Neural mechanisms of reactivation-induced updating that enhance and distort memory

Peggy L. St. Jacques a,b,1, Christopher Olm a,b, and Daniel L. Schacter a,b,1

*Department of Psychology and bCenter for Brain Science, Harvard University, Cambridge, MA 02138

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2013.

Contributed by Daniel L. Schacter, October 17, 2013 (sent for review September 10, 2013)

We remember a considerable number of personal experiences because we are frequently reminded of them, a process known as memory reactivation. Although memory reactivation helps to stabilize and update memories, reactivation may also introduce distortions if novel information becomes incorporated with memory. Here we used functional magnetic resonance imaging (fMRI) to investigate the neural mechanisms mediating reactivation-induced updating in memory for events experienced during a museum tour. During scanning, participants were shown target photographs to reactivate memories from the museum tour followed by a novel lure photograph from an alternate tour. Later, participants were presented with target and lure photographs and asked to determine whether the photographs showed a stop they visited during the tour. We used a subsequent memory analysis to examine neural recruitment during reactivation that was associated with later true and false memories. We predicted that the quality of reactivation, as determined by online ratings of subjective recollection, would increase subsequent true memories but also facilitate incorporation of the lure photograph, thereby increasing subsequent false memories. The fMRI results revealed that the quality of reactivation modulated subsequent true and false memories via recruitment of left posterior parahippocampal, bilateral retrosplenial, and bilateral posterior inferior parietal cortices. However, the timing of neural recruitment and the way in which memories were reactivated contributed to differences in whether memory reactivation led to distortions or not. These data reveal the neural mechanisms recruited during memory reactivation that modify how memories will be subsequently retrieved, supporting the flexible and dynamic aspects of memory.

Significance

Reactivation is a key process that updates memory by strengthening existing memories and incorporating relevant new information, thus supporting the dynamic and flexible nature of memory. This adaptive function, however, can sometimes contribute to memory distortions. The current study examines how neural mechanisms that operate during reactivation of memories for a museum tour contribute to enhancement of existing memories, while also supporting integration of novel information that can contribute to false memories. Our results reveal similarities and differences in the neural mechanisms of reactivation associated with subsequent true and false memories for real-world events, thereby illuminating how memories change over time as a consequence of reactivation—a process that has important implications for understanding the unreliability of eyewitness memories.

Author contributions: P.L.S.J. and D.L.S. designed research; P.L.S.J. and C.O. performed research; P.L.S.J. analyzed data; and P.L.S.J., C.O., and D.L.S. wrote the paper.

The authors declare no conflict of interest.

See QnAs 10.1073/pnas.1320211110.

1To whom correspondence may be addressed. E-mail: peggyjs@wjh.harvard.edu or dls@wjh.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1319630110/-/DCSupplemental.
Updating relies on the process of memory reactivation, or the activation of a latent memory trace when we are reminded of a past experience, which is a key process that shapes long-term memory representations by reorganizing them over distributed brain networks (39–41). In this paper we examine the effects of memory reactivation on the neural mechanisms mediating subsequent true and false memories. In a previous behavioral study, we showed that the quality of memory reactivation, as indexed by an individual’s subjective sense of recollection, modulates the extent to which reactivation strengthens subsequent true memories for a recently presented target event or creates false memories of novel information encountered for the first time after the target event has already occurred (42). We found that subsequent true memories were improved when targets were highly reactivated (i.e., the retrieval cues during reactivation matched the encoding experience) compared with memories that were reactivated at lower levels (i.e., the retrieval cues during reactivation mismatched encoding); however, subsequent false memories were also greater for lures that followed targets that were highly reactivated. A primary challenge of such research is to understand how neural mechanisms that operate during memory reactivation contribute to enhancement of existing memory traces while also supporting integration of novel information that can contribute to false memories.

Memory recruits a typical pattern of brain regions (43, 44), including frontoparietal network regions associated with controlled processes and default network regions linked to recovery of memory details (45). In particular, hippocampus, parahippocampal cortex, retrosplenial cortex, and posterior parietal cortex are associated with reactivation processes during retrieval that contribute to the quality of memory reactivation (46). Recovery of contextual information associated with the encoding experience that supports these recollection processes, however, can sometimes contribute to memory distortions as predicted by some computational models of memory (47). Reactivating (i.e., the retrieval cues during reactivation) overlapping memory representations (57, 58), in line with known hippocampal recruitment supports persistent changes and increased memory for the lures that followed. In sum, we found an increase in subsequent true and false memories for reactivated items compared with baseline items, and that the reactivation condition equally increased subsequent hits ($t(32) = 6.05, P < 0.0001$) and subsequent false alarms ($t(32) = 6.67, P < 0.0001$) compared with the baseline condition consisting of target and lure photographs that were not shown during session 2 (Fig. 1B). To link the reactivation-related increases in subsequent memory to the quality of reactivation, we examined reliving ratings, the subjective sense of reexperience or recollection, during session 2 according to the subsequent memory outcome in session 3. These results showed that stronger reliving of the target improved memory for the target, but also increased memory for the lure that followed (Fig. 1C). Participants made higher reliving ratings for photographs associated with subsequent hits versus subsequent misses ($t(32) = 13.68, P < 0.0001$) and with subsequent false alarms versus subsequent correct rejections ($t(32) = 3.96, P = 0.0003$). The link between reliving and subsequent memory was also evident on a trial-by-trial level, such that there was a significant within-participant correlation between reliving ratings and subsequent true memories ($r = 0.34, P < 0.0001$) and between reliving ratings and subsequent false memories ($r = 0.09, P = 0.0004$). Additionally, the influence of reliving on subsequent memory remained after controlling for how quickly the memory was retrieved and the relatedness of novel information (SI Results). Thus, the quality of reactivation during session 2 uniquely influenced the outcome of both subsequent true and subsequent false memories, such that on a trial-by-trial basis, stronger reliving improved memory for the targets and increased memory for the lures that followed. In sum, we found an increase in subsequent true and false memories for reactivated items compared with baseline items, and that the
quality of reactivation influenced the level of true and false memories. The current results support the findings of our previous behavioral study using a similar paradigm (42). For additional behavioral results, see SI Results.

Reactivation-related neural regions associated with subsequent memory.

The main goal of the fMRI analysis was to examine how neural mechanisms recruited during reactivation of memory for real-world events experienced during the museum tour influence updating processes that result in later enhancement or distortion in memory. Thus, we separated trials as a function of later true memories (i.e., hits minus misses) and later false memories (i.e., false alarms minus correct rejections; Materials and Methods). The subsequent memory analysis revealed that both subsequent true-memory and subsequent false-memory effects were associated with recruitment of left posterior parahippocampal, bilateral retrosplenial, and bilateral posterior inferior parietal cortices (Fig. 2 and Table S1).

Although reactivation-related updating in later memory recruited similar neural mechanisms for subsequent true and false memories, we found that the timing of these effects differed by examining neural recruitment during target (phase 1) and lure (phase 2) presentation. There was a significant interaction between trial phase (target photograph, lure photograph) and subsequent memory (true, false) in the recruitment of the posterior parahippocampal, retrosplenial, and posterior inferior parietal cortices (Fig. 2B and Table S2). The interaction was reflected by greater recruitment during target than during lure phases for subsequent true memories but equal recruitment of these regions across both trial phases for subsequent false memories. Thus, true and false subsequent memories were distinguished by neural mechanisms recruited when processing the lure photograph. For false memories, sustained involvement of the posterior parahippocampal, retrosplenial, and posterior inferior parietal cortices across memory reactivation and the presentation of novel information was associated with whether the lure photograph would be incorporated into later memories for the museum tour. In contrast, for true memories, reduced recruitment of these same brain regions during presentation of the lure was associated with accurate subsequent memories.

We conducted two additional follow-up analyses to determine whether the pattern of neural recruitment supporting memory updating was associated with memory reactivation. First, we examined how activation of these regions was related to trial-by-trial differences in the quality of memory reactivation by...
A challenge in memory research has been to understand the neurobiological mechanisms of memory updating and how neurobiological mechanisms of memory updating and how neurobiological mechanisms of memory updating and how...
Differential subsequent memory effects related to reactivation quality of memory condition interaction. Error bars in spatial representations in the image 0x1 to 19x816 of false memories (63). However, there is conflicting evidence regarding the nature of awake reactivation, with some studies showing that reactivation enhances or even protects memory from new information (54, 64, 65) and other studies indicating that reactivation incorporates new information into memory (49, 65–67). The present study significantly extends this work by demonstrating that reactivation can both protect true memories and support the incorporation of new information leading to false memories, depending upon neural recruitment during reactivation.

Memory updating was associated with the recruitment of a subset of retrieval-related regions that were sensitive to the quality of reactivation. Here we defined reactivation quality according to subjective recollection, and thus, our findings could also reflect the contribution of memory strength during reactivation (68). There was little overlap in neural recruitment supporting subsequent memory effects during reactivation compared with novel encoding. Indeed, many of these retrieval-related brain regions were deactivated compared with novel encoding during novel encoding tasks (69), suggesting less contribution from the typical subsequent memory regions during memory updating. One way that reactivation may support memory updating is via contextual reinstatement, by linking novel information that occurs during retrieval with the reinstated context of the reactivated memory (47, 49). Contextual reinstatement could underlie the link between the quality of reactivation and subsequent memory effects (42). In the current study, contextual reinstatement would include reconstruction of the rich 3D spatial setting of the museum stop from the photographic retrieval cue. In line with this idea, subsequent memory effects were associated with recruitment of posterior parahippocampal, retrosplenial, and posterior inferior parietal cortices, which contribute to the recovery of spatial context and scene construction (70, 71). Retrosplenial cortex is thought to support integration and translation between egocentric spatial representations in posterior parietal cortex and allocentric spatial representations in the MTL (71), and these processes are recruited to a greater extent when spatial context needs to be updated and manipulated (72). Contextual reinstatement may strengthen existing memory and/or integrate new information with novel information with baseline by the retrieval cues, in much the same way that retrieval practice is thought to support memory via elaboration of the memory trace (73). This restructuring of memories could facilitate re-encoding and reconsolidation processes that contribute to updating of long-term memory representations.

Neural recruitment sensitive to reactivation quality differed depending upon the outcome of subsequent memory. Our finding that subsequent false alarms were associated with greater involvement of anterior hippocampus and vmPFC builds upon accumulating evidence linking these regions to flexible memory processes that enable memories to be combined and used in novel ways. Hippocampus and vmPFC are part of a MTL subsystem of the default network (74) that supports the formation of mental models based on mnemonic content during both memory retrieval and the simulation and encoding of future events (36, 75, 76). In particular, the anterior portion of the hippocampus is associated with relational processes (77–79) that contribute to the ability to integrate memory details across experiences (80). Additionally, the anterior hippocampus represents global context through its connections with vmPFC and related schematic processes that could support the incorporation of novel information (59). Our results converge with these findings, but extend this work by demonstrating that the quality of reactivation is a key mechanism and also by showing that these processes support the formation of false memories. We also found that anterior hippocampus was sensitive to relatedness ratings more for false alarms than hits during presentation of the target (Table S5). One explanation is that target reactivation and related pattern completion processes may linger during the processing of the lure and affect the degree of relatedness of the lures (81), thus contributing to false memories. In line with this idea, we found a significant association between reliving and relatedness ratings, and that this association contributed more to false than true memories. In contrast, posterior hippocampus contributed equally to subsequent hits and false alarms, consistent with its
role in representing detailed contextual information that more generally supports recollection (59). Compared with subsequent false alarms, we found that subsequent hits were associated with posterior cingulate cortex and rostral medial PFC, which are core midline regions of the default network (74) that support self-referential processes (82, 83) and successful encoding of self-relevant information (84–86). These findings suggest that reactivation may enhance true memories by elaborating upon memory details related to oneself (87), but increase false memories via binding of novel information.

Our results add to the growing literature on the neural mechanisms supporting false memories (88). The memory distortion effects we observed here in some ways resemble those that occur in the postevent misinformation paradigm (2), in which erroneous information presented after encoding contributes to later false memories because of source-memory confusion. During encoding of misinformation, some studies have shown recruitment of anterior and posterior midline regions that protect true memories (51), and hippocampal recruitment that supports the formation of false memories (52). Our results converge with these findings, but further indicate that neural recruitment that differentiates subsequent true and false memories during presentation of novel information depends on the extent to which memory for the original experience is reactivated rather than source confusion between the presence or absence of a naturalistic event experienced in a real-world setting, which modulates both the extent of subsequent memory effects, and the neural recruitment associated with these memory effects. Similarly, in our previous study using the museum paradigm (42), we manipulated the quality of reactivation and showed that subsequent false memories were greater when lures followed targets where memories were highly reactivated (i.e., the retrieval cues during reactivation matched encoding experience) than for memories that were reactivated at lower levels (i.e., the retrieval cues during reactivation mismatched encoding). We suggested that reactivation not only reinforces our personal pasts (89)—voluntarily, as we share a memory with another individual, or involuntarily, as we spontaneously bring to mind a past event—our findings build upon accumulating evidence that memory is shaped by such retrievals (90, 91). Our data show that the quality of reactivation is one mechanism by which retrieval influences memory, and we suggest a link between contextual reinstatement via neural recruitment of the posterior parahippocampal, retrosplenial, and posterior inferior parietal cortices in the enhancement and distortion in later memory. The current study reveals neural mechanisms that support the formation of false memories for a lure. Thus, here we show that the real-world setting of reactivation has important implications for eyewitness memory and the law (92, 93). Indeed, Schacter and Loftus (92) argued that understanding the neural mechanisms of memory reactivation and reconciliation could provide a foundation for understanding how memories change over time, which in turn could eventually help better understand why eyewitness memories sometimes change in response to repeated questioning. Our findings also fit with an adaptive perspective on memory distortion (6), 20–23), in which reactivation allows for the incorporation of relevant new information that is essential for the operation of a dynamic memory system, but which comes at the cost of memory distortions. It is an open question as to whether all varieties of memory distortion can be conceived as costs associated with adaptive features of memory, but our results are consistent with the view that reactivation-related false memories reflect one downside to the generally beneficial process of memory updating.

Materials and Methods

Participants. There were 35 participants (18–30 y old, 19 women). All participants were right-handed and reported no history of neurological or psychiatric episodes or current use of medication known to affect cognitive function. Participants gave written consent for a protocol approved by the Harvard University Intuiotional Review Board. Two participants were excluded due to computer issues. Seven additional participants were excluded from the fMRI analysis only because of quality control issues (SI Materials and Methods).

Procedure. The study involved three sessions separated by 48 h: a museum tour, an fMRI scanning, and a recognition memory test. In session 1, participants were provided with an iTouch (Apple) outlining a self-guided audio tour of the adjoining Harvard Museum of Natural History and Peabody Museum and asked to wear a ViconRevue camera (Vicon, Oxford, UK), which automatically takes photographs every 15 s using a timer. The tour was composed of 208 museum stops (e.g., examining a display case, watching a video, etc.), and took ~4.5 h to complete. There were two versions of the tour, which were counterbalanced between participants. Photographs of museum stops from the alternate tour were used as lures. Museum stops in the two tour versions were matched to be similar in content (e.g., one video versus another video in the same exhibit; Fig. 1A), but selected to minimize overlap in the route through the museum. However, to increase the number of trials available for subsequent memory fMRI analysis, the museum tour included 41 stops that overlapped in both tour versions (i.e., participants in each tour version visited the same museum stop). Overlapping museum stops included items that would be unavoidable during each route through the museum (i.e., a large display case in the middle of the exhibit) and/or lures (a photograph of a lure from the same exhibit was unavailable (i.e., an odd numbered display case in an exhibit), and were included as partial trials in session 2 (i.e., trials in which the target photograph was not followed by a lure). Participants were instructed to complete only the museum stops described in the tour guide. Photographs from each participant’s camera were inspected to ensure that the participant adhered to the instructions, and if the camera captured a photograph of a unique museum stop from the alternate tour, it was deleted from further analysis. Photographs from each participant’s camera for each museum stop were selected to use in the later sessions. Lure photographs from the alternate tour were taken from a control set.

Session 2 took place in the MRI scanner. The scanning session included a 6-min run of quiet rest with eyes open, four runs of the reaction condition, another 6-min resting state run, four runs of the novel encoding condition in blocks of six (fMRI Methods), and another 6-min resting state run. The order of reactivation and encoding conditions was counterbalanced between participants.

During the reactivation condition, participants were asked to retrieve memories for the museum stops they visited during the tour. On each trial, participants were shown a photograph of a museum stop taken from their camera and instructed to retrieve their memory for that museum stop and then to rate the sense of reliving, which refers to the subjective sense of recollection or reexperience, on a five-point scale from low to high. The photograph remained on the screen for 5 s. For partial trials (~64 trials), fixation immediately followed. For full trials (~112 trials), a second photograph of a museum stop from the alternate tour that was not seen during the participant’s tour (i.e., lure photograph) immediately followed. Full and partial trials were included to separate neural recruitment during the trial (i.e., target lure) vs. comparison (i.e., lure photograph). Participants were instructed to indicate how related (i.e., “could it be taken from the same exhibit?”) the second photograph was to the preceding one, on a five-point scale from low to high. They were not told whether the lure photographs were from the tour or not. The photograph remained on the screen for 5 s. Trials were separated by a variable fixation (2.5–7.5 s) and distributed exponentially such that shorter intertrial intervals occurred more frequently than longer. Approximately 80% of the museum stops were shown during reaction (i.e., 176 trials), and the remaining museum stops were used for the baseline condition (i.e., 32 trials).

Session 3 involved an old/new recognition task. Participants were shown target and lure photographs taken from the museums they visited during the tour and asked to make a yes/no decision whether the photograph was a stop from their museum tour. Photographs consisted of reactivated targets (both partial and full trials, ~120 trials) and lures (i.e., photographs that were shown during scanning session, ~56 trials) and baseline (i.e., photographs that were not shown during scanning session) targets (~16 trials) and lures (~16 trials). Participants were warned that the lure photographs would look very similar to stops that they had conducted during the tour and to look carefully at each item before making their decision. Participants were allowed up to 6 s to make their decision, followed by a 6-s confidence rating on a five-point scale from low to high.

fMRI Methods. Image acquisition. Imaging was conducted on a 3T Siemens Magnetom TimTrio Scanner equipped with a 12-channel head coil at the
Center for Brain Science (Harvard University). A laptop computer running Cogent 2000 (Wellcome Department of Imaging Neuroscience, University College London, UK) software implemented in MATLAB (MathWorks) controlled stimulus display via a liquid crystal display projector, which projected onto a screen placed at the head of the bore. Participants viewed the screen through a mirror fastened to the head coil. Cushions were used to minimize head movement and earplugs dampened scanner noise. Participants made responses using a five-button box placed in their right hand.

Anatomical images were acquired using a high-resolution 3D magnetization-prepared rapid gradient echo sequence (176 sagittal slices, echo time $\text{TE} = 1.64$ ms, repetition time $\text{TR} = 2530$ ms, flip angle = $7^\circ$, voxel size = $1 \times 1 \times 1$ mm). Functional images were collected using a T2* gradient echo, echo-planar imaging sequence sensitive to blood oxygen level-dependent (BOLD) contrast ($\text{TR} = 2500$ ms, $\text{TE} = 30$ ms, flip angle = $90^\circ$, $3 \times 3 \times 3$ mm in-plane resolution). Whole-brain coverage was obtained with 39 contiguous slices, acquired parallel to the anterior-posterior commissure plane (3-mm slice thickness, 0.5-mm skip between slices).

**Image processing.** Imaging data were preprocessed and statistically analyzed using SPM8 (Wellcome Department of Imaging Neuroscience, London, UK). First, data were preprocessed to remove sources of noise and artifact. Preprocessing included slice-time correction to correct for differences in acquisition times for each whole brain volume; realignment within and across runs to correct for head movement; spatial normalization to the Montreal Neurological Institute template (resampled at $2 \times 2 \times 2$ mm voxels) and spatial smoothing (8-mm full-width at half maximum) using a Gaussian kernel.

**fmRI analysis.** Preprocessed data were analyzed using the general linear model (GLM) for each participant trial onset (i.e., time-locked to target or lure stimulus presentation) were modeled with a canonical hemodynamic response function and duration of $5$ s.

We used a compound trial approach (94) to separate the hemodynamic response associated with target and lure presentation within the same trial in the reactivation condition. Because the stimulus onsets for phase 2 (lures) occurred at a fixed interval after the stimulus onset of phase 1 (targets), partial trials were used to separate the BOLD responses of the phases. Approximately $36\%$ of the total trials were partial. A similar approach was used to separately analyze the pairs of novel photographs presented in the novel encoding condition (Fig. S4 and SI Materials and Methods).

To combine the compound trial approach and subsequent memory analysis on true and false memories, it was necessary to create two GLMs in the reactivation condition. The models differed in the ability to examine false alarm and correct rejection trial types in the two phases because the nature of the task design did not allow for partial trials within these categories. To address this issue, $50\%$ of the partial trials were randomly assigned to a new partial trial type to effectively separate the BOLD response of phase 1 and phase 2 for false alarms and correct rejections. One GLM was created to examine phase-1 effects only; it included a combined phase-2 trial type (full, false, misses, and false alarms, and correct rejections), phase-1 hits (full + partial trials), phase-1 misses (full + partial trials), phase-1 false alarms (full trials only), phase-1 correct rejections (full trials only), and phase-1 new partial trials. Another GLM was created to examine phase-2 effects only and included a combined phase-1 trial type, (full false alarms, full correct rejections, and new partial trials), phase-1 hits (full + partial trials), phase-2 misses (full + partial trials), phase-2 hits (full trials only), phase-2 misses (full trials only), phase-2 false alarms (full trials only), and phase-2 correct rejections (full trials only).

We then used a subsequent memory analysis. We examined neural activity during retrieval that was associated with subsequent true memories (i.e., subsequent hits minus subsequent misses) and subsequent false memories (i.e., subsequent false alarms minus subsequent correct rejections) separately for each stimulus onset. A minimum of eight trials per trial type was used as a cutoff for inclusion in the analysis. There were two primary analyses. First we conducted a 2 (trial phase: phase 1, phase 2) × $2$ (memory condition: true, false) ANOVA. The main interest in the trial phase × memory condition interaction was to determine the presence of variation in the magnitude of trial phase effects by memory condition (i.e., quantitative or non–cross-over interactions), rather than completely reverse effects (i.e., qualitative or cross-over interaction). Thus, here we focused on quantitative interactions by weighting the interaction effects in ANOVA, as implemented in SPM8, to examine differences in magnitude across levels of trial phase for true and false subsequent memories.

Second, we examined the influence of trial-by-trial variation in the quality of reactivation on neural activity by using a parametric modulation analysis on the reweighting in two models: (i) irrespective of subsequent memory performance and condition or (ii) differentiating reactivating sensitive neural activity for subsequent hits and subsequent false alarms. Additionally, we conducted ancillary parametric modulation analyses to examine neural recruitment sensitive to reaction time and relatedness ratings irrespective of memory performance, and a separate analysis that differentiated subsequent hits and subsequent false alarms for relatedness ratings. A threshold of $P < 0.001$, uncorrected, with a cluster size of 10 voxels was used for all reported analyses. Region of interest analyses for reporting percent signal change were performed in MarsBaR (version 0.43) by extracting all significantly active voxels in an 8-mm sphere centered on the coordinate from the relevant contrast.

**ACKNOWLEDGMENTS.** We thank Clifford A. Robbins, Annie Mitran, and Justin Kim for helpful assistance and the Harvard Museum of Natural History and Peabody Museum of Archaeology and Ethnology at Harvard University for cooperation. This work was supported by National Institute of Health Grant MH060941 (to D.L.S.) and by National Institute on Aging Grant National Research Service Award AG038079 and a L’Oréal USA for Women in Science Fellowship (both to P.L.S.J.).