

True and false memories, parietal cortex, and confidence judgments

Zhisen J. Urgolites,^{1,2} Christine N. Smith,^{1,2} and Larry R. Squire^{1,2,3,4}

¹Veterans Affairs San Diego Healthcare System, San Diego, California 92161, USA; ²Department of Psychiatry; ³Department of Neurosciences; ⁴Department of Psychology, University of California, San Diego, La Jolla, California 92093, USA

Recent studies have asked whether activity in the medial temporal lobe (MTL) and the neocortex can distinguish true memory from false memory. A frequent complication has been that the confidence associated with correct memory judgments (true memory) is typically higher than the confidence associated with incorrect memory judgments (false memory). Accordingly, it has often been difficult to know whether a finding is related to memory confidence or memory accuracy. In the current study, participants made recognition memory judgments with confidence ratings in response to previously studied scenes and novel scenes. The left hippocampus and 16 other brain regions distinguished true and false memories when confidence ratings were different for the two conditions. Only three regions (all in the parietal cortex) distinguished true and false memories when confidence ratings were equated. These findings illustrate the utility of taking confidence ratings into account when identifying brain regions associated with true and false memories. Neural correlates of true and false memories are most easily interpreted when confidence ratings are similar for the two kinds of memories.

The medial temporal lobe (MTL) plays an essential role in memory formation and in its retrieval for a time after learning (Squire and Zola-Morgan 1991; Eichenbaum and Cohen 2001). Recent studies have asked whether the MTL is differentially involved when memory judgments are accurate versus when they are inaccurate. In the first case, individuals believe their judgments to be correct, and they are right (i.e., hits, true memory). In the second case, individuals also believe their judgments to be correct, but they are wrong (i.e., false alarms, false memory). A few studies have found that the MTL can distinguish true from false memory, but most studies have not obtained this finding (for review, see Dennis et al. 2014). In the neocortex, regions in the parietal and prefrontal cortex have been reported to distinguish true from false memory (for review, see Schacter and Slotnick 2004).

A complication in many of these studies is that accurate memory judgments are ordinarily associated with higher confidence than inaccurate memory judgments. Yet, with rare exception (Kim and Cabeza 2007), studies investigating true and false memories have not taken confidence ratings into account. As a result, it has been unclear whether brain regions detected in these studies reflect a difference between high-confidence and low-confidence judgments or a difference between true and false memories.

Kim and Cabeza (2007) compared true and false memories separately for trials associated with high confidence and trials associated with low confidence. They found that regions in the MTL distinguished high-confidence true and false memories but not low-confidence true and false memories. This study and most others (but see Kirwan et al. 2009) that contrasted true and false memories used paradigms where the foils were either conceptually or perceptually related to the targets (e.g., Deese 1959; Roediger and McDermott 1995). These procedures increase the frequency of false memories. As it stands, little is known about the brain regions associated with true and false memories when confidence

ratings are taken into account and when the foils are unrelated to the targets. Note that false alarms made to foils that are unrelated to targets presumably reflect a different set of real-world conditions than false alarms made to foils that are related to targets.

We asked participants to make recognition memory judgments with confidence ratings in response to 240 previously studied scenes intermixed with an equal number of novel scenes (Fig. 1). The scenes were selected so that they were as distinct as possible from one another. We equated confidence ratings for true and false memory judgments and then looked for brain regions in the MTL and in the neocortex that distinguished true and false memories as well as brain regions that responded similarly for true and false memories.

Results

Behavioral performance

Participants scored well above chance on the recognition memory test (accuracy = hit rate/(hit rate + false alarm rate) = 79.8 ± 2.0% correct, $d' = 1.5 \pm 0.1$, $P_s < 0.001$). For hits, 14.1%, 19.0%, and 66.9% of the trials were given a confidence rating of 4, 5, and 6, respectively. For false alarms, 48.2%, 37.2%, and 14.5% of the trials were given a confidence rating of 4, 5, and 6, respectively.

fMRI analysis

The following analyses were carried out separately for voxels in the MTL and voxels in the whole brain (excluding the MTL). We first identified areas that distinguished true memory (hits) from false memory (false alarms). We noted, as is typically the case,

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first 12 months after the full-issue publication date (see <http://learnmem.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Corresponding author: lsquire@ucsd.edu

Article is online at <http://www.learnmem.org/cgi/doi/10.1101/lm.038349.115>.

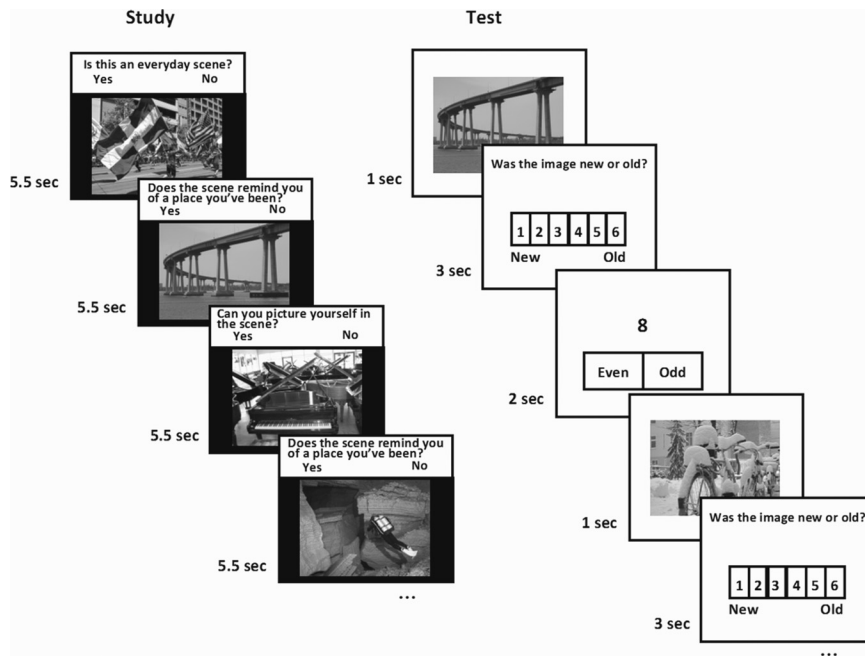


Figure 1. Task design. Participants studied a different set of 80 color photographs of scenes at each of four intervals before scanning (320 total). In the scanner, participants made recognition memory judgments with confidence ratings (1–6) in response to scenes (240 studied scenes intermixed with 240 novel scenes) and made odd/even judgments in response to digits (baseline trials). Scene trials were interleaved with 0–7 baseline trials.

that hits were associated with higher confidence ratings than false alarms (5.5 versus 4.6, $P < 0.001$). Accordingly, we contrasted hits and false alarms in two different ways. First, we contrasted activity in response to hits with activity in response to false alarms without taking account of confidence ratings. One MTL cluster in the left hippocampus ($X = -25.3$, $Y = -32.0$, $Z = -5.0$; 783 mm^3), 12 neocortical clusters (in prefrontal, temporal, parietal, and occipital cortex), and one cluster in the caudate nucleus exhibited activity that was greater for hits than for false alarms. Three additional neocortical clusters (in prefrontal cortex) exhibited activity that was greater for false alarms than for hits. These findings could have reflected a difference between true and false memories, or they could have reflected the difference between high and low confidence.

To test if activity in these 17 clusters was related to the level of confidence, we carried out a regression analysis to examine activity in each of these regions as a function of confidence level. We restricted these analyses to items judged to be old (i.e., for confi-

dence ratings 4, 5, and 6), irrespective of whether judgments were accurate or inaccurate. This analysis revealed that activity in all 17 regions was related to confidence ratings. Specifically, the 14 clusters that exhibited higher activity for hits than for false alarms exhibited activity that was positively related to confidence ratings ($P < 0.01$), and the three clusters in prefrontal cortex that exhibited higher activity for false alarms than for hits exhibited activity that was negatively related to confidence ratings ($P < 0.01$).

We next contrasted activity in response to hits with activity in response to false alarms when confidence ratings were matched for these two conditions. Separate analyses were carried out for high-confidence trials and low-confidence trials. When confidence ratings were high (and equated) for both true and false memories (confidence ratings = 5.2 and 5.2, respectively; accuracy = $74.9 \pm 2.2\%$ correct), only three clusters were identified, two in anterior parietal cortex bilaterally and one in right medial parietal cortex (Table 1; Fig. 2). Two of these three clusters overlapped minimally with two clusters in the original group of 17 clusters identified when confidence was

different for hits and false alarms (see clusters 3 and 4 in Table 1 and Fig. 2). All three clusters exhibited higher activity for true memory than false memory. No clusters were identified when confidence ratings were low (and equated) for true and false memories (confidence ratings = 4 and 4, respectively; accuracy = $45.7 \pm 2.4\%$ correct), probably because performance was no better than chance in this condition ($P = 0.10$).

Because two earlier studies (Kim and Cabeza 2007; Dennis et al. 2012) found that MTL activity distinguished true and false memories when strong memories were examined, we explored the possibility that clusters in the MTL might be observed if the probability threshold were more lenient. Accordingly, we compared true and false memories for high-confidence trials using a more lenient voxel-wise threshold ($P < 0.05$). With this more lenient probability threshold, we identified a cluster in the left hippocampus (540 mm^3 , uncorrected) where activity was higher for hits than for false alarms (Table 1; Fig. 2). This sub-threshold cluster partially overlapped with the hippocampal cluster identified

Table 1. Brain regions that differentiated true and false memories when confidence was equated

Brain region	Talairach coordinates			Volume (mm^3)	BA	Cohen's <i>d</i>
	LR	AP	IS			
True memory > false memory						
L hippocampus (1)	-28.6	-31.1	-4.2	540		0.53 ^a
R inferior parietal lobule (2)	38.7	-38.7	43.6	1215	40	0.67 ^b
L inferior/superior parietal lobules (3)	-39.7	-50.1	43.5	3294	7/40	0.74 ^c
R precuneus (4)	16.1	-51.8	32.1	1350	7	0.69 ^b

These clusters were identified by a contrast between true memory (hits) and false memory (false alarms) when confidence ratings associated with hits and false alarms were equated. Cluster numbers correspond to cluster numbers in Figure 2. Cohen's *d* reflects the difference in means, measured in standard deviations.

^aA sub-threshold cluster was detected only by using a voxel-wise threshold of $P < 0.05$, uncorrected.

^bCluster-wise $P < 0.01$.

^cCluster-wise $P < 0.001$.

(LR) left/right, (AP) anterior/posterior, (IS) inferior/superior, (BA) Brodmann area.

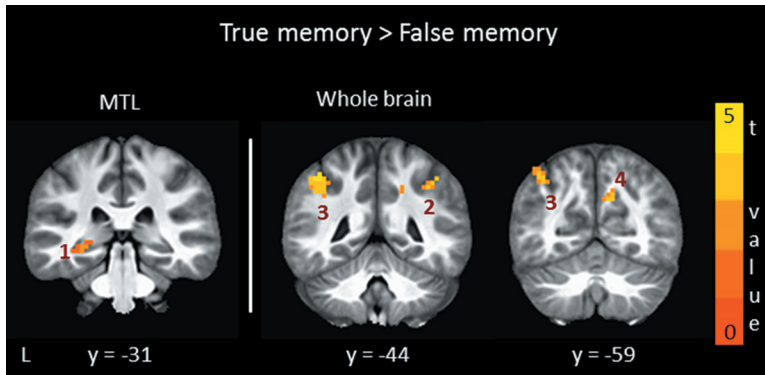


Figure 2. Brain regions in parietal cortex and MTL where activity was higher for true memory (hits) than false memory (false alarms) and confidence ratings were similar for these conditions (also see Table 1). Cluster numbers correspond to cluster numbers in Table 1. Note that the finding for the MTL cluster was sub-threshold (voxel-wise $P < 0.05$, uncorrected). L, left; y, anterior–posterior Talairach coordinate for each coronal section.

before confidence was equated. No clusters were identified in the MTL for low-confidence trials.

Next, we identified regions where activity was associated with both true memory (hits > correct rejections) and false memory (false alarms > correct rejections). Activity in these regions was related to a participant's judgment that the scene had been studied, regardless whether the judgment was correct or incorrect. To identify these regions, we first carried out an analysis without taking account of confidence ratings. No clusters were identified in the MTL, and one cluster was identified in left posterior parietal cortex.

We then asked where activity was similar for hits and false alarms when the confidence ratings associated with these two conditions were equated. We carried out analyses separately for high-confidence trials and low-confidence trials. No clusters were identified in the analysis of the low-confidence trials. For the high-confidence trials, two clusters in posterior parietal cortex (and no clusters in the MTL) were identified (Table 2; Fig. 3). One cluster (cluster 2 in Table 2 and Fig. 3) was new and independent of the cluster identified before confidence was equated. The other cluster (cluster 1 in Table 2 and Fig. 3) was adjacent to and partially overlapped with the cluster that had been identified before equating confidence ratings.

Table 2. Brain regions that exhibited similar levels of activity for true and false memories when confidence was equated

Brain region	Talairach coordinates			Volume (mm ³)	BA	Cohen's <i>d</i>
	LR	AP	IS			
L posterior cingulate cortex/precuneus (1)	-11.5	-52.8	18.5	1215	31	0.65 ^a
L precuneus (2)	-30.5	-73.2	36.4	1269	19	0.73 ^a

These clusters were identified by the analysis carried out when the confidence associated with hits and false alarms was equated. The analysis combined two single contrasts (hits > correct rejections and false alarms > correct rejections) that were carried out at the high-confidence level. Cluster numbers correspond to cluster numbers in Figure 3. Cohen's *d* reflects the average Cohen's *d* for the two contrasts that comprised the conjunction analysis, i.e., hits > correct rejections and false alarms > correct rejections.

^aCluster-wise $P < 0.01$.

(LR) left/right, (AP) anterior/posterior, (IS) inferior/superior, (BA) Brodmann area.

Discussion

Participants made recognition memory judgments in the fMRI scanner for previously studied scenes and novel scenes. We first identified brain areas that distinguished true memory (i.e., hits) and false memory (i.e., false alarms) without taking confidence ratings into account. This analysis identified the left hippocampus, 15 neocortical regions, and the caudate nucleus. Activity in all of these regions was linearly related to confidence levels. Thus, these regions likely distinguished high and low confidence rather than true and false memories per se.

When we equated the confidence ratings associated with true and false memories and repeated the same analysis, only three regions (all in parietal cortex) distinguished true and false

memories. Two of these regions minimally overlapped with regions identified when confidence was not equated, and one region was new. In addition, there was one sub-threshold MTL cluster (Table 1; Fig. 2). Thus, brain activity can distinguish true memory and false memory even when the two kinds of memory judgments are made with similar levels of confidence.

Last, we identified regions where activity was similar for true memory and false memory (and also higher than for correct rejections). One cortical region (and no MTL regions) exhibited similar activity for true memory and false memory before confidence was taken into account. When confidence was equated for hits and false alarms (and correct rejections), two regions in posterior parietal cortex exhibited similar activity for true and false memories (Table 2; Fig. 3). These included one new region and one region that partially overlapped with the region that had been identified before equating for confidence. Accordingly, brain activity in these two regions was related to the judgment that the scene had been presented previously, regardless whether the judgment was correct or incorrect.

The hippocampus and true and false memories

Hippocampal activity differentiated true and false memories when confidence ratings were different for the two conditions. This finding is consistent with the results of the only other study that compared hits and false alarms and where targets and foils were unrelated (Kirwan et al. 2009). In that study, hits were associated with higher confidence than false alarms, and the hippocampus exhibited higher activity for hits than for false alarms.

Other findings also indicate that hippocampal activity can differentiate true and false memories when confidence ratings are not controlled, so long as accuracy is relatively high as in our study [79.8% correct; accuracy = hit rate/(hit rate + false alarm rate)]. For example, Gutchess and Schacter (2011) and Kirwan et al. (2009) obtained a hippocampal finding when all hits were contrasted with all false alarms, and memory scores were high (accuracy = 70.7% and 75.4% correct, respectively). In contrast, studies that did not obtain hippocampal findings had lower accuracy rates (Schacter et al. 1996: 54.0% correct; Schacter et al. 1997: 61.1% correct; Cabeza et al. 2001: 52.4% correct; Slotnick and Schacter 2004: 53.4% correct; Garoff-Eaton et al. 2007: 59.3% correct; Lidaka et al. 2012: 54.1% correct).

The relationship between accuracy and the finding that hippocampal activity can differentiate true and false memory is less clear in studies where true or false memory refers to accurate or

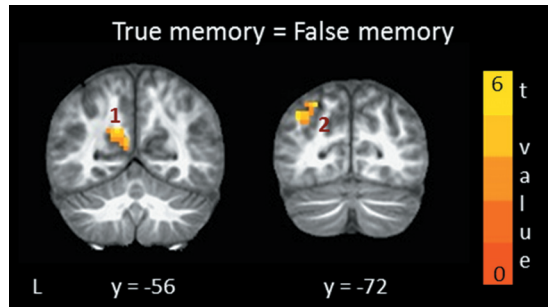


Figure 3. Brain regions in posterior parietal cortex where activity was similar for true memory (hits > correct rejections) and false memory (false alarms > correct rejections) and confidence ratings were similar for these conditions. Cluster numbers correspond to cluster numbers in Table 2. The *t*-value represents the average *t*-value for the two contrasts that comprised the analysis. L, left; y, anterior–posterior Talairach coordinate for each coronal section.

inaccurate retrieval of the context in which material was studied (hippocampal findings: Cansino et al. 2002: 69.8% correct where chance = 25%; Dobbins et al. 2002: hit rate = 80%, but false alarm rate was not available; Weis et al. 2004: 52.1% correct where chance = 25%; no hippocampal findings: Okado and Stark 2003: 78.8% correct; Kahn et al. 2004: 65.3% correct; Stark et al. 2010: 63.5% correct).

When confidence ratings were equated (and accuracy was high, 74.9% correct), hippocampal activity marginally distinguished true and false memories (i.e., voxel-wise probability $P < 0.05$ uncorrected). In contrast, no MTL regions distinguished true and false memories in the case of memory judgments made with low confidence (accuracy = 45.7% correct). These findings for the hippocampus are consistent with results from two other studies. In one study (Kim and Cabeza 2007), the hippocampus distinguished high-confidence true and false memories (Sure-Hits and Sure-False Alarms; accuracy = 71.0% correct) but not low-confidence true and false memories (Unsure-Hits and Unsure-False Alarms; accuracy = 43.5% correct). In the second study (Dennis et al. 2012), hippocampal activity distinguished true and false recollection (using a Remember, Know, New Paradigm; accuracy = 70.6% correct). Recollection is typically associated with high confidence and high accuracy (Wixted 2007).

The neocortex and true and false memories

The neocortex distinguished true and false memories when confidence ratings associated with the two conditions were similar. Specifically, parietal cortex (BA 7/40) exhibited higher activity for hits than for false alarms. The finding is in agreement with earlier work demonstrating a role for this region in the retrieval of contextual information (Curran 2000), a circumstance more likely to occur for true memory than for false memory. Kim and Cabeza (2007) also found regions in the parietal cortex that distinguished true and false memories when confidence was equated, but these regions exhibited higher activity for false memory than for true memory.

Other studies that did not equate confidence found that activity in posterolateral parietal cortex (BA 7/39/40) was higher for true than for false memory (Cabeza et al. 2001; Slotnick and Schacter 2004; Stark et al. 2010). We also detected clusters in the posterolateral parietal cortex for the contrast of hits and false alarms when confidence levels were not equated. Activity in these regions was linearly related to confidence levels. Accordingly, activity in posterolateral parietal cortex is likely related to confidence in the perceived oldness of the stimuli (also see Wheeler

and Buckner 2003, 2004) and not to true and false memories per se. Similarly, the activity we observed in prefrontal cortex (when confidence was not equated) is likely related to confidence and not to true and false memories. This activity was higher for false alarms than for hits (also see Schacter and Slotnick 2004) and was negatively related to confidence ratings.

Last, one region in posterolateral parietal cortex responded similarly for true and false memories before confidence was equated. In contrast, earlier studies that tested for regions that responded similarly for true and false memories (Kahn et al. 2004; Slotnick and Schacter 2004; Garoff-Eaton et al. 2006, 2007; Gutchess and Schacter 2011) found a number of brain regions in frontal, temporal, parietal, and occipital cortex. Note that in these earlier studies, the targets and foils were conceptually or perceptually related. This circumstance may increase the number of regions that respond similarly for true and false memories.

Conclusion

The MTL distinguished true and false memories when the level of confidence was different for each kind of memory. A partially overlapping region in the MTL distinguished true and false memories when confidence was high and also similar for the two conditions, but only at a lower threshold. Many neocortical regions also distinguished true and false memories when the memory judgments differed in confidence. However, only three regions (in parietal cortex) distinguished true and false memories when confidence was equated. In addition, one region responded similarly for true and false memories before confidence was equated. After confidence was equated, one partially overlapping region plus one new region (both in posterior parietal cortex) responded similarly for true and false memories. Our findings underscore the importance of taking confidence ratings into account when identifying brain regions associated with true and false memories. Brain activity associated with true and false memories are most easily interpreted when confidence ratings are equated.

Materials and Methods

Participants

Eighteen participants (10 female; mean age = 29 ± 1.4 yr; range = 21–42 yr) were recruited from the San Diego community.

Materials and procedure

A total of 800 color photographs of indoor/outdoor scenes were used (Fig. 1). The scenes were selected so that they were as distinct as possible from one another in order to minimize the possibility that participants would mistake one scene for another. Participants studied a different set of 80 scenes one month, one week, one day, and one hour before scanning. Data were combined across all study-test delays. (Results examining brain activity as a function of study-test delay will be reported elsewhere). Each scene was presented for 5.5 sec. One of three yes/no questions appeared above each scene to encourage deep encoding: “Is this an everyday scene?,” “Does the scene remind you of a place you have been?,” and “Can you picture yourself in the scene?.” Participants responded by pressing “yes” or “no” on a keyboard. Which 80 scenes were presented in each study session was randomized across participants.

During scanning, 240 studied scenes (60 from each study session) were intermixed with 240 novel scenes and 606 baseline trials (Fig. 1). For scene trials, a scene was presented for 1 sec, and participants then had 3 sec to make a recognition judgment with confidence ratings (1 = definitely new, 2 = probably new, 3 = maybe new, 4 = maybe old, 5 = probably old, and 6 = definitely old). Recognition judgments were made by selecting the appropriate number (1–6) on the screen with a mouse (Current

Designs, Philadelphia, PA). For baseline trials, participants saw a single digit (1–9) and indicated by mouse whether it was odd or even (2 sec) (Stark and Squire 2001). For both the scene and baseline trials, the starting location of the mouse cursor on the screen was randomized. Thus, rightward or leftward movements of the mouse on each trial were uncorrelated with rightward or leftward movements toward the right or left side of the scales (see Fig. 1). The test phase consisted of 6 scan runs. Each run (8.7 min duration; 261 images) consisted of 10 scenes from each of the four study sessions and an equal number of new scenes (80 scenes total). Zero to seven baseline trials (101 trials/run) were presented after each scene trial (mean = 2.5 trials). Each scene was equally likely to be studied before scanning or to serve as a target or a foil during scanning.

fMRI imaging

Imaging was carried out on a 3T General Electric scanner at the Center for Functional MRI (University of California San Diego). Functional images were acquired using a gradient-echo, echo-planar, T2*-weighted pulse sequence (2000 msec TR; 64 × 64 matrix size; 25 cm field of view; 3.9 × 3.9 mm in-plane resolution). The first five images acquired were not analyzed to allow for T1 equilibration. Thirty-six oblique coronal slices (slice thickness = 4.8 mm) were acquired perpendicular to the long axis of the hippocampus and covering the whole brain. Following the six functional runs, high-resolution structural images were acquired using a T1-weighted IR-SPGR pulse sequence (25.6 cm field of view; 172 slices; 1.0 mm slice thickness; 256 × 256 matrix size).

fMRI data analysis

fMRI data were analyzed using the AFNI suite of programs (Cox 1996). For preprocessing, functional data were corrected for field inhomogeneities with field mapping data collected before functional scanning, coregistered in three dimensions with the anatomical scan, slice-time corrected, resampled to 3-mm isotropic voxels, and co-registered through time to reduce effects of head motion (using `afni_proc.py`). Large motion events, defined as MR volumes in which there was > 0.3° of rotation or >0.6 mm of translation in any direction across functional images, were excluded from the deconvolution analysis by censoring the affected time points but without affecting the temporal structure of the data.

Behavioral vectors (described below) along with six vectors that coded for motion (three for translation and three for rotation), and three polynomial vectors that coded for linear, quadratic, and cubic drift in the MRI signal were used in deconvolution analyses of the fMRI time series data (3dDeconvolve software). The deconvolution method does not assume a shape of the hemodynamic response, and the fit of the data to each model was estimated for each time point independently (0–10 sec after trial onset). The resultant fit coefficients (β coefficients) represent activity (relative to the digit task baseline) in each voxel for a given time point and for each of the response categories. This activity was summed over the expected hemodynamic response (2–10 sec) after the trial onset and taken as the estimate of the response.

To combine the data across participants, we used Advanced Normalization Tools (ANTS), which implements SyN (symmetric normalization), a diffeomorphic registration algorithm (Klein et al. 2009). An averaged anatomical volume was constructed based on the Talairach-transformed structural scans from each participant. After spatial transformation of the anatomical scans, the functional maps (i.e., summed β weights) were also transformed to the same common anatomical space. After preprocessing and spatial transformation, the amount of spatial smoothing was 5.78 mm (calculated using 3dFWHMx). No additional spatial smoothing was carried out.

Two models were created. For each model, behavioral vectors were created that coded each retrieval trial according to the memory confidence rating (1–6) and the old/new status of the stimulus. When a participant had 5 or fewer trials for a behavioral vector, the participant was excluded from that analysis (see sec-

ond model below). One model ($N = 18$) categorized targets and foils presented during scanning according to their old/new status and according to the confidence ratings given during scanning: hits (confidence ratings 4–6 for targets), misses (confidence ratings 1–3 for targets), correct rejections (confidence ratings 1–3 for foils), and false alarms (confidence ratings 4–6 for foils). The mean number of trials per participant for each of the four trial types was 193 ± 5 , 47 ± 5 , 174 ± 6 , and 66 ± 6 , respectively.

As is typically the case in memory experiments, when participants indicated that they had viewed a scene before, they gave higher confidence ratings when they were correct (hits) than when they were incorrect (false alarms). We therefore created another model in order to analyze the data for hits and false alarms when confidence ratings were similar for the two trial types. This model separated trials for hits and false alarms into trials when confidence was similar but low (items identified as “maybe old”, i.e., confidence ratings of 4; $N = 17$) and trials when confidence was similar but high (items identified as “probably old” or “definitely old”, i.e., confidence ratings of 5 or 6; $N = 17$). Confidence ratings 5 and 6 were combined to create a single high-confidence condition because five participants gave too few false alarms with a confidence rating of 6 to allow for fMRI analysis based on this confidence rating alone. With this procedure the average confidence rating was 5.8 for hits and 5.2 for false alarms. To match confidence ratings more closely, we included all hit trials associated with a confidence rating of 5, and for each participant we randomly selected for analysis a proportion of the hit trials associated with a confidence rating of 6. This was done such that the ratio of 5 trials to 6 trials was the same for hits and false alarms (mean confidence rating for both high-confidence hits and false alarms was now 5.2). On average, 119 ± 9 hit trials associated with a confidence rating of 6 were eliminated for each participant. These trials were combined into one behavioral vector but were not analyzed further. The mean number of trials per participant for low and high-confidence conditions was 27 ± 3 and 49 ± 4 for hits and 33 ± 3 and 35 ± 5 for false alarms, respectively.

Vectors were also created for low-confidence and high-confidence correct rejections (for the conjunction analysis, see below). The low-confidence trials included items identified as “maybe new”, i.e., a confidence rating of 3 ($N = 18$). The high-confidence trials included items identified as “probably new” or “definitely new”, i.e., confidence ratings of 2 or 1 ($N = 18$). By inverting these confidence scales to align with items identified as old (1 = 6, 2 = 5, 3 = 4), average confidence ratings for the low- and high-confidence correct rejections were 4.0 and 5.4, respectively.

For the analyses of interest, we first identified brain regions that differentiated true memory (hits) and false memory (false alarms) by carrying out a paired *t*-test for all hits and false alarms. Next, we carried out paired *t*-tests to identify brain regions that differentiated true and false memories when the confidence ratings associated with hits and false alarms were similar. One test compared low-confidence true and false memories. The other test compared high-confidence true and false memories.

Finally, we used conjunction analyses to identify brain regions where activity was similar for true memory and false memory. We first carried out a conjunction analysis using all hits, false alarms, and correct rejections to identify voxels that were common for true memory (hits > correct rejections) and false memory (false alarms > correct rejections) (Garoff-Eaton et al. 2006; Gutchess and Schacter 2011). These voxels were then exclusively masked with (i.e., to eliminate) voxels where activity was significantly different for hits and false alarms (i.e., hits > false alarms or false alarms > hits), leaving only voxels that exhibited similar activity for true and false memories. Next, we carried out two conjunction analyses for trials where confidence ratings were similar for hits, false alarms, and correct rejections. One analysis was carried out for low-confidence true memory (low-confidence hits > low-confidence correct rejections) and low-confidence false memory (low-confidence false alarms > low-confidence correct rejections). The other analysis was carried out for high-confidence true memory (high-confidence hits > high-confidence correct rejections) and high-confidence false memory (high-confidence

false alarms > high-confidence correct rejections). At each confidence level, the identified voxels were exclusively masked with (i.e., to eliminate) voxels where activity was significantly different for hits and false alarms (i.e., hits > false alarms or false alarms > hits) at the corresponding confidence level, such that the remaining voxels were those that exhibited similar activity for true memory and false memory.

All analyses (including *t*-tests, conjunction analyses, and exclusive masks) were thresholded at a voxel-wise probability of $P < 0.01$. We corrected for multiple comparisons across voxels by using a minimum cluster extent threshold to account for false positives. Specifically, a Monte Carlo simulation using AlphaSim software was used to determine the minimum cluster size needed to obtain a cluster-wise probability threshold of $P < 0.05$. The correction also took into account the amount of spatial smoothing calculated from the data (5.78 mm). This correction was carried out separately for the MTL (1,214 3-mm isotropic voxels) and the whole brain (59,040 3-mm isotropic voxels). The MTL analyses included only regions in the MTL (i.e., the hippocampus and parahippocampal gyrus), which were hand drawn based on the averaged anatomical volume. The whole-brain analyses excluded the MTL regions. Based on the Monte Carlo simulation, we determined that clusters needed to contain at least 12 contiguous voxels in the MTL (324 mm³) and at least 33 contiguous voxels in the whole brain (891 mm³) to obtain a cluster-wise probability threshold of $P < 0.05$. For the conjunction analyses, each individual contrast was thresholded at $P < 0.01$ for a conjoint threshold of $P < 0.01$ (Nichols et al. 2005).

Acknowledgments

We thank Erin Light, Ashley Knutson, Jennifer Frascino, and Anna van der Horst for assistance. We also gratefully acknowledge the NVIDIA Corporation for their donation of a Quadro K5000 GPU which was used for this research. This work was supported by the Medical Research Service of the Department of Veterans Affairs, NIMH Grant 24600, and NIMH Training Grant 20002 to Z.J.U.

The authors declare no conflict of interest.

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.
- Cansino S, Maquet P, Dolan RJ, Rugg MD. 2002. Brain activity underlying encoding and retrieval of source memory. *Cereb Cortex* **12**: 1048–1056.
- Cox RW. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* **29**: 162–173.
- Curran T. 2000. Brain potentials of recollection and familiarity. *Mem Cognit* **28**: 923–938.
- Deese J. 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J Exp Psychol* **58**: 17–22.
- Dennis NA, Bowman CR, Vandekar SN. 2012. True and phantom recollection: an fMRI investigation of similar and distinct neural correlates and connectivity. *Neuroimage* **59**: 2982–2993.
- Dennis NA, Bowman CR, Turney IC. 2014. Neural correlates of false memories. In *The cognitive neuroscience of human memory* (ed. Duarte A, Addis DR, Barense M). Wiley-Blackwell.
- Dobbins IG, Rice HJ, Wagner AD, Schacter D. 2002. Memory orientation and success: separable neurocognitive components underlying episodic recognition. *Neuropsychologia* **41**: 318–333.
- Eichenbaum H, Cohen NJ. 2001. *From conditioning to conscious recollection: memory systems of the brain*. Oxford University Press, New York.
- Garoff-Eaton RJ, Slotnick SD, Schacter DL. 2006. Not all false memories are created equal: the neural basis of false recognition. *Cereb Cortex* **16**: 1645–1652.
- Garoff-Eaton RJ, Kensinger EA, Schacter DL. 2007. The neural correlates of conceptual and perceptual false recognition. *Learn Mem* **14**: 684–692.
- Gutchess AH, Schacter DL. 2011. The neural correlates of gist-based true and false recognition. *Neuroimage* **59**: 3418–3426.
- Iidaka T, Harada T, Kawaguchi J, Sadato N. 2012. Neuroanatomical substrates involved in true and false memories for face. *Neuroimage* **62**: 167–176.
- Kahn I, Davachi L, Wagner AD. 2004. Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. *J Neurosci* **24**: 4172–4180.
- Kim H, Cabeza R. 2007. Trusting our memories: dissociating the neural correlates of confidence in veridical versus illusory memories. *J Neurosci* **27**: 12190–12197.
- Kirwan CB, Shrago Y, Squire LR. 2009. Medial temporal lobe activity can distinguish between old and new stimuli independently of overt behavioral choice. *Proc Natl Acad Sci* **106**: 14617–14621.
- Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, et al. 2009. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* **46**: 786–802.
- Nichols T, Brett M, Andersson J, Wager T, Poline JB. 2005. Valid conjunction inference with the minimum statistic. *Neuroimage* **25**: 653–660.
- Okado Y, Stark CEL. 2003. Neural processing associated with true and false memory retrieval. *Cogn Affect Behav Neurosci* **3**: 323–334.
- Roediger HL, McDermott KB. 1995. Creating false memories: remembering words not presented in lists. *J Exp Psychol Learn* **21**: 803–814.
- Schacter DL, Slotnick SD. 2004. The cognitive neuroscience of memory distortion. *Neuron* **44**: 149–160.
- Schacter DL, Reiman E, Curran T, Yun LS, Bandy D, McDermott KB, Roediger HL III. 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.
- Slotnick SD, Schacter DL. 2004. A sensory signature that distinguishes true from false memories. *Nat Neurosci* **7**: 664–672.
- Squire LR, Zola-Morgan S. 1991. The medial temporal lobe memory system. *Science* **253**: 1380–1386.
- Stark CEL, Squire LR. 2001. When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proc Natl Acad Sci* **98**: 12760–12766.
- Stark CEL, Okado Y, Loftus EF. 2010. Imaging the reconstruction of true and false memories using sensory reactivation and the misinformation paradigm. *Learn Mem* **17**: 485–488.
- Weis S, Specht K, Klaver P, Tendolker I, Willmes K, Ruhlmann J, Elger CE, Fernández G. 2004. Process dissociation between contextual retrieval and item recognition. *Neuroreport* **15**: 2729–2733.
- Wheeler ME, Buckner RL. 2003. Functional dissociation among components of remembering: control, perceived oldness, and content. *J Neurosci* **23**: 3869–3880.
- Wheeler ME, Buckner RL. 2004. Functional-anatomic correlates of remembering and knowing. *Neuroimage* **21**: 1337–1349.
- Wixted JT. 2007. Dual-process theory and signal-detection theory of recognition memory. *Psychol Rev* **114**: 152–176.

Received February 3, 2015; accepted in revised form August 11, 2015.