.neubiorev.2020.07.024.

59. Kirwan, C.B., Ashby, S.R., and Nash, M.I. (2014). Remembering and imagining differentially engage the hippocampus: A multivariate fMRI investigation. Cogn. Neurosci. *5*, 177–185. https://doi.org/10.1080/17588928.2014.933203.

60. Ko, S.Y., Rong, Y., Ramsaran, A.I., Chen, X., Rashid, A.J., Mocle, A.J., Dhaliwal, J., Awasthi, A., Guskjolen, A., Josselyn, S.A., et al. (2025). Systems consolidation reorganizes hippocampal engram circuitry. Nature. https://doi.org/10.1038/s41586-025-08993-1.

61. Goto, A. (2022). Synaptic plasticity during systems memory consolidation. Neurosci. Res. *183*, 1–6. https://doi.org/10.1016/j.neures.2022.05.008.

62. Bota, A., Goto, A., Tsukamoto, S., Schmidt, A., Wolf, F., Luchetti, A., Nakai, J., Hirase, H., and Hayashi, Y. (2021). Shared and unique properties of place cells in anterior cingulate cortex and hippocampus. Preprint, https://doi.org/10.1101/2021.03.29.437441 https://doi.org/10.1101/2021.03.29.437441.

63. Tse, D., Langston, R.F., Kakeyama, M., Bethus, I., Spooner, P.A., Wood, E.R., Witter, M.P., and Morris, R.G.M. (2007). Schemas and Memory Consolidation. Science *316*, 76–82. https://doi.org/10.1126/science.1135935.

64. Nieto-Castanon, A., and Whitfield-Gabrieli, S. (2022). CONN functional connectivity toolbox: RRID SCR_009550, release 22 22nd ed. (Hilbert Press) https://doi.org/10.56441/hilbertpress.2246.5840.

65. Ghosh, S.S., Kent, J.D., Erramuzpe, A., DuPre, E., Snyder, M., Blair, R.W., Moodie, C.A., Esteban, O., Durnez, J., Isik, A.I., et al. (2018). fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat. Methods *16*, 111–116. https://doi.org/10.1038/s41592-018-0235-4.

66. Fischl, B. (2012). FreeSurfer. NeuroImage *62*, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021.

67. Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. *16*, 225–237. https://doi.org/10.3758/PBR.16.2.225.

68. Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., and Ghosh, S.S. (2011). Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. Front. Neuroinformatics *5*. https://doi.org/10.3389/fninf.2011.00013.

69. Gorgolewski, K.J., Esteban, O., Markiewicz, C.J., Ziegler, E., Ellis, D.G., Notter, M.P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., et al. (2018). Nipype. Software.

70. Tustison, N.J., Avants, B.B., Cook, P.A., Yuanjie Zheng, Egan, A., Yushkevich, P.A., and Gee, J.C. (2010). N4ITK: Improved N3 Bias Correction. IEEE Trans. Med. Imaging *29*, 1310–1320. https://doi.org/10.1109/TMI.2010.2046908.

71. Avants, B., Epstein, C., Grossman, M., and Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. *12*, 26–41. https://doi.org/10.1016/j.media.2007.06.004.

72. Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging *20*, 45–57. https://doi.org/10.1109/42.906424.

73. Reuter, M., Rosas, H.D., and Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. NeuroImage *53*, 1181–1196. https://doi.org/10.1016/j.neuroimage.2010.07.020.

74. Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. NeuroImage *9*, 179–194. https://doi.org/10.1006/nimg.1998.0395.

75. Klein, A., Ghosh, S.S., Bao, F.S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Chaibub Neto, E., et al. (2017). Mindboggling morphometry of human brains. PLOS Comput. Biol. *13*, e1005350. https://doi.org/10.1371/journal.pcbi.1005350.

76. Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., and Collins, D.L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage *47*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5.

77. Evans, A.C., Janke, A.L., Collins, D.L., and Baillet, S. (2012). Brain templates and atlases. NeuroImage *62*, 911–922. https://doi.org/10.1016/j.neuroimage.2012.01.024.

78. Greve, D.N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. NeuroImage *48*, 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060.

79. Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage *17*, 825–841. https://doi.org/10.1006/nimg.2002.1132.

80. Cox, R.W., and Hyde, J.S. (1997). Software tools for analysis and visualization of fMRI data. NMR Biomed. *10*, 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L.

81. Pruim, R.H.R., Mennes, M., Rooij, D. van, Llera, A., Buitelaar, J.K., and Beckmann, C.F. (2015). ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. NeuroImage *112*, 267–277. https://doi.org/10.1016/j.neuroimage.2015.02.064.

82. Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. NeuroImage *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048.

83. Behzadi, Y., Restom, K., Liau, J., and Liu, T.T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. NeuroImage *37*, 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042.

84. Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J., Calkins, M.E., Eickhoff, S.B., Hakonarson, H., Gur, R.C., Gur, R.E., et al. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. NeuroImage *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052.

85. Lanczos, C. (1964). Evaluation of Noisy Data. J. Soc. Ind. Appl. Math. Ser. B Numer. Anal. *1*, 76–85. https://doi.org/10.1137/0701007.

86. Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., and Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. Front. Neuroinformatics *8*. https://doi.org/10.3389/fninf.2014.00014.

87. Nilearn nilearn. https://doi.org/10.5281/zenodo.8397156 https://doi.org/10.5281/zenodo.8397156.

88. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830. https://doi.org/10.48550/arXiv.1201.0490.

89. Wetzels, R., and Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. Psychon. Bull. Rev. *19*, 1057–1064. https://doi.org/10.3758/s13423-012-0295-x.

964 **STAR★METHODS**

965 **KEY RESOURCES TABLE**

REAGENT OR RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
CONN Toolbox (version 22a)	Nieto-Castanon et al.[64]	RRID: SCR_009550
fMRIprep (v20.2.3)	Esteban et al.[65]	https://fmriprep.org/en/stable/workflows.html
Freesurfer (v7.1.1)	Fischl et al.[66]	https://surfer.nmr.mgh.harvard.edu
MATLAB (v R2021b)	The MathWorks, Natick, Massachusetts, United States (https://www.mathworks.com/products/matlab.html)	RRID: SCR_001622
MATLAB toolbox "SPM12"	https://www.fil.ion.ucl.ac.uk/spm/software/spm12/	RRID: SCR_007037
RStudio (2023.06.1+524)	RStudio Team	https://www.rstudio.com
R (v4.3.1)	R Core Team	N/A
Presentation (v23.0)	Neurobehavioral Systems, Berkely, California, United States (https://www.neurobs.com)	RRID: SCR_002521

966

967 **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**

968 **Human Participants**

969 Forty healthy, right-handed participants were randomly assigned to two groups. Both groups underwent
970 the same procedures but underwent distinct functional MRI sequences. Twenty participants were
971 examined with a functional MRI sequence with limited FOV covering the medial temporal lobe and
972 adjacent structures (age: 24.30 ± 3.87 (mean ± SD); 10 women). The remaining 20 participants were
973 examined with a whole-brain fMRI sequence (age: 23.8 ± 3.41 (mean ± SD); 11 women). Participants
974 were recruited through advertisements on university- and social-media-platforms. All participants
975 underwent a telephone screening. They reported normal or corrected-to-normal vision, no previous or
976 current neurological or psychiatric diseases, no claustrophobia, and no non-removable metals in or on the
977 body. All participants provided written informed consent for the experiment. Note that the participants'
978 consent was semi-informed. Participants were left unaware of the 24-hour category retrieval of the face-
979 object combinations learned on the first day. We did not inform them of a 24-hour retrieval test to prevent
980 their active rehearsing of the learning material. Instead, we informed participants that they are reinvited to
981 the second MRI session for them to learn and retrieve another set of face-object combinations.
982 Participants were fully debriefed at the end of the study. Participants were compensated with a
983 reimbursement of 150 Swiss Francs and digital copies of the anatomical scans of their own brains. The
984 study was approved by the ethics committee of the Canton of Berne (ID: 50008).

985 **METHOD DETAILS**

986 **Stimuli**

987 The stimulus set selected for this experiment consisted of 96 combinations of faces and objects. Face
988 stimuli were selected from the Flickr-Faces-HQ Dataset (FFHQ, https://github.com/NVlabs/ffhq-dataset).

Selection was restricted to Caucasian faces to reduce bias. Faces were neither standardized nor clipped, and paraphernalia (e.g., a hat or glasses) and the background (e.g., a forest or an office scene) were not removed from the pictures to preserve the stimuli's naturalistic appearance and to allow for better discriminability of faces. We selected images of everyday objects on white backgrounds from two categories: organic (e.g. an apple, a bird) and inorganic (e.g. a hammer, a couch). The set of 96 faces consisted of 48 male and 48 female faces. Half of the male faces were paired with an organic object and the other half with an inorganic object. The same pairing scheme was applied to the female faces.

Stimulus groups

Groups of four stimuli were created because of the constraints of the recognition task: During recognition, participants were presented with a face plus two object stimuli (the target object and a foil object). Their task was to recognize the object that was paired with the face during learning. All foil objects had been presented with different face stimuli during learning and were thus familiar. Recognition required participants to remember which familiar object had been presented with the current face during learning (associative recognition). For stimulus preparation, we created groups of two face and two object stimuli. The two face stimuli of a certain group would be presented with the same two objects from their group during the recognition task, one object being the target and the other the foil for the initially presented face and vice versa for the subsequently presented face. To exclude object bias in the groups, we collected data in a pilot study using an online survey. In this survey, 63 students decided which of two objects would fit better to either one of the two face stimuli using a rating scale from 0-100. Based on this survey, we included those groups in the main experiment with the least amount of object bias, i.e., groups that yielded a mean rating score close to 50. Table S19 shows the mean scores for each of the 48 groups used in the main experiment. The 48 groups provided 96 face-object pairs.

Randomization

Target and foil objects were randomly assigned to the face stimuli within each group. The presentation order and the interval between the two face-object combinations belonging to a group were semi-randomized such that the time interval is maximal between the two combinations that belong to one group. A maximal temporal distance between the two face-object combinations belonging to one group was important for participants' accuracy on the recognition task because on the recognition task the same two objects from each group were presented with each of the two faces from the group. Maximizing the distance between subsequent presentations increased the likelihood that participants would make their decision of which object was paired with a given face during learning solely based on memory and not based on their previous choice of an object. The order, in which the 96 face-object pairs were presented during learning, was randomized with a windowed shuffle for the 30-minute and the 24-hour category retrieval and for the recognition task. The windowed shuffle shuffled 4 list entries before moving the window to the next 4 entries. This ensured that the study-test retention interval was comparable for the 96 memories. Moreover, the shuffling also preserved the distance between the two face-objects pairs from the same group.

Experimental tasks

For all tasks, presentations of stimuli were separated by a jittered interstimulus-interval (ISI) of 500 – 2500 ms. A black fixation cross on a white background was presented during the ISI in the center of the screen. Participants were asked to focus on the fixation cross.

Associative learning task

First, the face stimulus was presented in the center of the screen for 1000 ms. Participants were instructed to look closely at the face and memorize it. Thereafter, the same face was presented again, this time for 2500 ms paired with an object that was drawn from one of two categories: organic or inorganic. Participants were asked to form an association between the presented face and the object. No instructions were given on how to form an association. Following associative learning, the face was presented in the center of the screen for 2750 ms for participants to indicate by button press whether the object the face had just been paired with was organic or inorganic. This immediate cued associative retrieval served to identify trials with successful learning and a correct conscious associative retrieval. Only these trials entered the analysis of the fate of engrams underlying subsequently forgotten memories.

Consolidation Task

A 5-minute pause was given following the immediate retrieval to allow for memory consolidation. During this pause we kept participants busy with a simple attention task (odd-even task) to avoid active rehearsal. We presented numbers between 0 and 100 on the screen for participants to indicate whether a number was odd or even. Corrective feedback was provided by a color cue: the number changed to green for correct and to red for incorrect odd-even responses.

Associative Category Retrieval Task

Each of the 96 face stimuli was presented in the center of the white screen as a retrieval cue for 1000 ms. Then, the face appeared on the screen again, this time for 2750 ms combined with the written instruction to indicate by button press whether the associated object was organic or inorganic. The index finger was assigned to the response button for inorganic and the middle finger was assigned to the response button for organic. Each retrieval trial was followed by a confidence rating regarding the just provided retrieval response. Participants were given 2500 ms to provide their confidence response using a 3-point scale (sure, unsure, guess). The unsure option was the largest category including all responses that were neither completely sure nor pure guesses. We evaluated only the extremes, i.e., the sure and the guess responses because we were interested in contrasting consciously remembered associations (sure responses) with forgotten associations (guess responses).

Associative Recognition Task

Each face was presented with the two objects from its group. The presentation duration was 3500 ms. One of the two objects was the target that had been presented along with the face during learning and the other was a familiar foil object that was presented with another face during learning. Participants were required to indicate which of the two objects had been presented with the face during learning. The two objects were presented twice (though paired with two distinct faces) during the recognition task. Each

object was once the target and once the foil. Following each recognition trial, participants gave their confidence assessment on a 3-point scale (sure, unsure, guess).

Technical Setup

The paradigm was presented with a Hyperion projector from Psychology Software Tools (Sharpsburg, USA) with a resolution of 1920 x 1080p Full HD, a refresh rate of 60 Hz, and dimensions of the projected image being set at 30 cm x 54 cm and a screen diagonal of 62cm. The distance between the participants' eyes and the projection screen was 1.5 m. The experiment was performed using Presentation® software (Version 23.0, Neurobehavioral Systems, Inc., Berkeley, CA) on a Windows computer. Responses were recorded in the MRI scanner with a hand sleeve with a wrist strap and a Response unit for response registration in the MRI scanner named "Celeritas Fiber Optic Response Unit - Right hand, 5 buttons" from Psychology Software Tools (Sharpsburg, USA). Participants used their index, middle, and ring finger of their right hand to submit responses during the task.

Experimental Procedure

The study was conducted on two consecutive days. All participants underwent the same protocol. Whenever participants were required to perform a behavioral task in the scanner, BOLD signal was recorded. The study consisted of 4 behavioral tasks and 4 BOLD fMRI runs. Between runs, fMRI data acquisition was paused. The first three runs were performed consecutively on the first day and the fourth run was performed on the following day. Between the two MRI sessions there was a 24-hour interval, during which participants were free to follow their usual routines. The experiment would usually start between midday and afternoon. The first MRI session lasted about 1 hour, and the second MRI session lasted about 45 minutes.

Day 1

On the first day, participants completed an online training at home to establish familiarity with the behavioral tasks. After arriving at and entering the MRI scanner, a t1-weighted anatomical image was recorded. During the anatomical scan participants rehearsed the behavioral tasks and the button press responses. Then, participants completed the associative learning of 96 face-object pairs, carried out the immediate category retrieval task and then took the odd-even task during the 5-minute pause. Finally, they performed the 30-minutae category retrieval, which marked the end of the first day of experimentation. To prevent active rehearsal of the learned material, participants were misinformed that they would learn another set of face-object pairs on the next day. Participants were asked not to perform cognitively demanding tasks between the two MRI sessions. At the end of the first experimentation day, participants received an e-mail which included a reminder for the second experimentation day as well as a link to a survey to assess sleep duration and sleep quality (Stanford Sleepiness Scale (SSS) and Stanford Sleepiness Test (SST)). Participants were asked to fill out the survey the next morning.

Day 2

When participants arrived at the MR center, they received a debriefing in written form, where we explained that they would not learn a new set of face-object combinations but retrieve the face-object

1100 pairs learned on the previous day. When asked about their expectations for the second day, 32.5% of
1101 participants reported that they had expected a repeated retrieval of the face-object pairs learned on the
1102 previous day. But none of participants reported to have actively rehearsed the face-object pairs.
1103 Participants entered the scanner again on the second day and we recorded a whole-brain anatomical t1-
1104 weighted scan. Next, participants performed the category retrieval task and then the recognition task for
1105 the face-object associations that they had learned on the previous day. Each category retrieval trial was
1106 immediately followed by its corresponding recognition trial.

1107 **MRI data acquisition**

1108 Both anatomical and functional MRI sequences were recorded on a 7T Siemens Magnetom Terra whole
1109 body MRI scanner (Siemens Medical Solutions, Erlangen, Deutschland) with a 32-channel head coil. The
1110 anatomical t1-weighted image acquisition followed a magnetization-prepared two rapid gradient echo
1111 (MP2RAGE) sequence with a time of repetition (TR) of 6000 ms; echo time (TE) 2.06 ms; flip angle (FA)
1112 = 0°; field of view (FOV) = 384 x 384 mm^2, spatial resolution of 0.63 x 0.63 x 0.63 mm^3, 256 sagittal
1113 slices, matrix points = 610. For the t2*-weighted functional MRI scans, we used two distinct sequences.
1114 The small field-of-view (FOV) fMRI sequence covered the bilateral medial temporal lobe with 0.8 mm
1115 isotropic voxel size to target hippocampal engram components and the other fMRI sequence covered the
1116 entire brain with a 1.1 mm isotropic voxel size to target neocortical engram components.

1117 ***Small FOV fMRI sequence***

1118 Isotropic functional T2* weighted blood oxygen level dependent (BOLD) sensitive multiple interleaved
1119 slice acquisition sequence with a TR = 2120 ms; TE = 24 ms; FA = 70°; FOV = 210 x 210 mm^2, spatial
1120 resolution of 0.8 x 0.8 x 0.8 mm^3, 28 transversal slices, matrix points = 263. We acquired a total of 2100
1121 volumes per participant. Day 1: associative learning, 5-minute pause with odd-even task, and associative
1122 category retrieval: 510, 150, 515; day 2: associative category retrieval task and associative recognition
1123 task: 925.

1124 ***Whole-brain fMRI sequence***

1125 Isotropic functional T2* weighted blood oxygen level dependent (BOLD) sensitive multiple interleaved
1126 slice acquisition sequence with a TR = 2700 ms; TE = 50 ms; FA = 90°; FOV = 210 x 210 mm^2, spatial
1127 resolution of 1.1 x 1.1 x 1.1 mm^3, 96 transversal slices; matrix points = 191. A total of 1650 volumes were
1128 collected per participant. Day 1: associative learning, 5-minute pause with odd-even task, and associative
1129 category retrieval: 400, 119, 405, day 2: associative category retrieval task and associative recognition
1130 task: 726.

1131 **QUANTIFICATION AND STATISTICAL ANALYSIS**

1132 **Statistical analysis of behavioral data**

1133 Behavioral data were preprocessed and analyzed with custom R scripts using RStudio and the tidyverse
1134 package on a MacBook Pro. Face-object pairs were removed from the analysis if participants answered
1135 incorrectly during immediate category retrieval, answered incorrectly but with high confidence (sure
1136 response) or if they pressed the wrong button. The significance threshold was set to $p < 0.05$, if not

1137 indicated otherwise. The Bayes factor was calculated using the “BayesFactor” R-library which makes use
1138 of the function described by Rouder[67] and colleagues.

1139 **Preprocessing of fMRI data**
1140 Whole-brain fMRI sequence: Results included in this manuscript for the whole-brain fMRI sequence come
1141 from preprocessing performed using fMRIPrep 20.2.3[65] (RRID: SCR _016216), which is based on Nipype
1142 1.6.1 [68,69] (RRID _SCR002502).

1143 **Anatomical data preprocessing**
1144 A total of 2 T1-weighted (T1w) images were acquired per participant within the input BIDS dataset. All of
1145 them were corrected for intensity non-uniformity (INU) with N4Bias FieldCorrection[70], distributed with
1146 ANTs 2.3.3 [71], (RRID:SCR_004757). The T1w-reference was then skull-stripped with a Nipype
1147 implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target
1148 template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter
1149 (GM) was performed on the brain-extracted T1w using fast[72] (FSL 5.0.9, RRID:SCR _002823). A T1w-
1150 reference map was computed after registration of 2 T1w images (after INU-correction) using mri _robust
1151 _template[73] (FreeSurfer 6.0.1). Brain surfaces were reconstructed using recon-all[74] (FreeSurfer 6.0.1,
1152 RRID:SCR _001847), and the brain mask estimated previously was refined with a custom variation of the
1153 method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of
1154 Mindboggle[75] (RRID:SCR _002438). Volume-based spatial normalization to two standard spaces
1155 (MNI152NLin2009cAsym, MNI152NLin6Asym) was performed through nonlinear registration with
1156 antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w
1157 template. The following templates were selected for spatial normalization: ICBM 152 Nonlinear
1158 Asymmetrical template version 2009c[76] [RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym],
1159 FSL’s MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration
1160 Model[77] [RRID:SCR _002823; TemplateFlow ID: MNI152NLin6Asym].

1161 **Functional data preprocessing (whole-brain fMRI sequence)**
1162 For each of the 4 BOLD runs acquired per participant (across all tasks and sessions), the following
1163 preprocessing was performed. First, a reference volume and its skull-stripped version were generated
1164 using a custom methodology of fMRIPrep. Susceptibility distortion correction (SDC) was omitted. The
1165 BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which
1166 implements boundary-based registration.[78] Co-registration was configured with six degrees of freedom.
1167 Head-motion parameters with respect to the BOLD reference (transformation matrices, and six
1168 corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering
1169 using mcflirt[79] (FSL 5.0.9). BOLD runs were slice-time corrected using 3dTshift from AFNI 20160207[80]
1170 (RRID:SCR _005927). The BOLD time-series were resampled onto the following surfaces (FreeSurfer
1171 reconstruction nomenclature): fsnative, fsaverage5. The BOLD time-series (including slice-timing
1172 correction when applied) were resampled onto their original, native space by applying the transforms to
1173 correct for headmotion. These resampled BOLD time-series will be referred to as preprocessed BOLD in

original space, or just preprocessed BOLD. The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Automatic removal of motion artifacts using independent component analysis (ICA-AROMA[81]) was performed on the preprocessed BOLD on MNI space after removal of non-steady state volumes and spatial smoothing with an isotropic Gaussian kernel of 6 mm FWHM (full-width half-maximum). Corresponding “non-aggresively” denoised runs were produced after such smoothing. Additionally, the “aggressive” noise-regressors were collected and placed in the corresponding confounds file. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power[82] (absolute sum of relative motions) and Jenkinson[79] (relative root mean square displacement between affines). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power[82] and colleagues). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor[83]). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi and colleagues in that insteadof eroding the masks by 2 pixels on BOLD space, the aCompCor masks are subtracted a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the FreeSurfer’s aseg segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks.

For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components’ time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each.[84] Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing

effects of other kernels.[85] Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.8.189[86] (RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

Copyright waiver

The above boilerplate text was automatically generated by fMRIPrep with the express intention that users should copy and paste this text into their manuscripts unchanged. It is released under the CC0 license.

Functional data preprocessing (small FOV sequence)

Due to the small FOV of the sequence the use of fMRIprep was not recommended. Preprocessing was done with SPM12 (Welcome Department of Cognitive Neurology, London, UK). The functional volumes were slice-time corrected and realigned to the mean functional volume image. Coregistration to the participants anatomical volume was carried out with a non-skull-stripped structural volume. The latter allowed for a more reliable coregistration because the skull provided for more anatomical information to properly match the small FOV with the anatomical template. Volumes were segmented and spatially normalized to the MNI/ICBM AVG152 Template provided by SPM. Functional volumes were smoothed with a 6 mm (FWHM) isotropic Gaussian Kernel.

Freesurfer segmentations

We created subject-specific ROIs for both the functional connectivity analyses as well as the RSA analyses with freesurfer. FMRIprep already applies *recon_all* during its workflow which computes ROIs for the neocortex. Hippocampal ROIs were created with freesurfers *segmentHA_T1* function. The creation of neocortical masks failed for 3 participants from the whole-brain fMRI group. They were thus excluded from the representational similarity analysis which used the masks for region of interest analyses.

GLM

In the following, both fMRI sequences (small FOV and whole-brain sequence) were always treated identically unless stated otherwise.

First Level

First-level analysis was performed using a general linear model (GLM) that modeled the BOLD-time-series using the sequence of event onsets convolved with the default hemodynamic response function (HRF) provided by SPM12. Events were modeled at a duration of zero. For each participant, 4 timeseries were estimated in the same first-level model. All stimulus presentations were modeled as events, the confidence ratings were modeled as events, and the button presses given during the category retrieval task and during recognition task were modeled as events. These events were categorized into 5 conditions depending on the responses given during the retrieval task: responses to every trial could be categorized as correct sure responses, incorrect unsure responses, correct unsure responses, incorrect guess responses or correct guess responses. For example, the 30-minute retrieval responses were used

for the categorization of corresponding trials during the encoding time-series, to investigate subsequent memory effects: the presentation of the face stimulus during encoding was modeled as an event and categorized based on the response during the retrieval of said face. Sure but incorrect answers were not estimated because their rate was too low for a robust estimate. Six movement regressors (x, y, z, pitch, yaw, roll) were added to the model for every session as well as a session constant which resulted in 129 estimated beta maps for every participant. Implicit masking with the default SPM threshold of 0.8 resulted in the exclusion of signal from the center of the brain, mainly in the whole-brain fMRI sequence. In both sequences, we used a binary brain mask, generated by SPMs segmentation algorithm, as an explicit mask and set SPM's masking threshold to [-inf], to include all voxels within the mask in the first level estimation. Note that all events were used for first level specification, including trials that were excluded during the behavioral analysis (e.g. false memories, see above), to allow for a better baseline estimation of the general linear model.

Second Level

Second-level analyses for group statistics were computed using a full factorial model with two factors. The first factor modeled 5 events of interest: the first presentation of the face stimulus during the learning task, the associative learning event, the presentation of the face stimulus during the 30-minute as well as the 24-hour category retrieval task, and the recognition task. The second factor modeled the retrieval quality for 3 different levels, correct sure responses, correct guessing responses, and incorrect guessing responses. This resulted in a 5 x 3 second-level full factorial design. The significance level for the calculation of the t-contrasts was set to $p < 0.001$, uncorrected, with a cluster extent threshold of $k = 10$ voxels, if not indicated otherwise.

Brain-behavior correlations

To isolate brain activity that was associated with task accuracy we used a multiple regression design within SPM. For all the models calculated we included the estimation of an intercept. The generated SPM contrast images were combined with a between-subjects behavioral covariate in a multiple-regression model. The behavioral covariate used for contrasts of guess responses was the z-scored percentage correct of all guess responses (guessing accuracy), for contrasts of sure responses the z-scored absolute number of correct sure responses (sure accuracy). For all second-level models we calculated both positive and negative brain-behavior correlations. The significance level set to $p<0.001$, uncorrected, with a cluster extent threshold of $k = 10$ voxels, if not indicated otherwise.

Stacked contrasts

To uncover both differences between retrieval quality (sure > guess) as well as retrieval timepoints (24-hour > 30-minute category retrieval, we calculated stacked contrasts by contrasting 4 parameter estimates of the corresponding retrieval timepoints and qualities within SPM12s result interface. The significance threshold was set to $p < 0.001$, uncorrected, with a cluster extent threshold of $k = 10$ voxels if not indicated otherwise.

1282 **Conjunction analyses**

1283 To uncover commonalities of brain-behavior correlations for consciously accessible and for forgotten
1284 memories, we calculated conjunction analyses. We used the above generated second level multiple
1285 regression models and extracted a binary mask of all the voxels that were significant under the threshold
1286 of $p<0.001$ and an extent of $k = 10$ voxels. Then we took these maps as inclusive masks in another
1287 second level model of interest and set the threshold again to $p<0.001$ with a cluster extent of $k = 10$. The
1288 significant clusters in the resulting t-maps only revealed the overlay of all the voxels that would be
1289 significant in both models independently of each other. To uncover commonalities of contrast analyses,
1290 we used the contrast maps produced by the second level full factorial model to calculate commonalities
1291 between activation patterns using SPM12. The significance threshold was set to $p < 0.001$, uncorrected,
1292 with a cluster extent threshold of $k = 10$ voxels if not indicated otherwise.

1293 **Functional connectivity analyses**

1294 To uncover task-related connectivity changes between the hippocampus and the neocortex we calculated
1295 general psycho-physiological interactions (gPPI) analyses. We used the conn toolbox (version 22a).
1296 Unlike the standard PPI analyses provided by SPM12, the conn toolbox (using a gPPI approach) can
1297 model effects of more than two conditions, to allow for interactions between conditions, and to make no
1298 assumptions how these conditions are related to the baseline condition. First level specifications were
1299 imported from SPM to the conn toolbox. Behavioral measures (sure and guessing accuracy) served as
1300 second-level covariates. We defined seed regions within the subdivided left and right hippocampus. More
1301 precisely, we used grey matter masks of the hippocampus head, body and tail generated by freesurfer.
1302 These seed regions were integrated as participant-specific ROIs. Conn's integrated setup and denoising
1303 pipeline was applied to the data with a bandpass filter between 0.008 – 0.09 and linear detrending but no
1304 despiking. We also included the second-level covariates to be able to correlate task related changes in
1305 connectivity with behavioral parameters like the accuracy of the guess responses given at the 30-minute
1306 and 24-hour category retrieval. We included this in a seed-to-voxel analysis with an underlying
1307 multivariate regression model. For conjunctions of multivariate regression functional connectivity
1308 analyses, we made use of the shared architecture of the two toolboxes, conn and SPM12. We used the
1309 contrast files created by the multivariate regression functional connectivity analyses to calculate
1310 commonalities between them using SPM12.

1311 **Representational similarity analysis**

1312 Methodology was modeled after the protocol described by Hebscher and colleagues.[57] One notable
1313 difference was how single-trial beta estimates were created. While Hebscher and colleagues used a
1314 least-squares-single (LSS, individual GLMs for each trial), we used a least-squares-all (LSA, one GLM
1315 with one condition per trial). We examined reinstatement of memory-specific patterns of neural activity by
1316 conducting a series of RSAs on patterns of estimated neural activity within ROIs. RSAs measured
1317 encoding-retrieval (or encoding-recognition) similarity by computing the Pearson correlation between
1318 patterns for all pairs of trials for each ROI. All trials were included in the analyses, independent of the

confidence rating given by the participant during the retrieval. Pairwise correlation values were Fisher transformed. We then subtracted the mean of all correlations calculated for non-matching trial-pairs from the mean correlation of matching trial-pairs. Matching trials refers to trials containing a presentation of the same face during encoding and retrieval. This metric was z-scored against a null distribution of correlation difference scores, calculated by randomly shuffling trials over 1000 permutations. Group-level statistics were performed on these z-scored pattern similarity scores. Finally, to correct for similarity based on face re-presentation and based on the mere task characteristics or encoding/retrieval effort, we subtracted the similarity values of encoding-retrieval pairs that yielded incorrect guess responses from the similarity values of encoding-retrieval pairs that yielded correct sure responses and correct guess responses. This way, we came up with similarity scores that reflected encoding-retrieval or retrieval-retrieval voxel pattern reinstatements. Pair-wise representational similarity analysis (RSA) was conducted with custom python scripts making use of the software package Nilearn[87]. Correlation values were calculated with the *pairwise_distances* function of the software package scikit-learn[88]. To assess the significance of voxel pattern reinstatement, we computed one-sample two-tailed t-tests on the z-scored pattern similarity values for each ROI. Similarity estimates were also correlated across subjects with retrieval accuracy. Results below a threshold of $p < 0.05$ were considered significant. In addition, we computed the Bayes Factor for all t-tests and across-subjects Pearson correlation using the “pingouin” python package, which computes Wetzels[89] Bayes factor.