The episodic engram transformed: Time reduces retrieval-related brain activity but correlates it with memory accuracy

Orit Furman, Avi Mendelsohn and Yadin Dudai

Supplemental Material	http://learnmem.cshlp.org/content/suppl/2012/11/07/19.12.575.DC1.html
References	This article cites 64 articles, 18 of which can be accessed free at: http://learnmem.cshlp.org/content/19/12/575.full.html#ref-list-1
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Research

The episodic engram transformed: Time reduces retrieval-related brain activity but correlates it with memory accuracy

Orit Furman,[1] Avi Mendelsohn, and Yadin Dudai

Department of Neurobiology, The Weizmann Institute of Science, Rehovot 76100, Israel

We took snapshots of human brain activity with fMRI during retrieval of realistic episodic memory over several months. Three groups of participants were scanned during a memory test either hours, weeks, or months after viewing a documentary movie. High recognition accuracy after hours decreased after weeks and remained at similar levels after months. In contrast, BOLD activity in a retrieval-related set of brain areas during correctly remembered events was similar after hours and weeks but significantly declined after months. Despite this reduction, BOLD activity in retrieval-related regions was positively correlated with recognition accuracy only after months. Hippocampal engagement during retrieval remained similar over time during recall but decreased in recognition. Our results are in line with the hypothesis that hippocampus subserves retrieval of real-life episodic memory long after encoding, its engagement being dependent on retrieval demands. Furthermore, our findings suggest that over time episodic engrams are transformed into a parsimonious form capable of supporting accurate retrieval of the crux of events, arguably a critical goal of memory, with only minimal network activation.

[Supplemental material is available for this article.]

That the episodic memory trace undergoes extensive changes over long periods of time is considered a given, yet the nature of these changes remains unclear. This question has cardinal implications concerning the understanding of brain mechanisms of memory in general and the stability, retrievability, and veracity of episodic memory in particular. The aforementioned changes are usually construed within the conceptual framework of the consolidation hypothesis, which posits that items in memory undergo maturational shifts affecting their quality and susceptibility to amnesic agents (Dudai 2004). On top of the universally accepted process of cellular consolidation, which concludes within hours, declarative memories are postulated to undergo systems consolidation, which lasts for weeks or longer, and appears to involve migration of information within memory networks (Frankland and Bontempi 2005; Dudai 2012).

Models of systems consolidation are classified on the basis of the postulated extent of hippocampal engagement. The standard consolidation theory (SCT) (McClelland et al. 1995; Squire et al. 2004) posits that memory ultimately becomes reliant solely on neocortex. The multiple trace theory (MTT) (Moscovitch and Nadel 1998) posits that consolidation generates in cortico-hippocampal networks multiple traces, converting memory to a widely distributed form resistant to hippocampal damage. Building on MTT, the trace transformation theory (Winocur et al. 2010; Winocur and Moscovitch 2011) proposes that episodic context-specific memory remains dependent on the hippocampus throughout, but time promotes the generation in the neocortex of a semantic gist-only version of the memory; the type of representation retrieved depends on the circumstances (Dudai 2012).

Our aim in this study was to examine brain signatures of retrieval of naturalistic episodes over time and correlate them with

memory performance, in order to unveil the potential transformations that the episodic engram undergoes during the time interval in which systems consolidation is postulated to take place. Data construed as relevant to systems consolidation do not yet allow delineation of the precise time window in which this hypothetical process takes place. It possibly begins already within the first minutes after encoding (Tambini et al. 2010), and in some cases may even advance considerably within hours to days (Tse et al. 2011). As to completion of the process, if it ever is completed (Dudai 2012), while some animal lesion studies suggest long-term associative memory becomes independent of hippocampal activity already after several weeks (Kim and Fanselow 1992), neuropsychological work in humans suggests that hippocampal dependence may last much longer (Nadel and Moscovitch 1997). Previous work we conducted using a movie memory paradigm revealed differences in memory performance between hours-to-weeks and months post–movie watching (Furman et al. 2007). In the present study, we therefore selected 3 h as an early time point; 3 wk as a time point in which the cortico–hippocampal dialogue assumed to subserve alterations in memory representations progresses significantly (Bontempi et al. 1999; Maviel et al. 2004; Frankland and Bontempi 2005) or systems consolidation is even practically completed in animal models (Kim and Fanselow 1992); and 3 mo as a time point tapping into older memory.

We used an audiovisual, time-evolving, social narrative in the form of a documentary movie as memoranda. We have previously found that a single viewing of such movies is sufficient to encode long-lasting bona fide episodic memories (Furman et al. 2007; Mendelsohn et al. 2010). We decided to probe memory using both recall and recognition because these processes may undergo distinct systems consolidation transformations (Winocur and Moscovitch 2011). In this protocol, episodic retrieval was elicited in two stages: in the first, "mental recall" phase (mRecall), participants were prompted by a movie frame and

[1]**Corresponding author**
E-mail orit.furman@weizmann.ac.il

specific question pertaining to an event in the movie to recall that event, but without providing an overt answer. The omission of the overt reply was because self-generated response via speaking or writing would have severely confounded fMRI signal. This mRecall phase was immediately followed by a content Recognition test, wherein two alternative verbal answers were presented with the same picture and question cue (Fig. 1C). To confirm the answer reached during the mRecall phase, participants made a correspondence report after each Recognition question. This report disclosed whether the answer chosen during Recognition was the same as the answer reached covertly during the mRecall phase (dubbed "matched" item) or differed from it ("nonmatched" item). We examined time-dependent changes in neural activity supporting retrieval for each retrieval phase/item type separately.

Different groups of participants were first scanned during movie watching and then scanned again while they completed the memory questionnaire after hours, weeks, and months. Since we wished to test the change of memory over time with minimal experience-dependent confounds, the participants in our protocol were not retested, because the retrieval experience itself provides relearning and reconsolidation experiences that are likely to affect subsequent performance and its supporting neural activity (Karpicke and Roediger 2008; Nader and Hardt 2009; Roediger and Butler 2011).

We identified a set of brain areas supporting retrieval consistently over a period of months, including medial temporal lobe (MTL) and neocortical areas, similar to the previously reported autobiographical memory (AM) network (Svoboda et al. 2006) (see also activation of this network during retrieval of movie memory)
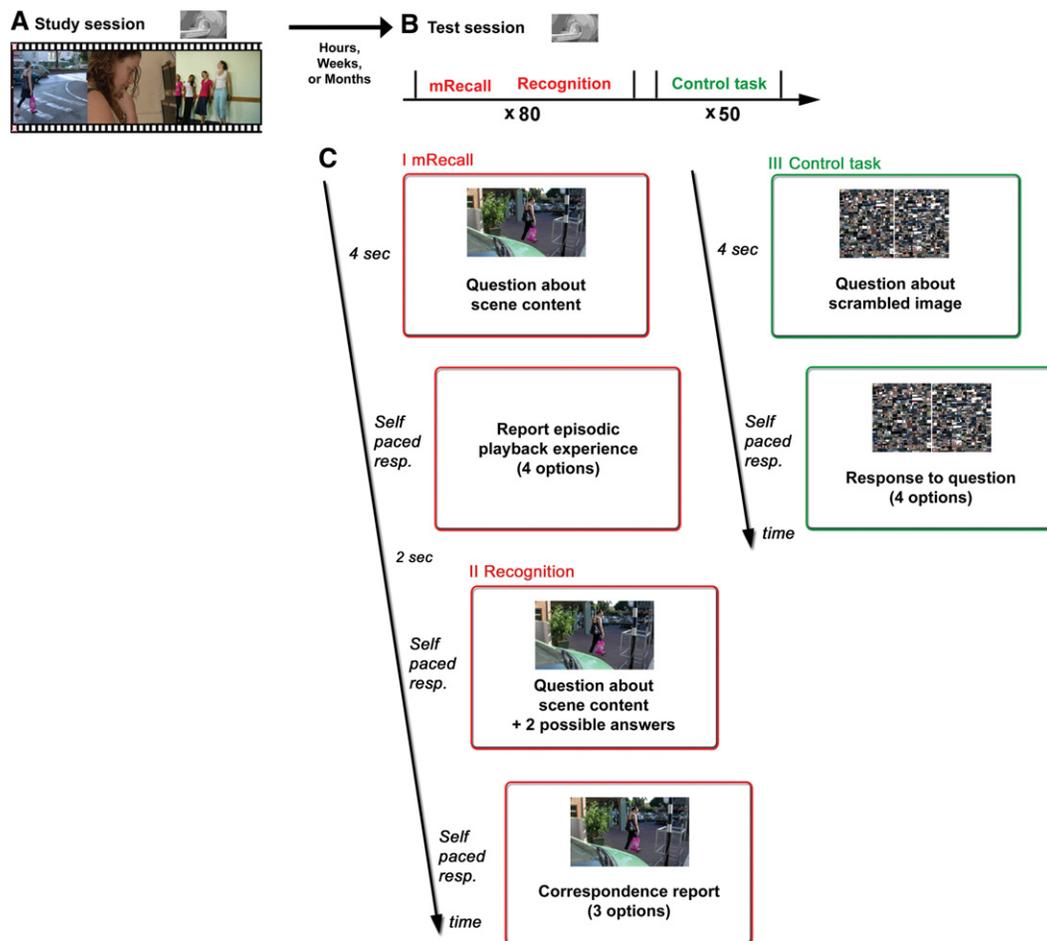
Figure 1. Experimental design. Participants completed two sessions, both conducted during fMRI brain scanning. During the Study session (*A*), participants watched a 27-min documentary movie. During the Test session (*B*), conducted either 3 h ($N = 13$), 3 wk ($N = 15$), or 3 mo ($N = 14$) after Study, participants completed a computerized interactive memory test. Each participant was tested only once. The computerized memory test (*C*) consisted of three different tasks. During the mental recall task (mRecall; see text), participants viewed a still image from the movie, accompanied by a question about scene content. After 4 sec, participants reported whether their attempt to answer this question elicited mental replay of the relevant movie scene. The participants were not, however, requested to provide an overt answer. During the Recognition task, participants were presented with the same image and question, chose one of two possible answers, and made a correspondence report. Participants were asked to choose between three predefined options to report whether their answer during Recognition corresponded with the answer they had covertly reached during the preceding mRecall task (option 1), their answer did not correspond with their covert mRecall answer but was judged to be the correct answer (option 2), or that their answer did not correspond with their mRecall answer and was a guess (option 3). During the Control task, participants passively viewed a scrambled image horizontally bisected into two, and after 4 sec were asked to report which image half was brighter or darker. The mRecall task was always followed by the Recognition task, while the Control task was randomly interspersed between memory retrieval tasks. In all tasks, question screens were shown until an answer was recorded, and no time limit was imposed. All tasks were preceded and followed by presentation of a fixation screen for 2 sec. An additional group of participants (No Movie, $N = 13$) completed the Test session without going through the Study session.

(Mendelsohn et al. 2010). After months, retrieval-related blood-oxygen-level-dependent (BOLD) activity in this network decreased, yet recognition accuracy remained similar to that detected after weeks. Furthermore, activity level in the retrieval network came to be positively correlated with accuracy only after months in mRecall, and for matched items but not for nonmatched items in Recognition.

We also find that the engagement of the hippocampus in retrieval is task-dependent, remaining similarly active at all time points during mRecall but decreasing over time in Recognition. Thus, over a period of months, the engram is transformed into a form that is capable of inducing episodic playback and supporting correct identification of narrative elements from an experienced event with only minimal network activation. The possibility should not be excluded that this potential energetic benefit might have served as a selective pressure in the evolution of the memory transformation process that is referred to as systems consolidation.

Results

Our movie memory paradigm included two sessions, Study and Test (Fig. 1). During Study, the participants viewed a 27-min in-house documentary movie detailing the daily life of a young Israeli woman. Participants were not told they would be tested on movie details. During Test, conducted once per group, hours, weeks, or months after Study, participants completed a custom-made computerized memory questionnaire that assessed memory performance. Both sessions were completed while undergoing fMRI brain scanning. During Test, the computerized memory questionnaire included a mental recall phase (mRecall) wherein participants reported if they experienced movie reenactment during memory retrieval, a content Recognition phase that tested accuracy of memory for narrative elements and details from the movie, and a correspondence report that assessed whether Recognition responses matched the answers reached covertly during mRecall. An additional group completed only the Test scan (No Movie group), to provide a reliable measurement of performance when using only information presented in the memory questionnaire.

Playback during mRecall phase

A distinctive feature of our fMRI study was the controlled probing for mental reenactment of experienced events (termed "playback"). This was possible because retrieval of details from our stimulus material, a real-life-like movie, spontaneously evoked mental reexperiencing of witnessed events. Note that playback of movie content was not a test requirement but arose when subjects retrieved movie information in order to answer test questions. For each item, subjects reported whether playback was experienced (PB or noPB) and whether they felt they knew the answer to the presented question (Ans or noAns), by choosing one of four possible options (termed PB/Ans; PB/noAns; noPB/Ans; noPB/noAns).

The proportion of items reported as evoking mental replay during the mRecall phase (PB items, combining PB/Ans and PB/noAns) remained stable at >70% across Hours (mean proportions from the entire data set ± SEM; 76.3% ± 3.7%), Weeks (72.7% ± 6%), and Months (73.2% ± 7%) (Fig. 2A). However, when dividing PB answers by self-rated answer knowledge, we found that the proportion of PB/Ans items significantly decreased with time (Hours, 52.2% ± 3.8%; Weeks, 34.8% ± 3.9%; Months, 28.7% ± 4.5%; $F_{(2,39)} = 7.13$, $P < 0.005$), while the proportion of PB/noAns items significantly increased over time (Hours, 24.1% ± 2.5%; Weeks, 37.8% ± 3.9%; Months, 44.5% ± 6.2%; $F_{(2,39)} = 3.47$, $P < 0.05$) (Fig. 2B). Similar analysis of items that did not elicit playback indicated no significant differences between groups (noPB/Ans:

Hours, 13% ± 2.6%; Weeks, 9.4% ± 2.8%; Months, 6.3% ± 3.9%; $F_{(2,39)} = 2.7$, NS; noPB/noAns: Hours, 10.6% ± 2.2%; Weeks, 17.8% ± 4.2%; Months, 20.4% ± 4.2%; $F_{(2,39)} = 1.5$, NS).

We next examined the proportion of correct PB items, based on the correctness of the responses made during the following Recognition phase. Again, we found that the proportion of correct PB/Ans items significantly declined with time (Hours, 47.3% ± 3.8%; Weeks, 28.6% ± 3.2%; Months, 23.5% ± 3.6%; $F_{(2,39)} = 9.56$, $P < 0.0005$), while the proportion of correct PB/noAns significantly increased over time (Hours, 16.5% ± 1.7%; Weeks, 25.8% ± 2.6%; Months, 29.8% ± 3.7%; $F_{(2,39)} = 3.45$, $P < 0.05$) (Fig. 2B). For analysis of PB items by correspondence, see the Supplemental Results.
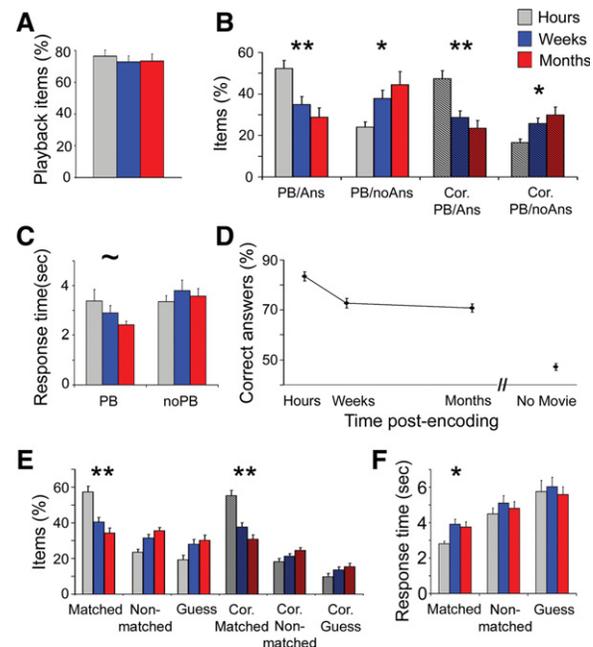
Figure 2. Memory performance over time. (*A*) Proportion of items eliciting self-reported mental replay of movie content (playback items) during the mRecall task. No significant difference was detected between groups. Here and below, error bars represent the standard error of the mean (SEM). (*B*) Proportion of playback items according to self-reported feeling of knowing the answer (PB/Ans) or lack of such knowledge (PB/noAns). Shaded bars represent the proportion of correct items that were determined based on answers chosen during the Recognition test. Whereas total PB was similar across groups, the Hours group had more PB/Ans than Weeks and Months and fewer PB/noAns responses. This was true for both total and correct PB answers. (*C*) Response time during mRecall task for PB/Ans and PB/noAns responses. The Hours group took longer to report playback without knowing the answer. (*D*) Forgetting curve. Correct answers were determined based on responses during the Recognition task. Performance after hours was higher than weeks and months, which did not differ significantly from one another. (*E*) Proportion of Recognition answers divided according to their correspondence with the covert answers reached during mRecall (matched, nonmatched, and guess) (see Fig. 1). Shaded bars represent the proportion of correct items according to the same division. The Hours group made more matched responses, but fewer nonmatched and guess responses than Weeks and Months. Weeks and Months did not differ from one another in any of the categories. (*F*) Response time of Recognition answers divided according to matched, nonmatched, and guess categories. Recognition responses took longer for nonmatched than matched items, but group differences were significant only for matched items, where the Hours group was faster than Weeks and Months in giving a response. Asterisks mark significant differences between groups in one-way analysis of variance; (*) $P < 0.05$; (**) $P < 0.005$; (~) $P < 0.08$.

Finally, analysis of response time (RT) for mRecall trials revealed a trend for reduction in RT for PB items over time (mean RT ± SEM; Hours, 3377 msec ± 464 msec; Weeks, 2898 msec ± 290 msec; Months, 2412 msec ± 150 msec; $F_{(2,39)} = 2.84$, $P < 0.07$) with no change in noPB items RT. When subdividing PB items, we found no change in time to respond PB/Ans but reduction in RT for PB/noAns items was close to significance ($F_{(2,38)} = 3.12$, $P < 0.06$), revealing that the Hours group took longer to report playback without knowing the answer (Fig. 2C).

We therefore find that a constant proportion of items evoked an experience of mental replay during retrieval over time, but we find clues that suggest that the nature of the replay experience changed as time went by.

Memory performance during Recognition phase

Memory for movie details and narrative was above chance level during all time points, yet declined over time ($F_{(2,39)} = 14.01$, $P < 0.0001$), revealing better memory after Hours than Weeks (Tukey post hoc, here and below, $P < 0.0005$) and Months ($P < 0.0005$). The participants in the No Movie control group were at chance level (mean ± SEM, 47.2% ± 1.3%, 95% confidence interval 37.5%–56.9%) when attempting to answer the memory questionnaire, as expected (Fig. 2D).

Correspondence between mRecall and Recognition phase answers changed over time, because more matched items and fewer nonmatched and guess items were reported after Hours than after Weeks and Months (Fig. 2E). Statistical analysis of the proportion of matched trials out of total trials was performed using one-way ANOVA and revealed a significant decline over time ($F_{(2,39)} = 16.36$, $P < 0.00005$; post hoc Tukey HSD hours > weeks, $P < 0.0001$; hours > months, $P < 0.0005$). A similar pattern of significant differences was found when analyzing proportion of correct matched trials out of total trials ($F_{(2,39)} = 21.22$, $P < 0.00001$, post hoc Tukey HSD hours > weeks, $P < 0.0005$; hours > months, $P < 0.0005$).

Finally, RT analysis of Recognition phase data using two-way repeated measures ANOVA revealed a significant main effect of category ($F_{(2,78)} = 88.7$, $P < 0.00005$). Post hoc tests confirmed that across groups, RT for matched items was shorter than RT for nonmatched items ($P < 0.0005$) and RT for guess items ($P < 0.0005$). Group differences assessed by separate one-way ANOVAs for each category were significant only for matched trials, which the Hours group answered significantly faster than Weeks and Months (Hours, 2807 ± 133 msec; Weeks, 3906 ± 283 msec; Months, 3756 ± 277 msec; $F_{(2,39)} = 5.56$, $P < 0.01$). All experimental groups responded significantly faster during correct compared with incorrect trials, in both mRecall and Recognition phases.

Our aim was to isolate brain correlates of retrieval of bona fide memory, and therefore we included in analysis only items that fulfilled two criteria: (1) items that were answered correctly, and (2) items that subjects reported remembering with confidence (i.e., excluding guesses). When analyzing memory performance under these conditions, we found that the proportion of such correct confident answers declined with time (Hours, 80.7% ± 2.7%; Weeks, 72.0% ± 2.8%; Months, 69.9% ± 3.2%; $F_{(2,39)} = 13.4$, $P < 0.0001$), and was higher at Hours compared with Weeks ($P < 0.005$) and Hours compared with Months ($P < 0.005$). No differences were found between performance of Weeks and Months groups in a multitude of behavioral measurements, including proportions of mRecall answers in each subcategory (Fig. 2B), proportion of correct responses and correct confident responses, correspondence category proportion of all items and of correct items (Fig. 2E), and in RT for all trial types.

We conclude that memory performance significantly declined between Hours and Months, but we did not detect a decline in memory accuracy between Weeks and Months post-viewing. We further conclude, from analysis of the mRecall phase data described above, that retrieval of episodic memories changed over time, such that months-old memories were less rich in contextual details and associations.

BOLD response amplitudes in the 'retrieval network'

We defined the set of brain areas supporting memory retrieval using the Inclusive model containing all three experimental groups (see Materials and Methods), contrasting activity during the mRecall phase to activity during the perceptual Control task. Functional clusters that survived a threshold of $P < 0.001$ corrected for multiple comparisons using cluster-size thresholding (see Materials and Methods) were included in further analysis reported below. The system thus delineated is referred to throughout as the "retrieval network" (see Fig. 3A). The retrieval network consisted of hippocampus, parahippocampal gyrus (PHG), middle and superior frontal gyri, dorsomedial (DMPFC), ventromedial (VMPFC) and ventrolateral prefrontal cortex (VLPFC), superior temporal sulcus (STS) and temporal pole (TP), temporal parietal junction (TPJ), and retrosplenial and posterior cingulate cortex (RSC-PCC) as well as fusiform gyrus and lateral occipital cortex (LO) (see Table 1 for details). These regions have previously been implicated in studies of autobiographical memory (Svoboda et al. 2006) and in a study of real-life-like memory retrieval, which found differential coactivation in a subset of these regions as a function of correct performance (Mendelsohn et al. 2010).

Multiple retrieval network ROIs showed decrease in BOLD signal over time during correct responses (i.e., using the Performance model, see Materials and Methods). These differences stemmed from a decrease in amplitude after Months relative to Hours and Weeks (Figs. 4, 5; Supplemental Figs. S3–S5). Supporting this pattern, we found significant interaction effects between Group and BOLD measurements, implying that the pattern of BOLD responses differed among Hours, Weeks, and Months. During correct mRecall trials, significant interactions were found in bilateral hippocampus and posterior PHG, left VLPFC, bilateral posterior STS, bilateral RSC-PCC, and left LO. Very similar results were found when including only matched correct mRecall trials in analysis (Supplemental Fig. S6) and when comparing matched correct mRecall only between Weeks and Months groups (Supplemental Fig. S7; Supplemental Table S4). During correct Recognition trials, we found significant interactions in bilateral hippocampus, PHG, left TP, STS, right RSC-PCC, medial parietal cortex, fusiform gyrus, and left LO.

Direct comparisons among groups revealed that the BOLD signal in several regions was significantly lower in the Months group relative to the Hours and Weeks groups. The regions showing this effect during correct mRecall trials were bilateral RSC-PCC, MTL (bilateral posterior PHG), and temporal cortex (bilateral posterior STS and left TP). During correct Recognition trials, a decrease in BOLD after Months was found in MTL (bilateral hippocampus and anterior PHG, right posterior PHG) and in temporal regions (right anterior and left posterior STS).

No regions in any analysis we conducted revealed higher BOLD signal for Months relative to Hours and Weeks. Similar to analysis of correct mRecall using the Performance model, ROI analysis using a different model for assessing playback items revealed multiple regions whose activity during playback mRecall events was also significantly lower for Months relative to Hours and Weeks (in PHG, L TP, STS, RSC, L TPJ, and L VLPFC but not in hippocampal ROIs) (see Supplemental Fig. S3). To rule out the possibility that the decline after Months can be explained by a decrease in confidence over time, we used an additional
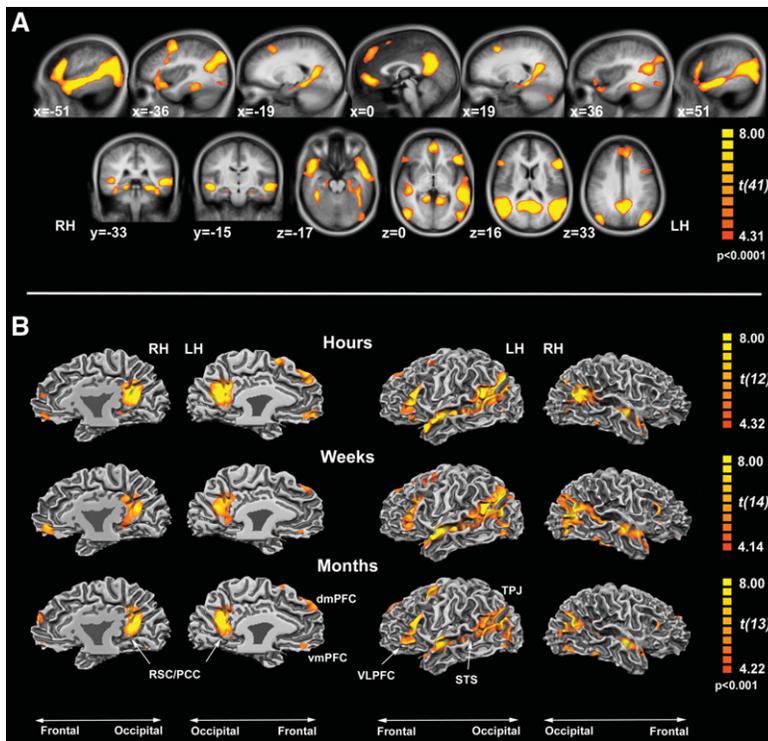
Figure 3. Brain system supporting episodic memory retrieval during mental recall. (*A*) Statistical map displayed on sagittal (*top*), coronal (*bottom left*), and axial (*bottom right*) sections of a group-averaged anatomical brain, depicting areas whose activity during mRecall was higher than activity during the Control task. When items were reported as guesses during the Correspondence report (see Fig. 1), the preceding mRecall items were excluded from analysis. The map depicts a contrast of activity during TR4 of mRecall and Control tasks, computed using a linear regression model that included participants from all three experimental groups (Hours, Weeks, Months). (For a similar map depicting a contrast of all TRs, see Supplemental Fig. S1. For details, see Materials and Methods.) The map depicts voxels exceeding a threshold of $P < 0.0001$, corrected at $\alpha = 0.05$ for cluster-level false-positive rate, minimal cluster size <1269 mm^3 (see Table 1 for list and coordinates of regions). (*B*) Statistical maps depict areas whose activity during the mRecall task was higher than activity during the Control task for each group separately, using the same criteria as in *A*. Medial (*left*) and lateral (*right*) views are displayed for each hemisphere on a single participant's brain (RH, right hemisphere; LH, left hemisphere). Contrast maps were thresholded at $P < 0.001$, uncorrected. Comparable activation patterns across the three experimental groups are seen in temporal, occipital, and prefrontal cortical regions. Single-group models detected comparable activity in parahippocampal gyrus (not shown) but not in hippocampus, where activity was detected only in the multigroup analysis. (RSC) Retrosplenial cortex; (PCC) posterior cingulate cortex; (dmPFC) dorsomedial prefrontal cortex; (vmPFC) ventromedial prefrontal cortex; (VLPFC) ventrolateral prefrontal cortex; (TPJ) temporoparietal junction; (STS) superior temporal sulcus.

model to separately analyze matched and nonmatched Recognition trials (grouping correct and incorrect trials) (see the Supplemental Material for details). We find a decreased BOLD response after Months in both matched and nonmatched Recognition trials in multiple regions (during both trials types, we found a decrease in PHG, STS, and R VLPFC; only during matched trials we found a decrease in hippocampus, fusiform, and L TPJ; only during nonmatched trials, we found a decrease in RSC-PCC and L TP) (Supplemental Figs. S4, S5).

Note that very similar retrieval network maps (mRecall > Control contrast) were generated from single-group GLM models (excluding hippocampus) (Fig. 3B). We did not find additional retrieval-related regions in any of the separate group analyses, and the opposite contrast (Control > mRecall) also yielded similar single-group maps. We thus rule out the possibility that different brain networks support retrieval at different time points. Moreover, the retrieval network ROIs recovered from single-group GLM models were consistent over time, and the response ampli-

tude analysis using group-specific ROIs gave results similar to those reported above (Supplemental Fig. S2).

To summarize, in MTL, temporal cortex, and medial posterior cortex, we found during correct retrieval a decrease in signal intensity in the Months group relative to the Hours and Weeks groups. This effect was not seen in prefrontal ROIs, where brain activity was similar among groups during both correct mRecall and Recognition trials. An exception to this pattern was found in VLPFC, where reduced activity after Months was significant on the left side during correct mRecall trials and on the right side during Recognition trials.

Brain–behavior correlations

We sought to take advantage of the variation in memory performance within our subject pool in order to further explore changes in retrieval-related brain activity over a time course during which systems consolidation is postulated to proceed. For each group, we computed the correlation between ROI brain activity and correct memory performance (see Materials and Methods). We found positive brain–behavior correlations in many regions of the retrieval network after Months, but not after Weeks and Hours. Positive correlations between brain activity and correct performance at Months were found in left hippocampus and right posterior PHG, left lateralized STS and TPJ, occipito-termporal regions, and several PFC regions (see Supplemental Tables S1, S2). These correlations were especially strong during matched Recognition trials (see Fig. 6), suggesting that the greater the cue-related brain activity, the higher was the correct confident memory performance at Months.

Nonparametric statistical tests, aimed at assessing the statistical significance of reported correlations (see Materials and Methods), revealed that only in the Months group were positive brain–behavior correlations statistically significant. No positive correlations were significant when using other time points of brain activity measures, when correlating memory performance with brain activity during the Control task, or when using the proportion of items eliciting mental replay as a measure of performance (Supplemental Table S3). Thus, although BOLD responses showed an overall decrease in the Months group compared with Hours and Weeks, only after Months did memory performance correlate positively with BOLD activation strength.

To summarize our findings, we report (1) a superior memory performance after hours, reduced to a similar performance weeks and months after study; (2) a reduction in BOLD activity amplitude during correct retrieval after Months relative to the Hours and Weeks groups; (3) positive brain–behavior correlations in many ROIs observed only for the Months group, indicating that the greater retrieval-related brain activity, the better is performance in the memory test.

Table 1. Regions encompassing the retrieval network

Region	X	Y	Z	mm³	t-value	P-value
R hippocampus	21	−20	−14	611	5.38	0.000003
L hippocampus	−21	−20	−14	651	6.63	$<10^{-6}$
R anterior PHG	21	−27	−11	324	6.25	$<10^{-6}$
L anterior PHG	−21	−22	−14	664	7.75	$<10^{-6}$
R posterior PHG	12	−43	2	3730	10.75	$<10^{-6}$
L posterior PHG	−12	−49	2	5149	10.55	$<10^{-6}$
R middle frontal gyrus (BA6)	21	17	49	2423	8.65	$<10^{-6}$
L middle frontal gyrus (BA6)	−36	2	52	11,169	9.00	$<10^{-6}$
L superior frontal gyrus (BA6)	−6	8	61	5870	7.74	$<10^{-6}$
DMPFC (BA 8-9)	−9	44	46	16,023	11.08	$<10^{-6}$
VMPFC (BA 10-11)	0	50	−2	7976	9.56	$<10^{-6}$
R VLPFC (BA 45-47)	48	26	10	6950	8.48	$<10^{-6}$
L VLPFC (BA 45-47)	−48	23	10	15,989	12.14	$<10^{-6}$
R temporal pole (BA 38)	45	20	−17	11,129	11.51	$<10^{-6}$
L temporal pole (BA 38)	−45	14	−17	13,473	12.67	$<10^{-6}$
R anterior STS (BA 21)	51	−10	−8	7573	17.34	$<10^{-6}$
L anterior STS (BA 21)	−54	−10	−5	10,658	16.38	$<10^{-6}$
R posterior STS (BA 21-22)	42	−53	16	12,067	10.99	$<10^{-6}$
L posterior STS (BA 21-22)	−54	−46	4	15,553	13.26	$<10^{-6}$
R TPJ (BA 39-19)	48	−64	10	20,967	14.81	$<10^{-6}$
L TPJ (BA 39-19)	−42	−70	25	29,886	15.50	$<10^{-6}$
R RSC-PCC (BA 29-30-23)	12	−49	10	10,230	13.88	$<10^{-6}$
L RSC-PCC (BA 29-30-23)	−9	−52	10	10,911	16.48	$<10^{-6}$
R PCC (BA 23-31)	1	−58	23	6858	13.05	$<10^{-6}$
L PCC (BA 23-31)	−9	−52	31	8487	14.71	$<10^{-6}$
R middle occipital gyrus (BA19)	48	−73	1	6491	12.63	$<10^{-6}$
L middle occipital gyrus (BA 19)	−45	−76	−8	8261	10.10	$<10^{-6}$
R fusiform gyrus (BA 19)	36	−37	−14	3506	9.52	$<10^{-6}$
L fusiform gyrus (BA 19)	−27	−34	−11	5247	9.99	$<10^{-6}$
R cerebellar vermis lob. VII	12	−79	−38	3536	6.06	$<10^{-6}$

Coordinates in Talairach space (Talairach and Tournoux 1988) and associated t scores for the peak voxel in ROIs that display differential BOLD activity between the mRecall task and Control task at TR4. The size of the whole ROI is indicated. The ROIs were obtained with a threshold of $t_{(41)} > 3.5$, $P < 0.001$, corrected at $\alpha = 0.001$ for cluster-level false-positive rate (minimal cluster size >1269 mm³). Some clusters were divided into several ROIs based on anatomy, resulting in ROI size smaller than the minimal cluster size. (BA) Brodmann area.

Discussion

We show that the passage of time is accompanied by distinct changes in episodic memory performance and in brain activity correlated with this performance. Overall, recognition memory for targeted details from movie events became less accurate over time. While the number of items reported to induce mental replay was equivalent after hours, weeks, and months, these reports became less indicative of retrieval accuracy over time. When comparing proportions of playback items reported to include (PB/Ans) or not include (PB/noAns) a feeling of knowing the answer, we find more PB/Ans and fewer PB/noAns responses after hours, with an opposite pattern after weeks and months. These findings led us to conclude that the richness and complexity of episodic memories decreases over time. Further support for changes in memory quality is gleaned from analysis of response time (RT): we find that for playback-inducing mRecall trials, RT is slowest after hours and is faster with time, whereas response in Recognition is quickest after hours and takes longer after weeks and months, in particular for matched items (items for which the covert answer reached at mRecall corresponds with the answer endorsed during Recognition test).

By way of construing the above findings, we speculate that replay of movie events includes mental reconstruction of relevant characters, their physical context, interactions, movements, and temporal propagation of events. In our protocol, mRecall and Recognition phases were presented consecutively while targeting the same detail, making it likely that replay-induced representations, retrieved during mRecall, affected processing during Recognition. If for a given cue, memory retrieval after hours produces more details and associations than are needed for responding, extra cognitive effort is required to sift through excessive details to reach a target answer (Brewer 1996), yielding longer mRecall RT.

To continue along this line of reasoning, we consider how an ensuing Recognition test could benefit from the excessive details brought to mind during mRecall, provided that the answer reached during mRecall is confirmed by one of the possible answers presented during Recognition. For such matched items, RT would be shorter than in nonmatched items, which require additional top-down processes of retrieval. Indeed, we find that regardless of memory age, matched items were shorter to respond to than nonmatched items (Fig. 2F). When examined for differences over time, we find that matched items RT is significantly shorter after hours than after weeks and months, but groups do not differ in nonmatched items RT. Therefore, recognition performance benefits from availability of richer representations after hours when mRecall and Recognition answers correspond, but not when new retrieval is needed. In summary, the analysis of performance parameters leads us to conclude that richer episodic representations are retrieved after hours than after weeks and months, supporting superior correct recognition after hours.

Next, we set out to determine how neural correlates of memory retrieval change over time to support and reflect these performance differences.

An extensive and consistent set of brain areas was activated at all time intervals (Table 1). This retrieval network consisted of MTL, medial anterior and posterior cortex, and lateral temporal and parietal cortex, as well as prefrontal and occipital cortical regions, similar to a previously defined network active during autobiographical memory retrieval (AM network) (Svoboda et al. 2006; McDermott et al. 2009). We defined this network based on a difference in brain activity between mRecall and control task at TR4, or 8 sec after event onset. This time point was chosen to capture differences in evoked hemodynamic response, which for randomly presented fMRI trials is expected to peak 8–12 sec after event onset (Friston et al. 1995). We find an identical network when contrasting all 10 TRs following onset of the mRecall task with all 10 TRs following onset of the control task (see Materials and Methods; Supplemental Fig. S1).

Whereas correct memory performance was highest after hours and decreased to a similar level after weeks and months, BOLD response in a large portion of the retrieval network significantly decreased after months compared with both hours and weeks, in which activation was similar. This decrease was especially prominent during the Recognition task in MTL, temporal, and parietal regions, but not in prefrontal cortex (excluding VLPFC). While these signal reductions might be interpreted as correlates of forgetting, or of more uncertainty and guessing, this interpretation is excluded when considering the similarity in memory performance after weeks and months. Furthermore,
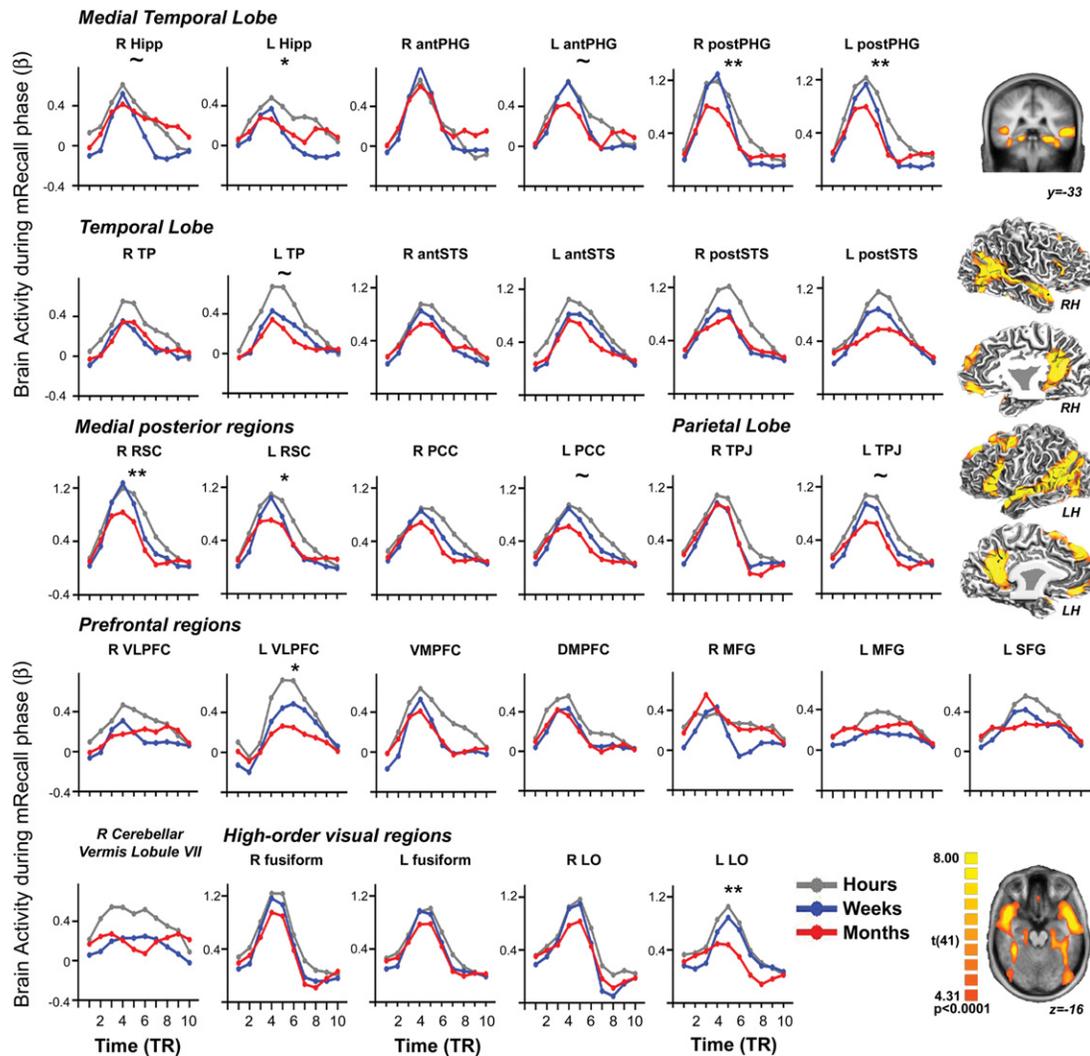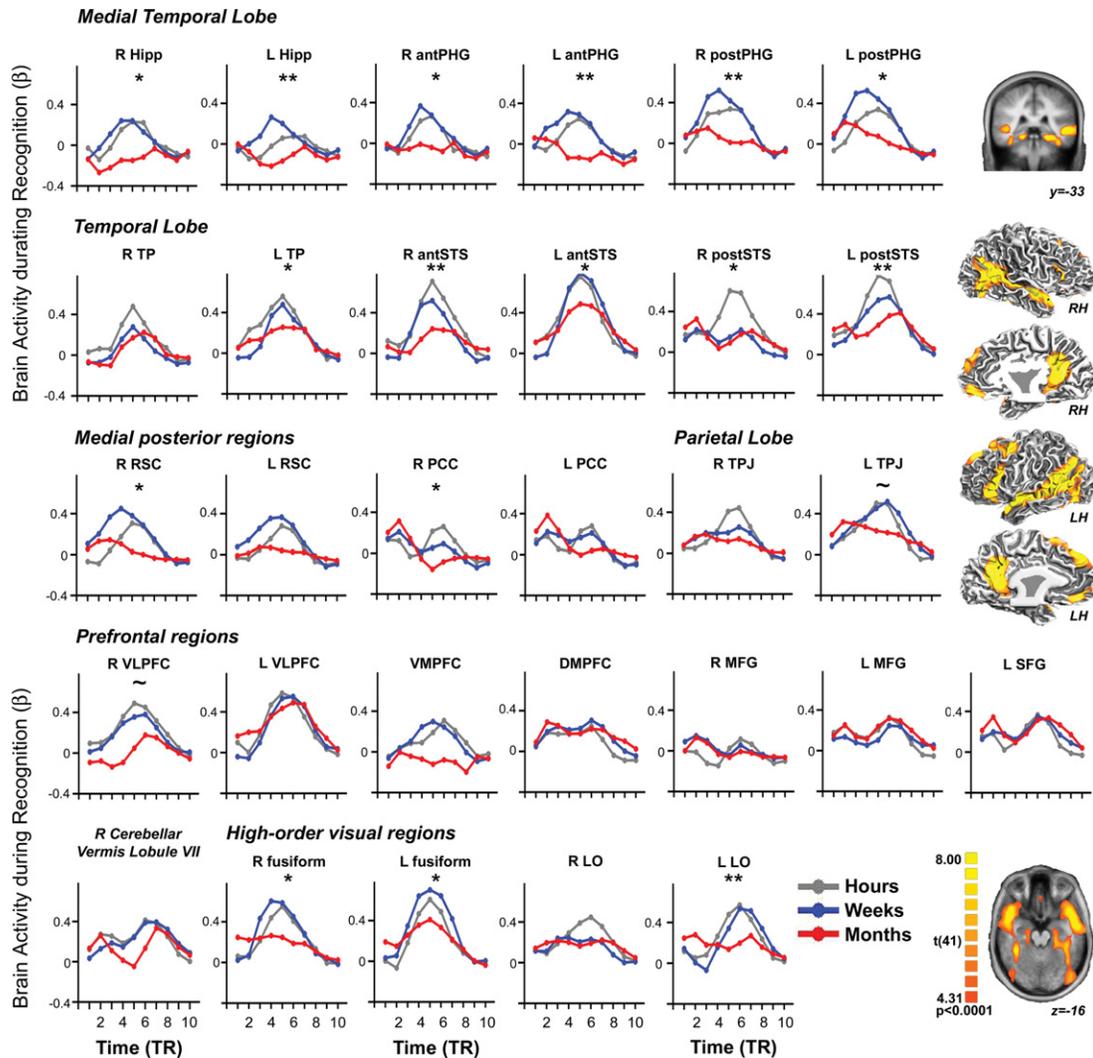
Figure 4. BOLD response in retrieval-related regions during correct mRecall trials decreases over months. ROIs defined from the retrieval network map (Fig. 3A) are displayed on a group-averaged anatomical brain as well as on medial and lateral views of each hemisphere on a single participant's brain; (RH) right hemisphere; (LH) left hemisphere. BOLD response during correct mRecall trials was estimated for an interval of 20 sec (10TR) using a finite impulse response (FIR) model (see Materials and Methods). Each plot compares response profiles of the Hours (gray), Weeks (blue), and Months (red) groups. Significant decrease in retrieval-induced BOLD response is detected in the Months group relative to the Hours and Weeks groups in multiple ROIs. Statistical comparisons of group responses were performed using a two-way analysis of variance (mixed ANOVA model, using study-test interval as group factor × activity estimate over TRs as a repeated measure factor). Significant interactions were found, implying differences between groups in BOLD response over time; (*) $P < 0.05$; (**) $P < 0.005$; (~) $P < 0.08$. (R), right; (L) left; (Hipp) hippocampus; (ant) anterior; (post) posterior; (PHG) parahippocampal gyrus; (TP) temporal pole; (STS) superior temporal sulcus; (RSC) retrosplenial cortex; (PCC) posterior cingulate cortex; (TPJ) temporoparietal junction; (VLPFC) ventrolateral prefrontal cortex; (VMPFC) ventromedial prefrontal cortex; (DMPFC) dorsomedial prefrontal cortex; (MFG) middle frontal gyrus; (SFG) superior frontal gyrus; (LO) lateral occipital cortex; (TR) time to repetition.

we did not retest participants, thus the reduced BOLD activity after months cannot be attributed to experience-dependent effects of test familiarity. One may raise the possibility that as study-test intervals grow longer, the more opportunities there are for spontaneous rehearsal of movie content, possibly promoting retrieval and, according to multiple trace theory (MTT), more widespread neural memory representations. It is difficult to control for amount of spontaneous out-of-lab rehearsal, and this problem is inherent to all real-life memory testing (Cohen 1996). However, during debriefing following memory testing, all long-term participants were asked in a written questionnaire about the circumstances during which they rehearsed or spontaneously recalled the movie. Most participants reported thinking about the

movie immediately after the encoding session and again after scheduling the next meeting. Finally, differences in original encoding efficiency cannot account for signal differences because participants were randomly assigned to groups, resulting in comparable age and education level (see the Supplemental Material for more details).

We suggest that reduced neural activity after months reflects an outcome of ongoing consolidation processes that change memory quality and organization while preserving the ability to retrieve accurate information. It is tempting to speculate that two aspects of bona fide episodic recollection, mental time-travel and correct judgment of the veracity of past events, evolved under different selective pressures. Whereas mental time-travel might

Figure 5. BOLD response in retrieval-related regions during correct Recognition trials decreases over months. ROIs defined from the retrieval network map (Fig. 3A) are displayed on a group average anatomical brain as well as on medial and lateral views of each hemisphere on a single participant's brain; (RH) right hemisphere; (LH) left hemisphere. BOLD response during correct Recognition trials was estimated for an interval of 20 sec (10TR) using a finite impulse response (FIR) model (see Materials and Methods). Each plot compares response profiles of the Hours (gray), Weeks (blue), and Months (red) groups. Significant decrease in retrieval-induced BOLD response is detected in the Months group relative to the Hours and Weeks groups in multiple ROIs. For statistical analysis description and notation, as well as abbreviations, see the legend to Figure 4.

subserve imagination (Tulving 1995; Dudai and Carruthers 2005; Addis et al. 2007), correct recognition seems of a more elementary survival value, e.g., distinguishing friend from foe. In balancing between usable information and its energetic cost (potentially tapped by the BOLD signal) (Bullmore and Sporns 2012), the system may have evolved to keep the critical ability to recognize correctly for longer times at a lesser expense. Supporting this view, previous results suggest that consolidation processes engaged during sleep change brain representation of an acquired motor skill by reducing cortical contributions, i.e., by decreasing BOLD signals measured during subsequent retrieval (Fischer et al. 2005).

Most human studies of systems consolidation focus on the time-dependent role of the hippocampus in declarative retrieval; standard consolidation theory (SCT) advocates a temporary role of the hippocampus (McClelland et al. 1995; Squire et al. 2004), while MTT maintains that hippocampus is required throughout (Moscovitch and Nadel 1998), provided that retrieval taxes episodic qualities such as vividness (Addis et al. 2004; Gilboa et al.

2004; Piolino et al. 2009). Evidence was reported supporting both MTT (Ryan et al. 2001; Gilboa et al. 2004; Viard et al. 2007, 2010) and SCT (Haist et al. 2001; Takashima et al. 2006, 2009; Smith and Squire 2009). Studies supporting MTT typically use cues requiring elaborative reenactment, while SCT is mostly supported by studies that probe memories of leanly associative and semantic nature (see Squire and Bayley 2007; Dudai 2012 for further discussion of this point).

We find different patterns of hippocampal activity depending on retrieval task. Whereas during correct mRecall trials, hippocampal activation is similar at all intervals, supporting MTT, during correct Recognition trials, activation drops after months, seeming to support SCT (Figs. 4 and 5, respectively). However, brain–behavior correlations reveal that Recognition performance depends on hippocampal activity after weeks and months (Supplemental Table S2), suggesting that hippocampal activity is important for both tasks at all examined time periods. Note that as consolidation processes might be longer or ongoing indefinitely
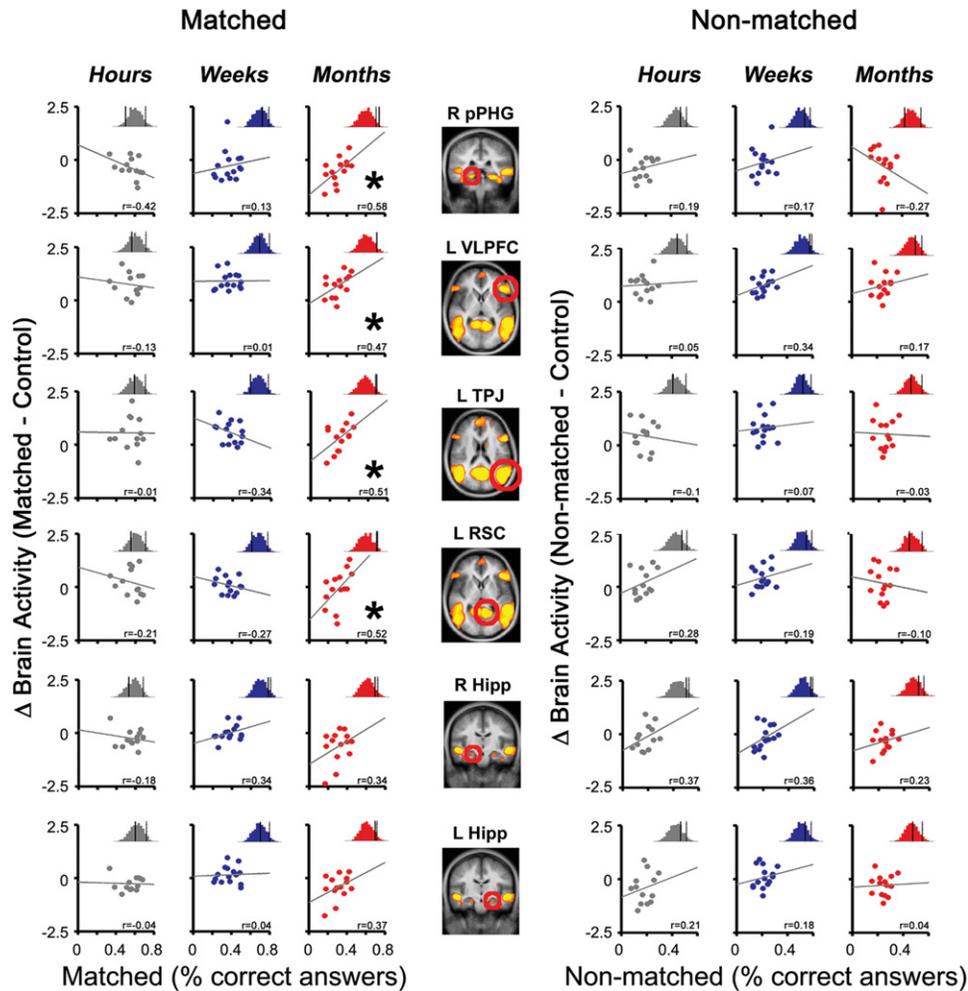
Figure 6. After months, brain activity and memory accuracy are positively correlated during matched but not nonmatched Recognition responses. ROIs defined from the retrieval network map (Fig. 3A) are displayed on group averaged anatomical brain (*middle* column). Each row of plots represents data obtained from the corresponding ROI. BOLD activity differences between memory retrieval tasks and the control task were computed for each subject at TR4 (ordinate, including only correct trials and excluding guesses) and correlated with correct memory performance (abscissa). Data were correlated separately for matched (*left* columns) and nonmatched (*right* columns) Recognition trials and for each group (gray, Hours; blue, Weeks; red, Months). To determine the statistical significance of the correlations, a nonparametric test was performed by randomly assigning memory performance measures for each subject (the histogram inset in each scatterplot; the abscissa denotes correlation values from −1 to 1, and the ordinate denotes the number of occurrences from 0 to 150. The vertical solid line depicts correlation value using veridical data, the dashed vertical line depicts the 95th percentile, and the asterisk denotes significance of the nonparametric test). A significant positive brain–behavior correlation is detected only during matched trials and only in the Months group: The higher the difference in brain activity between memory retrieval and control task, the greater is memory performance. (pPHG) Posterior parahippocampal gyrus; (VLPFC) ventrolateral prefrontal cortex; (TPJ) temporoparietal junction; (RSC) retrosplenial cortex; (Hipp) hippocampus.

(Dudai 2012), we cannot predict whether this hippocampal dependency will remain at longer time points, e.g., years postencoding. Taken together, our finding of modular engagement of hippocampus according to task demands is in line with the trace transformation theory proposed by Winocur et al. (Winocur et al. 2010; Winocur and Moscovitch 2011). This hypothesis posits that an item in episodic memory remains dependent on the hippocampus as long as this item is available in its richly detailed form, that with time the hippocampus supports the development in neocortex of a schematic gist-only version of the memory, and that the memory version that dominates depends on the circumstances of retrieval (Wiltgen et al. 2010; Winocur et al. 2010). Our data call for further refinement of the trace transformation theory, because we find that detailed hippocampal-dependent declarative memories are transformed

over a period of months to a form that is less detailed yet maintains the capacity to elicit recollection by mental playback and to support correct performance.

Additionally, we find greater signal decline over time during Recognition than mRecall in STS, RSC, and fusiform gyrus. These cortical regions, while holding diverse cognitive correlates, are all implicated in functions related to retrieval of real-life memories such as scene reconstruction, interpretation of social communication, and relating objects to their context in complex scenes (Bar 2004; Hassabis and Maguire 2007; Hasson et al. 2008; Henderson et al. 2008; Summerfield et al. 2009). Because these regions (especially RSC) (see Vann et al. 2009) are reported to be involved in reenactment and imagery of scenarios, functions that are more likely to be engaged during the mRecall phase than during Recognition, it is likely that activity during mRecall may be

high at all time points. The decline in activity after months during Recognition can thus be construed as a decrease over time in richness of evoked memory details.

Human prefrontal regions play a pivotal role in declarative memory retrieval (Moscovitch and Winocur 2002; Gilboa 2004; Svoboda et al. 2006), although an increase in their activity as a function of time since the event is not typically detected in studies of AM (Maguire et al. 2001; Gilboa et al. 2004; but see Steinvorth et al. 2006) unlike animal studies in which increased PFC activity over time is reported (Bontempi et al. 1999; Frankland et al. 2004; Rudy et al. 2005). We find that left VLPFC activity decreases in a graded fashion over time during correct mRecall as well as playback mRecall trials (Fig. 4; Supplemental Fig. S3), as previously reported (Maguire et al. 2001). Activity in this region is associated with strategic retrieval of information from posterior cortical association regions (Fletcher and Henson 2001; Petrides 2002) and was shown to be modulated by the amount of retrieved alternatives from which targeted memories are to be selected (Gold and Buckner 2002; Badre et al. 2005). We therefore suggest that decline in activity over time in left VLPFC during mRecall might reflect excessive monitoring and selection processes engaged after hours, when representations are more detail-laden, but not after longer time intervals. In addition, we find that right VLPFC activity significantly declines after months during confident Recognition trials, either matched or nonmatched (Supplemental Figs. S4, S5), and a trend is seen during correct Recognition trials (Fig. 5).

Other PFC regions (ventromedial, dorsomedial, and bilateral premotor cortex) did not show time-dependent differences in activity, suggesting that the diverse processes they support are required for memory retrieval of complex naturalistic socially meaningful stimuli, regardless of memory age and strength (Gusnard et al. 2001; Gallagher and Frith 2003; Mar 2004; Mitchell et al. 2004; Summerfield et al. 2009).

Notably, in addition to changes in BOLD amplitude, we find time-dependent shifts in brain–behavior correlations: Brain and behavior measure-ments revealed a trend toward negative correlations at hours (with significant negative correlations in L VLPFC during correct mRecall, in L postPHG and L RSC during correct Recognition), were uncorrelated at weeks, and were significantly positively correlated at months. Put differently, when tested after hours, the higher the activations in multiple regions, the worse memory performance was, whereas after months, the higher the activations, the better memory performance was. Positive brain–behavior correlations were especially prominent in left temporal and frontal regions, as well as bilateral occipital and medial posterior cortical areas during both mRecall and matched item Recognition phases. Our interpretation for this set of findings is that after hours, the extent of brain activity is not predictive or indicative of performance since the activation during this time point might reflect retrieval of both relevant and nonrelevant information. Thus, although brain activity is decreased at months (as described above), the extent of activity is highly related to memory performance. This finding, repeating in mRecall and Recognition and relating to matched but not to nonmatched Recognition items, strengthens our claim that the episodic engram changes over time to a form that contains fewer details but is nonetheless efficient for correct recognition performance. We suggest that at shorter time intervals after encoding many more items are retrieved than are needed, which may recruit excessive brain circuits not necessarily related to successful retrieval that might even interfere with performance.

All in all, our findings indicate that over time, brain activity supporting episodic memory performance changes in concordance with transformation of highly associative representations into memories that are less context-ridden yet support

mental replay and correct recognition performance. The pattern of hippocampal participation as a function of time and task that we observe in our naturalistic episodic memory paradigm can be construed to be in line with the recently proposed trace transformation theory of systems consolidation (Winocur et al. 2010). We further suggest that between weeks and months after encoding of new experiences, while recollection of memory becomes less contextualized and more specific to the targeted information, the correspondence between brain activity and memory performance is augmented, increasing retrieval specificity. Indeed, the possibility could be raised that the transformation of the engram into a distilled form that allows retrieval of the crux of past events by permitting recognition of what did or did not happen using parsimonious brain activation might have been the outcome of the adaptive selective pressure that had led to the evolution of systems consolidation.

Materials and Methods

Participants

A total of 62 participants took part in this fMRI study (34 females, mean age 26.3 yr, range 18–38). The number of participants per study-test interval was 3 h, $n = 13$; 3 wk, $n = 15$; 3 mo $n = 14$. An additional group completed only the Test scan (No Movie group, $n = 13$). The experimental protocol was approved by the Institutional Review Board of the Sourasky Medical Center, Tel-Aviv. All participants had normal or corrected-to-normal vision, provided written informed consent, and were remunerated for their participation (further information provided in Supplemental Material).

Study material and memory test

The stimulus used at Study was a 27-min documentary movie depicting the daily life of a young actress (E.S.), produced in-house for exclusive use in memory research in our laboratory (see Mendelsohn et al. 2009, 2010). The movie depicted E.S. going about her daily routines as well as carrying out more salient activities such as rehearsing for a play, teaching a drama class in high school, and rollerblading with friends. Participants were not told they would be tested on the movie details.

The Test session, in which memory performance was assessed, consisted of a computerized interactive questionnaire (Presentation software, Neurobehavioral Systems), which included a mental recall (mRecall) phase, a Recognition phase, and a correspondence report (Fig. 1). Prior to testing, participants received written instructions and completed a short demo to ensure understanding of task demands (see Supplemental Material). During the mRecall phase, a still image capturing a frame from the movie was presented for 4 sec, along with a question probing memory for a particular event that occurred in the relevant scene (e.g., "Where is E.S. heading from the rehearsal room?" "What is the first exercise in the drama class?"). Subjects were required to view the image and question passively for 4 sec, after which they reported whether and to what extent their attempt to answer the question elicited mental replay of the relevant scene (referred to as playback). Subjects did not give an overt reply because self-generated response via speaking or writing requires movement during scanning. A playback report was performed by choosing one of the following four options:

1. "I experienced playback & I know the answer."
2. "I experienced playback & I don't know the answer."
3. "No playback & I know the answer."
4. "No playback & I don't know the answer."

This phase was self-paced. Subsequently, the Recognition task was presented, using the same image and question as in the mRecall phase, but this time with two answers to the question, one correct and one false. The participants' task was to choose the answer they

thought was most accurate. Following their response, participants reported correspondence between Recognition and mRecall phase answers, using a prespecified rating system that was explained during prescan instructions (see Supplemental Material for reproduction of instructions). While this was not an explicit rating of answer confidence, the correspondence report assessed response confidence indirectly, considering that confidence during the Recognition phase could derive from the relation between the options presented and the covert answer reached during the mRecall phase. Participants indicated one option if the answer chosen during Recognition was also the answer they had reached covertly during mRecall ("matched"); a second option indicated that the answer they chose as correct was dissimilar from the answer they had reached during mRecall ("nonmatched"), and a third possible option ("guess") if they had no preference between the suggested optional answers.

In total, 78 questions were composed to test memory for narrative elements of the documentary movie, at an approximate sampling rate of one question per 20 sec of film. Questions about the narrative pertained to sequence of events (e.g., "Where did E.S. go after teaching her class in high school?") and to detailed aspects of a particular scene (e.g., "What is the composer's comment on the song they are rehearsing?"). Answer screens were presented until the response was recorded.

An additional Control task was used, appearing randomly between memory retrieval tasks (average of 25 trials per scan). In this task, participants viewed a colored, horizontally bisected, scrambled image for 4 sec and were instructed to determine which half was brighter by pressing a designated button on the response box. Four possible options were given:

1. "Right side, low confidence."
2. "Right side, high confidence."
3. "Left side, low confidence."
4. "Left side, high confidence."

The scrambled images were created by randomly scrambling movie stills.

FMRI experimental procedures
During Study, the documentary movie was projected to the subjects via an LCD projector to a screen behind the MRI scanner and was viewed via a mirror mounted on the head coil. Audio was fed into the standard Siemens bed system and was delivered via MRI-compatible head phones (MR Confon GmbH). Participants were not informed of a pending memory questionnaire and were not asked to memorize movie content explicitly.

During Test, carried out 3 h (dubbed Hours), 3 wk (Weeks), or 3 months (Months) after Study, participants were scanned while answering the memory questionnaire (see above) using an MRI-compatible response box (fiber optic response pad, Current Designs Inc.). The test was divided into two consecutive scans and was followed by an anatomical scan. After scanning, participants were debriefed about their retrieval strategies and memory performance during the test.

MRI acquisition
Scanning was performed on a 3T Trio Magneton Siemens scanner located at the Human Brain Imaging Center of the Weizmann Institute of Science. During each fMRI scan, a time series of volumes was acquired using a T2*-weighted gradient-echo echo planar imaging (EPI) pulse sequence (TR 2000 msec, TE 30 msec, flip angle 80°, 35 oblique slices without gap, 30° from ACPC, $3 \times 3 \times$ 4-mm voxels, FOV 240 mm). All images were acquired using a head coil (12–channels head matrix coil, Siemens Medical Solutions). In addition, T1-weighted high-resolution ($1 \times 1 \times$ 1 mm) anatomical images were acquired for each subject with a magnetization prepared rapid acquisition gradient-echo (MP-RAGE) pulse sequence (TE 2.98 msec, TR 2300 msec, TI 900

msec, α 9°) to allow accurate 3D reconstruction and volume-based statistical analysis.

Analysis of memory performance data
Memory performance was calculated from the Recognition phase of the entire memory test; the first question at the beginning of each scan was discarded, and the two scans were concatenated. Correct mRecall trials were designated based on answers to the Recognition task immediately following.

Analysis of fMRI data
FMRI data were preprocessed and analyzed using the BrainVoyager QX (v1.8.6 and v1.10.4) software package (Brain Innovation) and in-house supplementary software (MATLAB; The Mathworks). Preprocessing of functional scans included slice-time correction, head movement detection and realignment (scans with head movement larger than 2 mm were rejected), high-pass temporal filtering, and removal of linear trends. Participants' data were transformed into Talairach space, resampled into $3 \times 3 \times$ 3-mm voxels, and spatially smoothed by a Gaussian filter of 6-mm full width at half maximum (FWHM).

BOLD response analysis
The experiment was constructed as an event-related design with relatively long events (8 sec on average). For analysis of BOLD response, we applied finite impulse response (FIR) models, for which conditions were modeled from trial onset to 20 sec (10 TRs) post-onset. Such models assume that activity in each TR is the sum of overlapping BOLD-response patterns (Shulman et al. 1999; Miezin et al. 2000). Two models were used for analysis: (1) The Inclusive model, designed to delineate the retrieval-related network, containing four conditions: non-guess mRecall events, non-guess Recognition events, Control events, and Guess events; and (2) the Performance model, used for region-of-interest (ROI) analysis of activity in retrieval-related regions, containing the same four conditions as in the Inclusive model, but dividing the mRecall and Recognition conditions into correct and incorrect responses, resulting in a total of six conditions. For additional BOLD response analyses using more models incorporating playback and correspondence parameters, see the Supplemental Material.

The Inclusive model, which incorporated data from the Hours, Weeks, and Months groups, was used to delineate the retrieval-related brain networks by contrasting activity between the mRecall and Control tasks at the 4th TR (8 sec after mRecall phase question screen was presented). This time point was chosen to capture differences in evoked hemodynamic response, which for randomly presented fMRI trials is expected to peak 8–12 sec after event onset (Friston et al. 1995). We find an identical network when contrasting all 10 TRs following onset of mRecall trials with all 10 TRs following onset of the Control task (Supplemental Fig. S1). The contrast was performed using random-effects analysis of all groups in order to allow systematic comparison of identical ROIs among the three groups. Statistical maps were corrected for multiple comparisons using cluster-size thresholding at confidence of $\alpha = 0.001$, resulting in minimum cluster size of 1269 mm^3. Correction was implemented by the "ClusterThresh" plug-in in BrainVoyager QX (Forman et al. 1995; Goebel et al. 2006). ROIs were defined from the functional clusters that survived correction for multiple comparisons. Because some clusters were large and contiguous (Fig. 3A), they were subdivided into ROIs according to the anatomical landmarks and boundaries established in the literature (Duvernoy et al. 1991; Mai et al. 1997). We assigned ROI labels according to anatomical location of the entire cluster and not only according to the peak voxel location. Clusters that spanned several anatomical locations were subdivided into separate ROIs. The resulting network was used for further ROI analysis and is referred to throughout as the retrieval network.

Note that very similar retrieval network maps were generated when constructing statistical maps for each group separately (excluding hippocampus) (Fig. 3B), and BOLD response analysis

using group-specific ROIs gave results similar to those reported below, using identical ROIs (Supplemental Fig. S2).

For each ROI of the retrieval network, we extracted an average BOLD signal from the entire ROI per subject, using the Performance model. We did not compute further contrasts using this model, but only averaged the BOLD response during correct mRecall and correct Recognition trials separately for the Hours, Weeks, and Months groups. We did not compare correct to incorrect trials in analyses because (a) trial number was insufficient to allow reliable results, and (b) incorrect trials may stem from cognitive processes that differ widely both within and between participants and thus may introduce a great deal of noise in analyses of neural activity. We compared the BOLD response between groups per trial type using a two-way mixed ANOVA with Group (i.e., Hours, Weeks, and Months) as a between-subject factor and BOLD estimate per TR as a repeated-measures within-subject factor. BOLD response was considered to differ among groups when the ANOVA yielded a significant interaction between Group and BOLD estimate over time (Zacks et al. 2001). ROIs showing significant interaction were subjected to further analysis that directly compared BOLD responses at TR4 of correct mRecall trials and TR4 of correct Recognition trials separately using one-way between-group ANOVAs, complemented by post hoc Fisher's LSD t-tests.

Brain–behavior correlation analysis

To examine inter-subject variability in memory performance and its correspondence to brain activity, we computed Pearson correlations between these two factors. As an estimate for memory performance, we used the percent of correct answers as determined in the Recognition phase (see above). As an estimate for memory-related brain activity, we used the difference in BOLD activity at TR4 (i.e., 8 sec post-onset) between the memory retrieval and control tasks (i.e., correct mRecall minus Control at TR4, and correct Recognition minus Control at TR4). This procedure was performed for each subject separately.

For testing the statistical significance of the computed correlations, a nonparametric permutation test was applied. In this test, for each of the groups' data sets and in each ROI, the level of memory performance (i.e., percent of accurate answers) was randomly assigned to subjects 1000 times. Thus, for each ROI, 1000 correlation coefficients were calculated between retrieval-related brain activity and inter-subject shuffled permutations of memory performance values. Finally, these randomly generated correlation values were plotted on a histogram, along with the original correlation value. The original, nonpermutated correlation value was considered significant if it exceeded a threshold of $P < 0.05$ in relation to the permutated histogram.

Acknowledgments

References

Addis DR, Moscovitch M, Crawley AP, McAndrews MP. 2004. Recollective qualities modulate hippocampal activation during autobiographical memory retrieval. *Hippocampus* **14:** 752–762.

Addis DR, Wong AT, Schacter DL. 2007. Remembering the past and imagining the future: Common and distinct neural substrates during event construction and elaboration. *Neuropsychologia* **45:** 1362–1377.

Badre D, Poldrack RA, Paré-Blagoev EJ, Insler RZ, Wagner AD. 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* **47:** 907–918.

Bar M. 2004. Visual objects in context. *Nat Rev Neurosci* **5:** 617–629.

Bontempi B, Laurent-Demir C, Destrade C, Jaffard R. 1999. Time-dependent reorganization of brain circuitry underlying long-term memory storage. *Nature* **400:** 671–675.

Brewer WF. 1996. What is recollective memory? In *Remembering our past. Studies in autobiographical memory* (ed. D Rubin), pp. 19–65. Cambridge University Press, New York.

Bullmore E, Sporns O. 2012. The economy of brain network organization. *Nat Rev Neurosci* **13:** 336–349.

Cohen G. 1996. *Memory in the real world*, 2nd ed. Psychology Press, Hove, UK.

Dudai Y. 2004. The neurobiology of consolidations, or, how stable is the engram? *Ann Rev Psychol* **55:** 51–86.

Dudai Y. 2012. The restless engram: Consolidations never end. *Ann Rev Psychol* **35:** 222–247.

Dudai Y, Carruthers M. 2005. The Janus face of Mnemosyne. *Nature* **434:** 567. doi: 10.1038/434567a.

Duvernoy HM, Cabanis EA, Vannson JL. 1991. *The human brain: Surface, three-dimensional sectional anatomy and MRI*. Springer-Verlag, New York.

Fischer S, Nitschke MF, Melchert UH, Erdmann C, Born J. 2005. Motor memory consolidation in sleep shapes more effective neuronal representations. *J Neurosci* **25:** 11248–11255.

Fletcher PC, Henson RN. 2001. Frontal lobes and human memory: Insights from functional neuroimaging. *Brain* **124:** 849–881.

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC. 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magn Reson Med* **33:** 636–647.

Frankland PW, Bontempi B. 2005. The organization of recent and remote memories. *Nat Rev Neurosci* **6:** 119–130.

Frankland PW, Bontempi B, Talton LE, Kaczmarek L, Silva AJ. 2004. The involvement of the anterior cingulate cortex in remote contextual fear memory. *Science* **304:** 881–883.

Friston KJ, Frith CD, Frackowiack RSJ, Turner R. 1995. Characterizing dynamic brain responses with fMRI: A multivariate approach. *Neuroimage* **2:** 166–172.

Furman O, Dorfman N, Hasson U, Davachi L, Dudai Y. 2007. They saw a movie: Long-term memory for an extended audiovisual narrative. *Learn Mem* **14:** 457–467.

Gallagher HL, Frith CD. 2003. Functional imaging of 'theory of mind.' *Trends Cogn Sci* **7:** 77–83.

Gilboa A. 2004. Autobiographical and episodic memory—one and the same? Evidence from prefrontal activation in neuroimaging studies. *Neuropsychologia* **42:** 1336–1349.

Gilboa A, Winocur G, Grady CL, Hevenor SJ, Moscovitch M. 2004. Remembering our past: Functional neuroanatomy of recollection of recent and very remote personal events. *Cereb Cortex* **14:** 1214–1225.

Goebel R, Esposito F, Formisano E. 2006. Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortical aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* **27:** 392–401.

Gold BT, Buckner RL. 2002. Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron* **35:** 803–812.

Gusnard DA, Akbuda E, Shulman GL, Raichle ME. 2001. Medial prefrontal cortex and self-referential mental activity: Relation to a default mode of brain function. *Proc Natl Acad Sci* **98:** 4259–4264.

Haist F, Bowden Gore J, Mao H. 2001. Consolidation of human memory over decades revealed by functional magnetic resonance imaging. *Nat Neurosci* **4:** 1139–1145.

Hassabis D, Maguire EA. 2007. Deconstructing episodic memory with construction. *Trends Cogn Sci* **11:** 299–306.

Hasson U, Furman O, Clark D, Dudai Y, Davachi L. 2008. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron* **57:** 452–462.

Henderson JM, Larson CL, Zhu DC. 2008. Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: An fMRI study. *Brain Cogn* **66:** 40–49.

Karpicke JD, Roediger HL. 2008. The critical importance of retrieval for learning. *Science* **319:** 966–968.

Kim JJ, Fanselow MS. 1992. Modality-specific retrograde amnesia of fear. *Science* **256:** 675–677.

Maguire EA, Henson RN, Mummery CJ, Frith CD. 2001. Activity in prefrontal cortex, not hippocampus, varies parametrically with the increasing remoteness of memories. *Neuroreport* **12:** 441–444.

Mai JK, Assheuer J, Paxinos G. 1997. *Atlas of the human brain*. Academic Press, San Diego.

Mar RA. 2004. The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia* **42:** 1414–1434.

Maviel T, Durkin TP, Menzaghi F, Bontempi B. 2004. Sites of neocortical reorganization critical for remote spatial memory. *Science* **305:** 96–99.

McClelland JL, McNaughton BL, O'Reilly RC. 1995. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev* **102:** 419–457.

McDermott KB, Szpunar KK, Christ SE. 2009. Laboratory-based and autobiographical retrieval tasks differ substantially in their neural substrates. *Neuropsychologia* **47:** 2290–2298.

Mendelsohn A, Furman O, Navon I, Dudai Y. 2009. Subjective vs. documented reality: A case study of long-term real-life autobiographical memory. *Learn Mem* **16:** 142–144.

Mendelsohn A, Furman O, Dudai Y. 2010. Signature of memory: Brain coactivations during retrieval distinguish correct from incorrect recollection. *Front Behav Neurosci* **4:** 18. doi: 10.3389/fnbeh.2010.00018.

Miezin FM, Macotta L, Ollinger JM, Petersen SE, Buckner RL. 2000. Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage* **11:** 735–759.

Mitchell JP, Macrae CN, Banaji MR. 2004. Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *J Neurosci* **24:** 4912–4917.

Moscovitch M, Nadel L. 1998. Consolidation and the hippocampal complex revisited: In defense of the multiple-trace model. *Curr Opin Neurobiol* **8:** 297–300.

Moscovitch M, Winocur G. 2002. The frontal cortex and working with memory. In *Principles of frontal lobe function* (ed. DT Stuss, RT Knight), 188–209. Oxford University Press, New York.

Nadel L, Moscovitch M. 1997. Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol* **7:** 217–227.

Nader K, Hardt O. 2009. A single standard for memory: The case for reconsolidation. *Nat Rev Neurosci* **10:** 224–234.

Petrides M. 2002. The mid-ventrolateral prefrontal cortex and active mnemonic retrieval. *Neurobiol Learn Mem* **78:** 528–538.

Piolino P, Desgranges B, Eustache F. 2009. Episodic autobiographical memories over the course of time: Cognitive, neuropsychological and neuroimaging findings. *Neuropsychologia* **47:** 2314–2319.

Roediger HL, Butler AC. 2011. The critical role of retrieval practice in long-term retention. *Trends Cog Sci* **15:** 20–27.

Rudy JW, Biedenkapp JC, O'Reilly RC. 2005. Prefrontal cortex and the organization of recent and remote memories: An alternative view. *Learn Mem* **12:** 445–446.

Ryan L, Nadel L, Keil K, Putnam K, Schnyer D, Trouard T, Moscovitch M. 2001. Hippocampal complex and retrieval of recent and very remote autobiographical memories: Evidence from functional magnetic resonance imaging in neurologically intact people. *Hippocampus* **11:** 707–714.

Shulman GL, Ollinger JM, Akbudak E, Conturo TE, Snyder AZ, Petersen SE, Corbetta M. 1999. Areas involved in encoding and applying directional expectations to moving objects. *J Neurosci* **19:** 9480–9496.

Smith CN, Squire LR. 2009. Medial temporal lobe activity during retrieval of semantic memory is related to the age of the memory. *J Neurosci* **29:** 930–938.

Squire LR, Bayley PJ. 2007. The neuroscience of remote memory. *Curr Opin Neurobiol* **17:** 185–196.

Squire LR, Stark CEL, Clark RE. 2004. The medial temporal lobe. *Annu Rev Neurosci* **27:** 279–306.

Steinvorth S, Corkin S, Halgren E. 2006. Ecphory of autobiographical memories: An fMRI study of recent and remote memory retrieval. *Neuroimage* **30:** 285–298.

Summerfield JJ, Hassabis D, Maguire EA. 2009. Cortical midline involvement in autobiographical memory. *Neuroimage* **1:** 1188–1200.

Svoboda E, McKinnon MC, Levine B. 2006. The functional neuroanatomy of autobiographical memory: A meta-analysis. *Neuropsychologia* **44:** 2189–2208.

Takashima A, Petersson KM, Rutters F, Tendolkar I, Jensen O, Zwarts MJ, McNaughton BL, Fernández G. 2006. Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proc Natl Acad Sci* **103:** 756–761.

Takashima A, Nieuwenhuis IL, Jensen O, Talamini LM, Rijpkema M, Fernández G. 2009. Shift from hippocampal to neocortical centered retrieval network with consolidation. *J Neurosci* **29:** 10087–10093.

Talairach J, Tournoux P. 1988. *Co-planar stereotaxic atlas of the human brain*. Thieme Medical Publishers, New York.

Tambini A, Ketz N, Davachi L. 2010. Enhanced brain correlations during rest are related to memory for recent experiences. *Neuron* **65:** 280–290.

Tse D, Takeuchi T, Kakeyama M, Kajii Y, Okuno H, YTohyama C, Biot H, Morris RGM. 2011. Schema-dependent gene activation and memory encoding in neocortex. *Science* **333:** 891–895.

Tulving E. 1995. Memory and consciousness. *Can Psychol* **26:** 1–12.

Vann SD, Aggleton JP, Maguire EA. 2009. What does the retrosplenial cortex do? *Nat Rev Neurosci* **10:** 792–802.

Viard A, Piolino P, Desgranges B, Chételat G, Lebreton K, Landeau B, Young A, De La Sayette V, Eustache F. 2007. Hippocampal activation for autobiographical memories over the entire lifetime in healthy aged subjects: An fMRI study. *Cereb Cortex* **17:** 2453–2467.

Viard A, Lebreton K, Chetelat G, Desgranges B, Landeau B, Young A, De La Sayette V, Eustache F, Piolino P. 2010. Patterns of hippocampal–neocortical interactions in the retrieval of episodic autobiographical memories across the entire life-span of aged adults. *Hippocampus* **20:** 153–165.

Wiltgen BJ, Zhou M, Cai Y, Balaji J, Karlsson MG, Parivash SN, Li W, Silva AJ. 2010. The hippocampus plays a selective role in the retrieval of detailed contextual memories. *Curr Biol* **20:** 1336–1344.

Winocur G, Moscovitch M. 2011. Memory transformation and systems consolidation. *J Intern Neuropsychol Soc* **17:** 1–15.

Winocur G, Moscovitch M, Bontempi B. 2010. Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia* **48:** 2339–2356.

Zacks JM, Braver TS, Sheridan MA, Donaldson DI, Snyder AZ, Ollinger JM, Buckner RL, Raichle ME. 2001. Human brain activity time-locked to perceptual event boundary. *Nat Neurosci* **4:** 651–655.