

Brief Communication

Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms

Craig E.L. Stark,^{1,4} Yoko Okado,² and Elizabeth F. Loftus³

¹Department of Neurobiology and Behavior, Center for the Neurobiology of Learning and Memory, University of California at Irvine, Irvine, California 92629, USA; ²Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, Maryland 21218, USA; ³Department of Psychology and Social Behavior, Criminology, Law and Society, Cognitive Sciences, School of Law, University of California at Irvine, Irvine, California 92629, USA

Many current theories of false memories propose that, when we retrieve a memory, we are not reactivating a veridical, fixed representation of a past event, but are rather reactivating incomplete fragments that may be accurate or distorted and may have arisen from other events. By presenting the two phases of the misinformation paradigm in different modalities, we could observe sensory reactivation of the auditory and visual cortex during the retrieval phase. Overall, true and false memories showed similar brain activation, but could be distinguished by this reactivation. This was true only in the early regions of the sensory cortex.

False memories can arise from misattribution, suggestion, or bias, and are a normal part of everyday memory (Schacter 1999). Examining these memory distortions can help us better understand normal memory and its reconstructive nature. False memories, although inaccurate, are often just as compelling as true memories. Subjective responses such as high confidence (Loftus et al. 1995; Roediger and McDermott 1995), fast reaction times (Loftus et al. 1989), resilience to warnings (McDermott and Roediger 1998), and claims that the memories are detailed recollections rather than simply familiar (Schacter et al. 1998) are remarkably similar for true and false memories. Behavioral evidence suggests, however, that true memories can be accompanied by the retrieval of greater sensory details than false memories (Johnson and Raye 1981; Norman and Schacter 1997). Neuroimaging studies have begun to investigate the similarities and differences between true and false memories, with numerous studies observing differences in the sensory signatures for true and false memories (Schacter et al. 1996, 1997; Gonsalves and Paller 2000; Cabeza et al. 2001; Fabiani et al. 2001; Okado and Stark 2003; Slotnick and Schacter 2004, 2006). Notably, as Slotnick and Schacter (2004, 2006) have argued, it may well be the case that early sensory regions exhibit differences in activity for true versus false memories, but that any information carried by this difference is not available to conscious recollection, and therefore is not a basis for explicit memory judgments.

In previous work (Okado and Stark 2005), we used the misinformation paradigm to create false episodic memories in a setting that approximates real-world scenarios (Loftus et al. 1978). Participants see a vignette and are later presented with new information about the vignette that contains misinformation. For example, they might see a man put a stolen wallet in his jacket pocket, but later be exposed to the false detail that he placed it in his pants pocket. Previous research demonstrated a role for successful versus unsuccessful encoding of the two separate presentations in determining whether a true or false memory would result

(Okado and Stark 2005). Consistent with the general memory literature, the activity in the medial temporal lobe during each of these encoding phases predicted whether the information from that phase would be remembered and incorporated into the memory for the vignette.

Here, we examined the retrieval phase, as the neural correlates of true and false retrieval in the misinformation task has not been explored. In so doing, we tested several aspects of the hypothesis put forth by Slotnick and Schacter (2004, 2006) that false memories would be dissociable from true memories at earlier but not later stages in perceptual processing, with early processing areas reflecting sensory traces for perceived events and later stages correlating with their conscious judgments.

To address this, we used the well-studied sensory reactivation phenomenon (Nyberg et al. 2000; Wheeler et al. 2000; Vaidya et al. 2002; Wheeler and Buckner 2003). In typical studies, words are presented at study with either pictures or of sounds associated with the object. At test, when the word is presented for simple recognition or to recall whether a picture or sound was shown, fMRI activity increases in the visual or auditory cortex, respectively, reactivating the regions used during encoding.

In the present experiment, 25 participants (age 18–40, right-handed, neurologically intact, fluent in English) were recruited from the Johns Hopkins University community and compensated for their time. The experiment consisted of four phases. In the first phase (Original Event; outside the scanner), each participant was shown 11 unique vignettes (see Okado and Stark 2005), composed of 50 color images (3.5-sec duration, 0.5-sec ITI). Each vignette contained 12 critical slides that were later to be inaccurately described in the misinformation phase and six neutral slides that were later to be vaguely described in the misinformation phase. Two versions of each critical slide were developed to allow for appropriate counterbalancing. A day later, in the second phase (Misinformation), participants heard a recorded narrative that they were led to believe was an accurate description of the events seen the day before (one ~3-sec sentence per slide). The majority of the sentences had descriptions that were entirely consistent with the corresponding slide. For the 12 critical slides in each vignette, however, the auditory descriptions were inaccurate.

⁴Corresponding author.

E-mail cestark@uci.edu; **fax** (949) 824-2447.

Article is online at <http://www.learnmem.org/cgi/doi/10.1101/lm.1845710>.

For all six neutral slides in each vignette, the auditory descriptions were vague and not in full detail.

Approximately 15 min later, the participants were placed into the MRI scanner for the third phase, consisting of 24 recognition memory probes for each vignette. Here, a brief sentence was presented visually (7 sec, 0.5-sec ITI) that described a scene, and participants were asked: "Did you see this happening in the original event?" Yes/no responses were recorded with a response box. The 24 questions consisted of 12 directed at the critical events (seven describing the Misinformation version and five describing the Original Event version to slightly increase our opportunity to observe false memories), six directed at the consistent events (three target and three foil versions), and six directed at the neutral targets (three target and three foil versions).

The fourth phase consisted of a surprise source memory test administered outside the scanner to give us a robust assessment of true and false memories. Participants were shown the recognition memory question again, along with their answer, and asked what the source was for their prior recognition memory response. Five options were given: "Remembered seeing it in the slides," "Remembered it was in both the slides and the narrative," "Remembered hearing it in the narrative," "Remembered seeing it in the slides but hearing a different description in the narrative," and "Could not remember and guessed."

The MRI data was collected on a Philips Gyroscan 3T MRI scanner using a whole-brain SENSE coil (SENSitivity Encoding) that reduces acquisition time and distortion due to magnetic susceptibility (Pruessmann et al. 1999). Thirty-five axial T2*-weighted echoplanar functional images were collected per volume (AC-PC alignment, TE = 30 msec, flip angle = 75°, in-plane resolution = 3 × 3 mm, thickness = 3 mm plus a 1-mm interslice gap, TR = 1.875 sec, four TR initial skip, 1056 vol in total). After the functional scans, a whole-brain 1-mm structural scan was acquired for anatomical localization.

fMRI data were analyzed using AFNI (Cox 1996) by first co-registering all volumes in 3D. Six vectors coding for the motion were used to censor time points corrupted by significant motion (>3° or 2 mm) from additional analyses. fMRI data from the 11 runs were concatenated and subjected to both whole-brain and restricted medial temporal lobe (MTL) analyses. In both, data were analyzed using a deconvolution technique based on multiple linear regression (3dDeconvolve) using behavioral vectors coding for the trial type. The sum of the resulting β coefficients corresponding to the expected hemodynamic response region (~2–11 sec) was taken as the model's estimate of the response to each trial type. Initial whole-brain spatial normalization of the structural and statistical data was done to the atlas of

Talairach and Tournoux (1988), and additional fine-tuning for the MTL analysis was done using ROI-LDDMM (Miller et al. 2005). A 4-mm Gaussian blur was imposed and the data were resampled to 2.5-mm isotropic voxels in the process prior to group analyses. Both voxel-wise and cluster-size thresholds were used to establish a final α level of $P < 0.05$ for the two analyses (MTL analysis corrected for the smaller volume), correcting for multiple comparisons by using Monte Carlo simulation (AlphaSim).

Recognition memory performance is shown in Figure 1A, and demonstrates a clear misinformation effect. The Critical Hit rate (accurately responding "yes" to the recognition memory probe for a critical item when the Original Event version was shown) was 0.61 ± 0.02 . Of key interest are the Critical False Alarm (FA) rate (falsely responding "yes" to the recognition memory probe for a critical item when the Misinformation version was shown) and the Neutral FA rate (falsely responding "yes" to the recognition memory probe for an entirely unstudied neutral foil item). The Critical FA rate of 0.46 ± 0.03 was significantly higher than the Neutral False Alarm rate of 0.27 ± 0.02 ($t_{(24)} = 10.6$, $P < 0.001$), demonstrating a misinformation effect. Notably, it was also significantly higher than the 0.29 False Alarm rate to Consistent trials ($t_{(24)} = 6.8$; $P < 0.001$).

Data from the source memory task (Fig. 1B) were used to filter trials and isolate robust true and false memories (Zhu et al. 2010) by including only those Critical Hit (true) and FA (false) trials that were subsequently classified with the source response of "saw it in the slides." This was done to remove guesses and to provide a conservative way of characterizing true and false memories. It also gives us the cleanest contrast between them, as, in both conditions, participants only remember (or believe to remember) seeing the item. Fifty-five percent of the Critical Hits and 42% of the Critical FAs were so classified. These trials, believed to be the best representation of true and false memories, were used in the True versus False memory contrast in the fMRI data analysis. Reaction times to these two key trial types did not differ ($P > 0.2$). However, given this filtering, only 16 participants had sufficient trials for a meaningful analysis of the fMRI data (at least 10 trials in each condition) in this contrast.

The fMRI signals for these two trial types were contrasted to identify differences between true and false recognition (Fig. 2). True memory was associated with greater activity than false memory in ostensibly early visual cortical regions (BA 17/18), including the striate cortex. One area included the left lingual gyrus and the cuneus and a second was in the cuneus. Additional activity was observed in the left inferior parietal lobule (BA 40), the left cingulate (BA 31), and the anterior cingulate. In contrast, false memory was associated with increased activity only in the left superior temporal gyrus (BA 22/42) of the auditory cortex. These results demonstrate that the manipulation of the misinformation paradigm was successful in engaging distinct sensory reactivation at time of retrieval for true and false recognition. True recollections based on visual information reactivated the visual cortex, whereas false recollections based on auditory information reactivated the auditory cortex. In addition, true recognition recruited more regions compared with false recognition, which may be indicative of more information recovery and processing associated with true memories when compared with false memories. Importantly, these results are consistent with the hypothesis put forth by Slotnick and Schacter (2004, 2006)

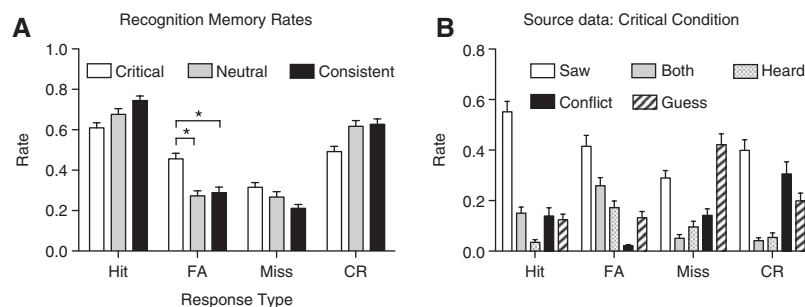


Figure 1. (A) Behavioral data from the recognition memory phase for all three trial types and for each response category. Rates shown are out of the total number of trials and do not factor out the 6.6% "no response" rate. Significance in the key comparison to demonstrate a false memory is indicated with an asterisk. (B) Source memory response rates for the Critical condition. Error bars indicate the standard error of the mean.

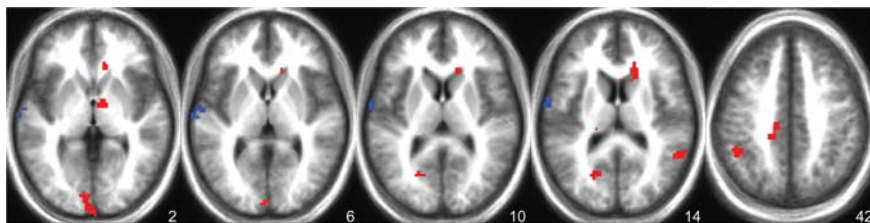


Figure 2. Regions showing a difference in activity for true versus false memory (red = $T > F$; blue = $F > T$) during recognition memory testing. In the paradigm, true memories were retrievals of information that had been encoded visually, and false memories were of information that had been encoded auditorially. Numbers on each image indicate the inferior–superior location of the slice in Talairach coordinates, and the *left* side of the image corresponds to the left side of brain.

that, at least for visual regions, it is only the early regions (BA 17/18) that distinguish true from false memories. We should note that these differences are based on averaging across multiple trials and participants, and that this does not imply that this could be used to distinguish true from false memories at finer levels of analysis.

Figure 3 shows the result of contrasts between both true and false memories and a common baseline. In the first (Fig. 3, top), Critical Hits were contrasted with Critical Correct Rejections (CR; filtered to remove guesses). Both represent accurate performance with no difference in reaction time ($P > 0.05$). Critical Hits were tied to correctly endorsing viewed items, and Critical CRs were tied to correctly identifying items that were only narrated. According to the sensory reactivation hypothesis, this should elicit activity in the visual cortex, as participants retrieve the image during Critical Hits but not during Critical CRs. Critical Hits were associated with increased activity in both early and later portions of the visual cortex. One region was located in the left lingual gyrus (BA 18, but outside of the cuneus/BA 17), and a second was located in the left middle occipital gyrus and middle temporal gyrus (BA 19 and BA 39). Thus, true recollections of visually based events were accompanied by greater visual activity compared with correct rejections of false, auditory-based events. In addition, frontal regions—including the anterior cingulate, left middle frontal gyrus, and right medial frontal gyrus—also showed significant differences, indicative perhaps of post-retrieval

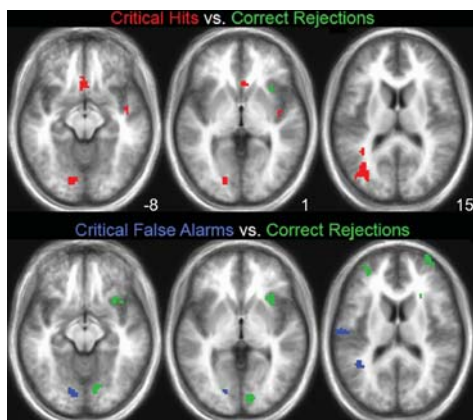


Figure 3. Regions exhibiting significantly different activity for Critical Hits (red) versus Critical Correct Rejections (green) are shown in the *top* panel ($H > CR = \text{red}$; $CR > H = \text{green}$). The *bottom* panel shows the corresponding contrast between Critical False Alarms (blue) and Critical Correct Rejections. Numbers on each image indicate the inferior–superior location of the slice in Talairach coordinates, and the *left* side of the image corresponds to the left side of brain.

monitoring and evaluation of recovered information associated with true recognition (e.g., Hayama and Rugg 2009). In the MTL analysis, two MTL regions (the right hippocampus and left parahippocampal cortex) were also more engaged for Critical Hits compared with Critical CRs (consistent with numerous studies in the literature) (for review, see Wais 2008). The reverse of this contrast showed greater activity for Critical CRs than Critical Hits in frontal regions (medial frontal gyrus and right superior frontal gyrus) and in the right inferior parietal lobe. These results raise the possibility of greater effortful post-retrieval processing in these regions associated with correct rejection of false, auditory-based events compared with true recognition.

When Critical FAs were contrasted with Critical CRs (Fig. 3, bottom), greater activity for the Critical FAs was observed in the auditory cortex, including the left superior temporal gyrus (BA 41/42/43 and BA 22). Thus, false recollections of auditory-based events were accompanied by greater auditory activity compared with correct rejections of these same events. Activity in the right insula (BA 13) was observed as well. Similar to what we observed in the Critical Hits versus CR contrast, we again observed activity in the early visual cortex (just outside of the cuneus, BA17) within the left lingual gyrus (BA 18). Here, while inaccurate, participants believed they had seen a picture of the event. In the MTL analysis, the right hippocampus and left parahippocampal cortex were also more active for Critical FAs compared with Critical CRs. Thus, falsely retrieving a memory resulted in more MTL activity than an accurate correct rejection. As this contrast did not yield activity in the overall true versus false contrast (in which both conditions elicited an “old” response), concluding that the MTL differentiated true from false memories is not justified. Its activity was elevated when the participant mistakenly believed an item had been seen previously relative to when the participant believed the item to be entirely new or to have been only heard. We should note that several studies have observed MTL activity correlated with the actual old–new status rather than the perceived old–new status in even standard recognition memory tasks (Daselaar et al. 2006; Kirwan et al. 2009). In these, standard recognition memory tasks were used, and it is quite possible that the false alarms generated in these tasks are qualitatively different from the false alarms generated using the misinformation paradigm. If the false memories here are indeed richer than those in simple list-learning paradigms (as one might expect), and if true memories are richer than the false alarms used in these (e.g., confidence ratings in Kirwan et al. 2009 showed very few high-confidence responses for false alarms and many for hits), we might account for this discrepancy.

The reverse of this contrast showed significantly elevated activity for the Critical CRs relative to the Critical FAs in a corresponding area of the right visual cortex (lingual gyrus). The reason for the inversion of the effect in the right hemisphere is not entirely clear. One possibility is that, for CRs, the actual study item was retrieved (and judged to be only similar to the probe), and that this is richer (and hence more active) than for FAs (e.g., Gonsalves and Paller 2000; Okado and Stark 2003). Alternatively, it is always possible that it is less related to the visual information per se, but driven more by the other aspects of the complex set of processes involved during challenging retrieval tasks (Gallo 2006). Frontal regions—including the bilateral superior and middle frontal gyri, right inferior frontal gyrus, right medial frontal gyrus, and left anterior cingulate—also showed

significant differences, as did the right inferior parietal lobe. These results raise the possibility of greater effortful post-retrieval processing in these regions associated with correct rejection of false, auditory-based events compared with false recollections of these same events. Caution is warranted here, given the general problem of reverse inferences made on fMRI data (Poldrack 2008).

Together, the contrasts with the baseline (Critical CR) condition are consistent with the observation by Slotnick and Schacter (2004, 2006) that activity for both true and false memories differs from some form of baseline in early and later portions of visual cortex. Here, the early region did not extend into the striate cortex, but was somewhat more anterior (BA 18). Our results extend their work, demonstrating the auditory cortex in what is likely both early (BA 41) and later (BA 22, BA 42, and BA 43) processing regions. As noted above, when comparing true and false memories directly, our results are also consistent with their work, demonstrating that in, the visual cortex, only the early regions (BA 17/18) distinguished true from false memories.

Overall, the results are consistent with the general view that memories are reconstructed by fragments of previously encoded pieces of information that are reactivated at the time of retrieval. The auditory cortex was reactivated during the retrieval of information encoded auditorally, even when this was falsely being judged as coming from a visually presented source. In the visual cortex, earlier, more posterior regions differentiated true from false, but later, more anterior regions did not. This report is the first time this finding has been observed in a misinformation task, a more real-world (albeit less experimentally tractable) method of investigating false memories.

Acknowledgments

We would like to acknowledge NSF BCS0544959 for support of this project and Monica Lopez-Gonzalez for assistance with data collection.

References

Cabeza R, Rao SM, Wagner AD, Mayer AR, Schacter DL. 2001. Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proc Natl Acad Sci* **98**: 4805–4810.

Cox RW. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* **29**: 162–173.

Daselaar SM, Fleck MS, Prince SE, Cabeza R. 2006. The medial temporal lobe distinguishes old from new independently of consciousness. *J Neurosci* **26**: 5835–5839.

Fabiani M, Stadler MA, Wessels PM. 2001. True but not false memories produce a sensory signature in human lateralized brain potentials. *J Cogn Neurosci* **12**: 941–949.

Gallo DA. 2006. *Associative illusions of memory*. Psychology Press, New York.

Gonzales B, Paller KA. 2000. Neural events that underlie remembering something that never happened. *Nat Neurosci* **3**: 1316–1321.

Hayama HR, Rugg MD. 2009. Right dorsolateral prefrontal cortex is engaged during post-retrieval processing of both episodic and semantic information. *Neuropsychologia* **47**: 2409–2416.

Johnson MK, Raye CL. 1981. Reality monitoring. *Psychol Rev* **88**: 67–85.

Kirwan CB, Shragar Y, Squire L. 2009. Medial temporal lobe activity can distinguish between old and new stimuli independently of overt behavioral choice. *Proc Natl Acad Sci* **106**: 14617–14621.

Loftus EF, Miller DG, Burns HJ. 1978. Semantic integration of verbal information into a visual memory. *J Exp Psychol Hum Learn* **4**: 19–31.

Loftus EF, Donders K, Hoffman HG, Schooler JW. 1989. Creating new memories that are quickly accessed and confidently held. *Mem Cognit* **17**: 607–616.

Loftus EF, Feldman J, Dashiell R. 1995. The reality of illusory memories. In *Memory distortion: How minds, brains, and societies reconstruct the past*, pp. 47–68. Harvard University Press, Cambridge, MA.

McDermott KB, Roediger HL. 1998. Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *J Mem Lang* **39**: 508–520.

Miller MI, Beg MF, Ceritoglu C, Stark CEL. 2005. Increasing the power of functional maps of the medial temporal lobe using large deformation metric mapping. *Proc Natl Acad Sci* **102**: 9685–9690.

Norman KA, Schacter DL. 1997. False recognition in young and older adults: Exploring the characteristics of illusory memories. *Mem Cognit* **25**: 838–848.

Nyberg L, Habib R, McIntosh AR, Tulving E. 2000. Reactivation of encoding-related brain activity during memory retrieval. *Proc Natl Acad Sci* **97**: 11120–11124.

Okado Y, Stark CEL. 2003. Neural processing associated with true and false memory retrieval. *Cogn Affect Behav Neurosci* **3**: 323–334.

Okado Y, Stark CEL. 2005. Neural activity during encoding predicts false memories created by misinformation. *Learn Mem* **12**: 3–11.

Poldrack RA. 2008. The role of fMRI in cognitive neuroscience: Where do we stand? *Curr Opin Neurobiol* **18**: 223–227.

Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. 1999. SENSE: Sensitivity encoding for fast MRI. *Magn Reson Med* **42**: 952–962.

Roediger HL, McDermott KB. 1995. Creating false memories: Remembering words not presented in lists. *J Exp Psychol Learn Mem Cogn* **12**: 803–814.

Schacter DL. 1999. The seven sins of memory. *Am Psychol* **54**: 182–203.

Schacter DL, Reiman E, Curran T, Yun LS, Bandy D, McDermott KB, Roediger HL. 1996. Neuroanatomical correlates of veridical and illusory recognition memory: Evidence from positron emission tomography. *Neuron* **17**: 267–274.

Schacter DL, Buckner RL, Koutstaal W, Dale AM, Rosen BR. 1997. Late onset of anterior prefrontal activity during true and false recognition: An event-related fMRI study. *NeuroImage* **6**: 259–269.

Schacter DL, Norman KA, Koutstaal W. 1998. The cognitive neuroscience of constructive memory. *Annu Rev Psychol* **49**: 289–318.

Slotnick SD, Schacter DL. 2004. A sensory signature that distinguishes true from false memories. *Nat Neurosci* **7**: 664–672.

Slotnick SD, Schacter DL. 2006. The nature of memory related activity in early visual areas. *Neuropsychologia* **44**: 2874–2886.

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

Vaidya CJ, Zhao M, Desmond JE, Gabrieli JD. 2002. Evidence for cortical encoding specificity in episodic memory: Memory-induced re-activation of picture processing areas. *Neuropsychologia* **40**: 2136–2143.

Wais PE. 2008. fMRI signals associated with memory strength in the medial temporal lobes: A meta-analysis. *Neuropsychologia* **26**: 3185–3196.

Wheeler ME, Buckner RL. 2003. Functional dissociation among components of remembering: Control, perceived oldness, and content. *J Neurosci* **23**: 3869–3880.

Wheeler ME, Petersen SE, Buckner RL. 2000. Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci* **97**: 11125–11129.

Zhu B, Chen C, Loftus EF, Lin C, He Q, Chen C, Moyzis RK, Lessard J, Dong Q. 2010. Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Pers Individ Dif* **48**: 889–894.

Received April 21, 2010; accepted in revised form July 27, 2010.



Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigms

Craig E.L. Stark, Yoko Okado and Elizabeth F. Loftus

Learn. Mem. 2010, **17**:

Access the most recent version at doi:[10.1101/lm.1845710](https://doi.org/10.1101/lm.1845710)

References

This article cites 30 articles, 8 of which can be accessed free at:
<http://learnmem.cshlp.org/content/17/10/485.full.html#ref-list-1>

License

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](#).
