Behavioral/Systems/Cognitive

Trusting Our Memories: Dissociating the Neural Correlates of Confidence in Veridical versus Illusory Memories

Hongkeun Kim[1] **and Roberto Cabeza**[2]

[1]Department of Rehabilitation Psychology, Daegu University, Daegu 705-714, South Korea, and [2]Center for Cognitive Neuroscience, Duke University, Durham, North Carolina 27708-0999

Although memory confidence and accuracy tend to be positively correlated, people sometimes remember with high confidence events that never happened. How can confidence correlate with accuracy but apply also to illusory memories? One possible explanation is that high confidence in veridical versus illusory memories depends on different neural mechanisms. The present study investigated this possibility using functional magnetic resonance imaging and a modified version of the Deese-Roediger-McDermott false-memory paradigm. Participants read short lists of categorized words, and brain activity was measured while they performed a recognition test with confidence rating. The study yielded three main findings. First, compared with low-confidence responses, high-confidence responses were associated with medial temporal lobe (MTL) activity in the case of true recognition but with frontoparietal activity in the case of false recognition. Second, these regions showed significant confidence-by-veridicality interactions. Finally, only MTL regions showed greater activity for high-confidence true recognition than for high-confidence false recognition, and only frontoparietal regions showed greater activity for high-confidence false recognition than for high-confidence true recognition. These findings indicate that confidence in true recognition is mediated primarily by a recollection-related MTL mechanism, whereas confidence in false recognition reflects mainly a familiarity-related frontoparietal mechanism. This account is consistent with the fuzzy trace theory of false recognition. Correlation analyses revealed that MTL and frontoparietal regions play complementary roles during episodic retrieval. In sum, the present study shows that when one focuses exclusively on high-confidence responses, the neural correlates of true and false memory are clearly different.

Key words: **fMRI; false memory; prefrontal cortex; medial temporal lobe; human memory; memory confidence**

Introduction

Although memory confidence and memory accuracy are usually positively correlated (Lindsay et al., 1998), in certain situations, we can remember with high confidence events that never happened (Schacter, 2001). How can confidence correlate with accuracy but apply also to illusory memories? One possible explanation is that high confidence in veridical versus illusory memories depends on different neural mechanisms. In the present study, we investigated this idea using functional magnetic resonance imaging (fMRI) and a modified version of the Deese-Roediger-McDermott (DRM) false-memory paradigm (Deese, 1959; Roediger and McDermott, 1995).

In this paradigm, participants study a list of words that are all related to a critical word that is not presented, and at test, they show a strong tendency to falsely recognize the critical lure. According to a popular false-memory theory (Reyna and Brainerd, 1995; Schacter et al., 1996c), true recognition of studied list items is supported by retrieval of item-specific information, whereas false recognition of the critical lures reflects mainly the retrieval of semantic gist information. Thus, true recognition may be accompanied by vivid remembering of specific contextual details, or recollection, whereas false recognition is largely based on a feeling of oldness in the absence of contextual details, or familiarity (Brainerd and Reyna, 2002). Thus, high confidence in true recognition is likely to involve brain regions associated with recollection, whereas high confidence in false recognition is more likely to involve regions associated with familiarity.

Recollection has been strongly associated with medial temporal lobe (MTL) regions, and familiarity has been associated with prefrontal cortex (PFC) and parietal regions. Lesions in MTL regions such as the hippocampus have been shown to impair recollection rather than familiarity in both humans (Yonelinas et al., 2002) and animals (Fortin et al., 2004), and functional neuroimaging studies often find recollection-related activations within MTL (for review, see Eichenbaum et al., 2007). Although familiarity has been also linked to some MTL regions, such as the perirhinal cortex, its associations with PFC and parietal regions are supported by both lesion and neuroimaging evidence. For example, Duarte et al. (2005) found that patients with dorsolateral PFC lesions were significantly impaired in familiarity. Functional neuroimaging studies have associated familiarity with both PFC and parietal activations (Cansino et al., 2002; Yonelinas et al., 2005).

Received Feb. 27, 2007; revised Sept. 4, 2007; accepted Sept. 5, 2007.

This work was supported by a Daegu University research grant in 2007 (H.K.) and by National Institutes of Health Grants AG19731 and AG23770 (R.C.). We thank Amber Baptiste for participant recruitment and Rakesh Arya for technical assistance.

Correspondence should be addressed to Hongkeun Kim, Department of Rehabilitation Psychology, Daegu University, 2288 Daemyung-dong, Nam-gu, Daegu 705-714, South Korea. E-mail: hongkn1@gmail.com.

DOI:10.1523/JNEUROSCI.3408-07.2007

 0270-6474/07/2712190-08$15.00/0

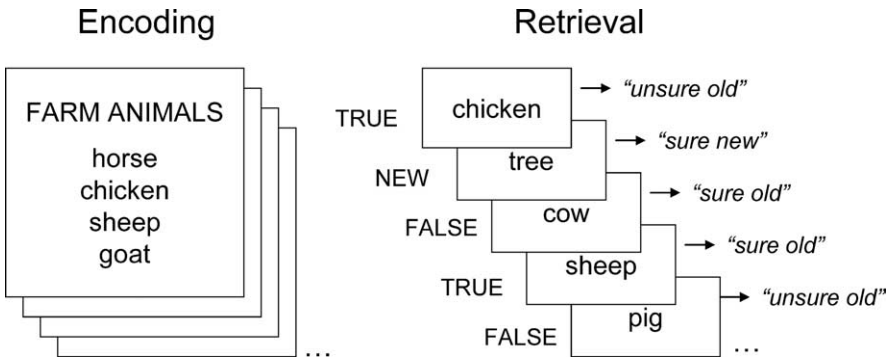

Figure 1. Behavioral paradigm. The encoding task was a category judgment task. The retrieval task was an old–new recognition test with confidence ratings that included studied words (true words), nonstudied words from studied categories (false words), and nonstudied, unrelated words (new words).

Thus, the goal of the present study was to investigate the hypothesis that confidence in true recognition is mediated primarily by recollection-related MTL activity, whereas confidence in false recognition is mediated mainly by familiarity-related frontoparietal activity. To this end, we investigated whether separate contrasts would show that high-confidence recognition responses elicit greater MTL activity in the case of true recognition, but greater frontoparietal activity in the case of false recognition. Also, we investigated whether these regions would show confidence (high vs low) × veridicality (true vs false) interactions. Furthermore, we kept high-confidence constant and directly compared true and false recognition. Finally, we investigated the relationship between MTL and frontoparietal regions using correlation analyses.

Materials and Methods

Subjects. Sixteen young adults participated in the experiment. They were healthy, right-handed, native English speakers, with no history of neurological or psychiatric episodes. All subjects gave informed consent to a protocol approved by the Duke University Institutional Review Board. In an effort to identify neural correlates of confidence in true recognition and false recognition, the analyses focused on the following four trial types: high-confidence true recognition, low-confidence true recognition, high-confidence false recognition, and low-confidence false recognition. Five subjects were excluded from the analyses because of sparse number (<10) of trials in one of the four trial types. Thus, the reported results are based on the data from the remaining 11 subjects (five female; age range, 18–30).

Stimulus materials. The method we used was an adaptation of DRM false-memory paradigm (Deese, 1959; Roediger and McDermott, 1995). The materials were 72 categorical six-word lists selected from category norms (Battig and Montague, 1969; Yoon et al., 2004). Each list consisted of the six most typical instances (e.g., cow, pig, horse, chicken, sheep, goat) of a natural/artificial category (e.g., farm animal), with minor exceptions. In each list, the third to the sixth typical instances were used as encoding stimuli (true words); the first and the second typical instances were used as "critical lures" (false words) in the test phase. Additionally, semantically unrelated words, matched in letter number, frequency, and concreteness to the category words, were used as control words (new words) in the test phase. The categories were carefully chosen so that their instances did not overlap. Thus, both "farm animal" and "wild animal" categories were included in the stimulus set, but "four-legged animal" was not included.

Task procedures. The paradigm is illustrated by Figure 1. During the study phase, participants viewed one by one 82 "mini" word lists, each consisting of a category name and four of the most typical members of the category. Each list was presented for 4 s. The subjects' task was to decide whether all four or only three instances belonged to the category by pressing one of the two keys in a response box using their right hand. In 72 "critical" trials used in fMRI analyses, all four words were member of the category, whereas in 10 "catch" trials, only three of the four words belonged to the category. The test phase, which started ~10 min after completion of the study phase, consisted of six scans. There were a total of 288 true-word, 144 false-word, and 144 new-word trials across all scans. Trials were presented in a predetermined, pseudorandom order. In each trial, a word was shown for 2 s, followed by a fixation cross for 1 s. A fixation period, ranging from 1.5 to 4.5 s, was interspersed across both study and test trials to "jitter" the onset times of trials and allow event-related fMRI analyses. All words in the test phase were displayed in white color against black background. Subjects responded by pressing one of four keys according to whether the word was judged to be "sure old," "unsure old," "unsure new," or "sure new."

fMRI procedures. MRI scanning was conducted using a 4-T GE magnet. Scanner noise was reduced with earplugs, and head motion was reduced with foam pads and headbands. Stimuli were presented with liquid-crystal display goggles. Anatomical scanning started with a T2-weighted sagittal localizer series. The anterior commissure (AC) and posterior commissure (PC) were identified in the midsagittal slice, and 34 contiguous oblique slices were prescribed parallel to the AC–PC plane. High-resolution T1-weighted structural images were collected with a 500 ms repetition time (TR), a 14 ms echo time (TE), a 24 cm field of view (FOV), a 256^2 matrix, 68 slices, and a slice thickness of 1.9 mm. Functional images were acquired using an inverse spiral sequence with a 1500 ms TR, a 6 ms TE, a 24 cm FOV, a 64^2 matrix, and a 60° flip angle. Thirty-four contiguous slices were acquired with the same slice prescription as the anatomical images. Slice thickness was 3.75 mm, resulting in cubic 3.75 mm^3 isotropic voxels.

Although scanning took place during encoding and retrieval, the present study focuses on fMRI data collected during retrieval. The fMRI data collected during encoding was reported previously (Kim and Cabeza, 2007). Image processing and analyses were performed using SPM2 software (www.fil.ion.ucl.ac.uk/spm/). After discarding the first six volumes, the functional images were slice-timing corrected and motion corrected, and then spatially normalized to the Montreal Neurological Institute (MNI) templates implemented in SPM2. The coordinates were later converted to Talairach and Tournoux (1988) space. Subsequently, the functional images were spatially smoothed using an 8 mm isotropic Gaussian kernel, and resliced to a resolution of 3.75 mm^3 isotropic voxels.

Trial-related fMRI activity was first modeled by convolving a vector of the onset times of the stimuli with a canonical hemodynamic response function (HRF). The general linear model (GLM), as implemented in SPM2, was used to model the effects of interest and other confounding effects (e.g., head movement and magnetic field drift). Trials were coded based on item status (true words, false words, new words) and subjects' responses (sure old, unsure old, unsure new, sure new). In an effort to identify neural correlates of confidence in true recognition and false recognition, four critical trial types were selected a priori for further analyses: (1) high-confidence true recognition (i.e., sure old responses to true words), (2) low-confidence true recognition, (3) high-confidence false recognition (i.e., sure old responses to false words), and (4) low-confidence false recognition. For each subject, statistical parametric maps (SPM) pertaining to the effects of interest were identified and subsequently integrated across subjects using a random-effects model.

Preliminary fMRI analysis. In addition to analyses relevant to the main hypothesis, which, as described below, focus on differential effects of confidence on true and false recognition, we conducted a preliminary analysis to identify similar effects of confidence on true and false recognition. This contrast identified the main effects of confidence (high > low and low > high) on all recognition trials (i.e., collapsed over true and false recognition) at a threshold of $p < 0.005$ with >15 contiguous vox-

Table 1. Behavioral results: number of trials and reaction time

		Response			
	Condition	Sure old	Unsure old	Unsure new	Sure new
Number of trials					
	True	138 (27)	74 (26)	51 (21)	21 (21)
	False	28 (12)	48 (18)	46 (13)	19 (19)
	New	6 (6)	23 (16)	78 (28)	35 (28)
Reaction time (ms)					
	True	1239 (155)	1487 (175)	1538 (202)	1573 (257)
	False	1307 (165)	1513 (215)	1545 (172)	1528 (190)
	New	1514 (422)	1522 (272)	1477 (198)	1480 (212)

Values are across-subject means (SD).

els, and then masked out voxels showing veridicality × confidence interactions at a very lenient threshold ($p < 0.15$). It is worth noting that the more lenient the threshold of the mask, the more strict the test for the absence of an interaction.

Main fMRI analyses. As stated in the Introduction, our main hypothesis posits that confidence in true recognition is mediated primarily by recollection-related MTL activity, whereas confidence in false recognition is mediated mainly by familiarity-related frontoparietal activity. We examined this hypothesis across three analyses: (1) contrasts between high- and low-confidence trials performed separately for true and false recognition; (2) a test of veridicality (true vs false)-by-confidence (high vs low) interactions; and (3) a direct contrast between high-confidence true recognition and high-confidence false recognition. All effects were assessed using an uncorrected threshold of $p < 0.005$ with an extent threshold of >15 contiguous voxels, which yields a false positive probability of 0.00001 per voxel according to Monte Carlo simulations of spatially correlated data (Forman et al., 1995). In addition to SPM-based contrasts, follow-up region-of-interest (ROI) analyses were performed for certain significant clusters in the group contrast analyses. From each subject and ROI, the mean parameter estimate across all significant voxels was extracted for the four critical trial types, respectively. These parameter estimates were subject to repeated-measures ANOVAs.

In addition to these three main analyses, to investigate the relationship between MTL and frontoparietal regions we performed correlation analyses including these two sets of regions. These analyses were performed in three steps. First, from each subject and a pair of ROIs (e.g., left MTL and left PFC), a within-subject Pearson correlation coefficient was computed across the mean parameter estimates of the four critical conditions. Second, the mean of this correlation was computed across subjects, and finally, significance of this mean correlation was tested using a random effects approach. For all ROI analyses, the significance threshold was set at $p < 0.05$, two-tailed.

Results

Behavioral performance

Category judgment at the study phase was very accurate (mean, 95% correct). Behavioral results at the test phase are summarized in Table 1. At the test phase, combined across high- and low-confidence recognition, the proportion of hits for true words (73.7%) was significantly greater than the proportion of false alarms for false words (52.5%; $t_{(10)} = 5.92$, $p < 0.001$), which in turn was significantly greater than the proportion of false alarms for new words (19.8%; $t_{(10)} = 10.50$, $p < 0.001$). Critically for our fMRI analyses, there was a substantial number of high-confidence false alarms to false words (20%). In contrast, high-confidence false alarms to new words were scarce (4%) and significantly fewer ($t_{(10)} = 7.25$; $p < 0.001$). Thus, the paradigm was effective in eliciting high-confidence false memories. As expected, reaction times were longer for low- than for high-confidence responses (true words: $t_{(10)} = 8.21$, $p < 0.001$; false words: $t_{(10)} = 6.49$, $p < 0.001$).

Common effects of confidence on true and false recognition

To identify similar effects of confidence on true recognition and false recognition activity, we tested the contrasts [(high-confidence true recognition and false recognition) > (low-confidence true recognition and false recognition)] and [(low-confidence true recognition and false recognition) > (high-confidence true recognition and false recognition)] (supplemental Table 1, available at www.jneurosci.org as supplemental material). The most notable regions identified in these contrasts are illustrated in Figure 2. A right dorsolateral/anterior PFC region [Brodmann area (BA) 46/10] (Fig. 2*A*) and a dorsal precuneus region (BA 7) (Fig. 2*B*) showed common low-confidence activity across true recognition and false recognition. In contrast, a posterior cingulate region (BA 31) (Fig. 2*C*) showed common high-confidence activity across true recognition and false recognition. The general involvement of right dorsolateral PFC in low-confidence responses fits very well with fMRI evidence that this region is involved in monitoring and/or decision processes across different cognitive functions (Henson et al., 2000; Fleck et al., 2006). The dorsal precuneus was associated with low-confidence responses, whereas the posterior cingulate was associated with high-confidence responses. Although this dissociation is tangential to the main goals of this study, it is worth mentioning that it is consistent with the results of two recent fMRI studies of episodic retrieval (Yonelinas et al., 2005; Daselaar et al., 2006).

Differential effects of confidence on true and false recognition

The main hypothesis states that confidence in true recognition is mediated mainly by recollection-related MTL activity, whereas confidence in false recognition is mediated mainly by familiarity-related frontoparietal activity. To investigate this hypothesis, we performed three fMRI analyses. First, we compared high- and low-confidence responses separately for true recognition and for false recognition (Table 2). Consistent with our hypothesis, high-confidence responses were associated with MTL activity in the case of true recognition (Fig. 3*A*, top) but with frontoparietal activity in the case of false recognition (Fig. 3*B*, bottom). In contrast, low-confidence true recognition was associated with frontoparietal regions (Fig. 3*A*, bottom) and low-confidence false recognition with MTL regions (Fig. 3*B*, top).

An interesting secondary finding during true recognition was a dissociation between a more dorsal posterior parietal region (BA 7; in or around the intraparietal sulcus), which showed greater activity for low- than high-confidence true recognition (Fig. 3*A*, bottom, blue parietal region), and a more ventral parietotemporal region (BA 39), which showed greater activity for high- than low-confidence true recognition (yellow parietal region in the same image). This dissociation was confirmed by a significant region × confidence interaction ($F_{(1,10)} = 47.53$, $p < 0.001$). This finding is consistent with evidence linking parietotemporal cortex to recollection (Wheeler and Buckner, 2004; Yonelinas et al., 2005; Daselaar et al., 2006), and with evidence that dorsal and ventral posterior parietal regions play different roles in episodic retrieval (for review, see Wagner et al., 2005).

Second, although the results of separate analyses on true and false recognition supported our hypothesis, to confirm the differential involvement of MTL and frontoparietal regions in true versus false recognition, we entered both conditions into the same model and searched for voxels showing confidence (high vs low) × veridicality (true vs false) interactions across the whole brain. Consistent with our hypothesis, significant interactions were found almost exclusively in MTL and frontoparietal regions

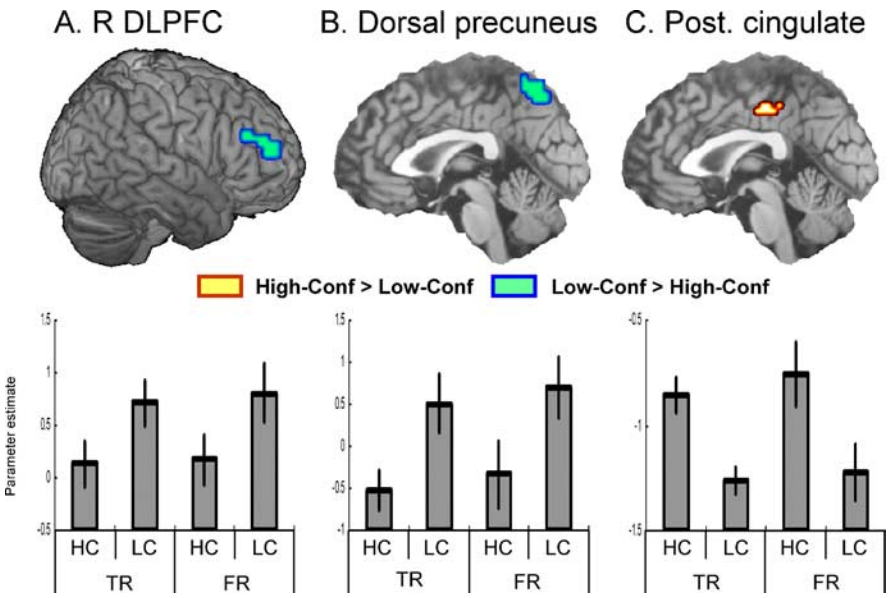

Figure 2. Right (R) dorsolateral/anterior PFC (***A***) and dorsal precuneus regions (***B***) showed greater activity for low- than high-confidence (Conf) response for both true recognition and false recognition. In contrast, posterior (Post.) cingulate region (***C***) showed greater activity for high- than low-confidence response for both true recognition and false recognition. The bar graphs display mean parameter estimates across all significant voxels. Error bars show ± 1 SE. HC, High confidence; LC, low confidence; TR, true recognition; FR, false recognition.

Table 2. Brain regions showing significant differences between high- versus low-confidence activity in true recognition and false recognition

			Talairach				
	H	BA	*x*	*y*	*z*	Voxels	*t*
True recognition							
High-confidence activity > low-confidence activity							
MTL							
Hippocampus	R	—	27	−27	−8	23	7.79
	L	—	−27	−27	−11	59	7.63
Amygdala	R	—	27	3	−16	20	4.79
Anterior cingulate	L	32	−8	33	−5	41	6.73
	R	32	15	33	6	39	4.94
Posterior cingulate	B	31	15	−54	30	235	6.67
Subcallosal gyrus	B	25	4	10	−10	33	4.78
Parietotemporal cortex	L	39	−46	−80	25	56	5.50
Postcentral gyrus	R	5	23	−38	65	23	4.04
Low-confidence activity > high-confidence activity							
PFC							
Anterior	R	10/46	38	46	19	68	6.33
Ventrolateral	L	44/45	−49	15	30	111	6.21
Dorsolateral	L	6	−27	4	57	34	5.84
	R	6	19	8	65	61	7.87
Dorsomedial	B	6/8	−8	8	65	136	6.30
Lateral PPC	L	7	−27	−65	49	78	6.21
Dorsal precuneus	B	7	11	−65	53	78	6.47
False recognition							
High-confidence activity > low-confidence activity							
PFC							
Dorsolateral	L	8	−34	20	38	20	5.33
	R	8	34	17	42	27	5.37
Dorsomedial	B	8/9	−8	29	47	64	5.47
Ventrolateral	L	44/45	−38	16	20	16	4.10
Lateral PPC	L	39/40	−30	−50	34	174	7.53
Anterior cingulate	L	32	−4	45	8	16	5.10
Central cingulate	B	24	0	1	28	18	5.35
Low-confidence activity > high-confidence activity							
MTL: PHC	R	36	19	−30	−23	21	6.40
Dorsal precuneus	B	7	0	−65	61	50	4.66
Thalamus	B	—	11	−27	8	17	4.01

H, Hemisphere; L, left; R, right; B, bilateral; BA, Brodmann area; MTL, medial temporal lobe; PFC, prefrontal cortex; PPC, posterior parietal cortex; PHC, parahippocampal cortex.

(Table 3, Fig. 4). Within MTL, bilateral posterior hippocampal and parahippocampal activity was greater for high- than low-confidence true recognition, but it was greater for low- than high-confidence false recognition (Fig. 4*A*). Within bilateral frontoparietal regions, in contrast, activity was greater for low- than for high-confidence true recognition, but it was greater for high- than low-confidence false recognition (Fig. 4*B*). Thus, the pattern of frontoparietal activity was a virtual mirror image of the pattern of MTL activity (Fig. 4*A*,*B*, compare bar graphs). These dissociations were confirmed by a significant region × confidence × veridicality interaction between left MTL and both left PFC ($F_{(1,10)} = 74.67$, $p < 0.001$) and left dorsal parietal cortex ($F_{(1,10)} = 57.28$, $p < 0.001$) and between right MTL and both right PFC ($F_{(1,10)} = 143.39$, $p < 0.001$) and right dorsal parietal cortex ($F_{(1,10)} = 55.63$, $p < 0.001$).

Finally, to more directly test the hypothesis that high-confidence true recognition is driven by MTL-mediated recollection whereas high-confidence false recognition is driven by frontoparietal-mediated familiarity, we compared high-confidence true and false recognition directly to each other across all voxels in the brain. Consistent with the hypothesis, significant differences were found exclusively in frontoparietal and MTL regions (Table 4). As illustrated by Figure 5, the results directly supported our hypothesis. Compared with high-confidence false recognition, high-confidence true recognition elicited greater activity in bilateral MTL regions (posterior hippocampus and parahippocampal cortex) and no other brain regions (Fig. 5*A*). Compared with high-confidence true recognition, high-confidence false recognition elicited greater activity in frontoparietal regions (Fig. 5*B*), including anterior (BA 10) and posterior ventrolateral (BA 44) PFC regions and bilateral dorsal parietal regions (BA 40/7), and no other brain regions.

The activation patterns in Figure 4 suggest that (1) consistent with the notion of a frontoparietal network, posterior ventrolateral PFC and dorsal parietal regions behaved very similarly to each other, and (2) they behaved very differently than MTL regions. To investigate these ideas, correlation analyses were conducted among the two frontal, two parietal, and two MTL regions displayed in Figure 4. From each subject and a pair of ROIs, a within-subject Pearson correlation was computed across the four critical conditions, the mean of

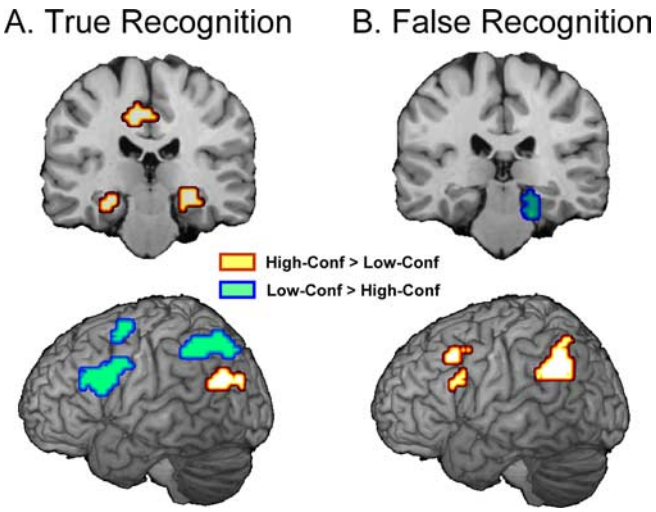

Figure 3. ***A***, For true recognition, high-confidence (Conf) responses (yellow) were associated with MTL activity, and low-confidence responses (blue) were associated with frontoparietal activity. A ventral posterior parietal region was associated with high-confidence activity (yellow) and was dissociated from the more dorsal posterior parietal region (nearby blue region). ***B***, For false recognition, high-confidence responses (yellow) were associated with the frontoparietal activity and low-confidence responses (blue) with MTL activity. The results support the hypothesis that high confidence in true recognition is mediated by MTL (recollection), whereas high confidence in false recognition is mediated by frontoparietal network (familiarity).

Table 3. Brain regions showing confidence (high vs low)-by-veridicality (true vs false) interactions

			Talairach				
	H	BA	*x*	*y*	*z*	Voxels	*t*
HC-TR > LC-TR and LC-FR > HC-FR							
MTL: hippocampus/PHC	L	36	−34	−30	−11	18	5.06
	R	36	19	−34	−14	35	7.52
Insula	R	—	46	10	−10	20	6.17
LC-TR > HC-TR and HC-FR > LC-FR							
PFC							
Ventrolateral	L	44/45	−42	19	20	97	6.28
Dorsolateral	R	8/9	38	20	34	51	6.32
Dorsomedial	B	8	−8	28	34	94	5.92
Lateral PPC	L	40	−30	−50	34	78	4.61
	R	40	42	−45	45	18	4.28

HC, High confidence; LC, low confidence; TR, true recognition; FR, false recognition. For other abbreviations, see Table 2.

this correlation was computed across subjects, and statistical significance of this mean correlation was tested using a random effects approach. These analyses were done separately for the left and right hemispheres. The results yielded two main findings (Fig. 6). First, consistent with the idea of a frontoparietal network, PFC activity was positively correlated with dorsal parietal activity. Second, consistent with the idea of a complementary relationship between the frontoparietal network and MTL, PFC and parietal activations were negatively correlated with MTL activations. All correlations were significant except the correlation between left MTL and dorsal parietal cortex ($p = 0.11$). The results of correlation analyses suggest the existence of a parallel relationship between PFC and parietal regions and a complementary relationship between this frontoparietal network and MTL regions.

Discussion

The goal of the present study was to test the hypothesis that confidence in true memories reflects mainly recollection-related MTL activity, whereas confidence in false memories reflects mainly familiarity-related frontoparietal activity. The present study yielded three findings supporting this hypothesis. First, compared with low-confidence responses, high-confidence responses were associated with MTL activity in the case of true recognition but with frontoparietal activity in the case of false recognition (Fig. 3). Second, these regions showed significant confidence-by-veridicality interactions (Fig. 4). Finally, only MTL regions showed greater activity for high-confidence true recognition than for high-confidence false recognition, and only frontoparietal regions showed the reverse pattern (Fig. 5). Together, these findings demonstrate a clear dissociation between the neural correlates of confidence in true and false recognition.

The dissociation between regions supporting confidence in true versus false recognition is consistent with the fuzzy trace theory (Brainerd and Reyna, 1990, 2002; Schacter et al., 1996c) account of false memories in the DRM paradigm. This theory assumes that studying a list of associates leads to formation of verbatim traces, which contain item-specific information, as well as gist traces, which contain the general meaning of the list. At the test phase, true memories are supported mainly by retrieval of verbatim traces, whereas false memories are supported primarily by retrieval of gist traces. Assuming that verbatim retrieval is mediated by regions associated with recollection, such as MTL, and gist processing is mediated by regions associated with familiarity, such as frontoparietal regions, our findings fit very well with the fuzzy trace theory. In terms of this theory, our findings suggest that the reason why we trust veridical memories is because of the quality of verbatim traces recovered by MTL, whereas the reason why we trust illusory memories is because of the quality of gist traces processed by a frontoparietal network.

The dissociation between the neural correlates of confidence in true and false recognition is not incompatible with functional neuroimaging evidence that true and false recognition activity may overlap (Cabeza et al., 2001; Schacter et al., 1996b). First, although we focused on differences, several regions showed similar effects of confidence for true and false recognition (Fig. 2). Second, overlaps in the neural correlates of true and false recognition are likely to be greater when using longer associative study lists, which elicit higher levels of false recognition, including illusory or "phantom" recollection of critical lures (Brainerd and Reyna, 2002). Phantom recollection, which include participants' tendency to "remember" critical lures in the standard DRM paradigm (Roediger and McDermott, 1995), tends to occur in conditions involving very high levels of false recognition (>80%) rather than the moderate levels observed in this current study (52.5%). According to one view (Lampinen et al., 2005), phantom recollection reflects "content borrowing" (that is, the misattribution of details belonging to list items to the critical lure). In contrast, the shorter categorical study lists investigated in this study yielded lower levels of false recognition, possibly emphasizing the role of familiarity in false recognition and differences with recollection-based true recognition.

We interpret MTL activity during high-confidence true recognition as reflecting mainly recollection. This assumption is supported by behavioral evidence that recollection-based recognition responses are almost always made with high confidence (Yonelinas, 2001), as well as with lesion and functional neuroimaging evidence linking MTL to recollection. MTL lesions yield significant recollection deficits (Fortin et al., 2004; Yonelinas et al., 2002), and MTL activity tends to be greater for remember than "know" responses (Eldridge et al., 2000; Yonelinas et al., 2005), for correct than incorrect source memory (Cansino et al.,

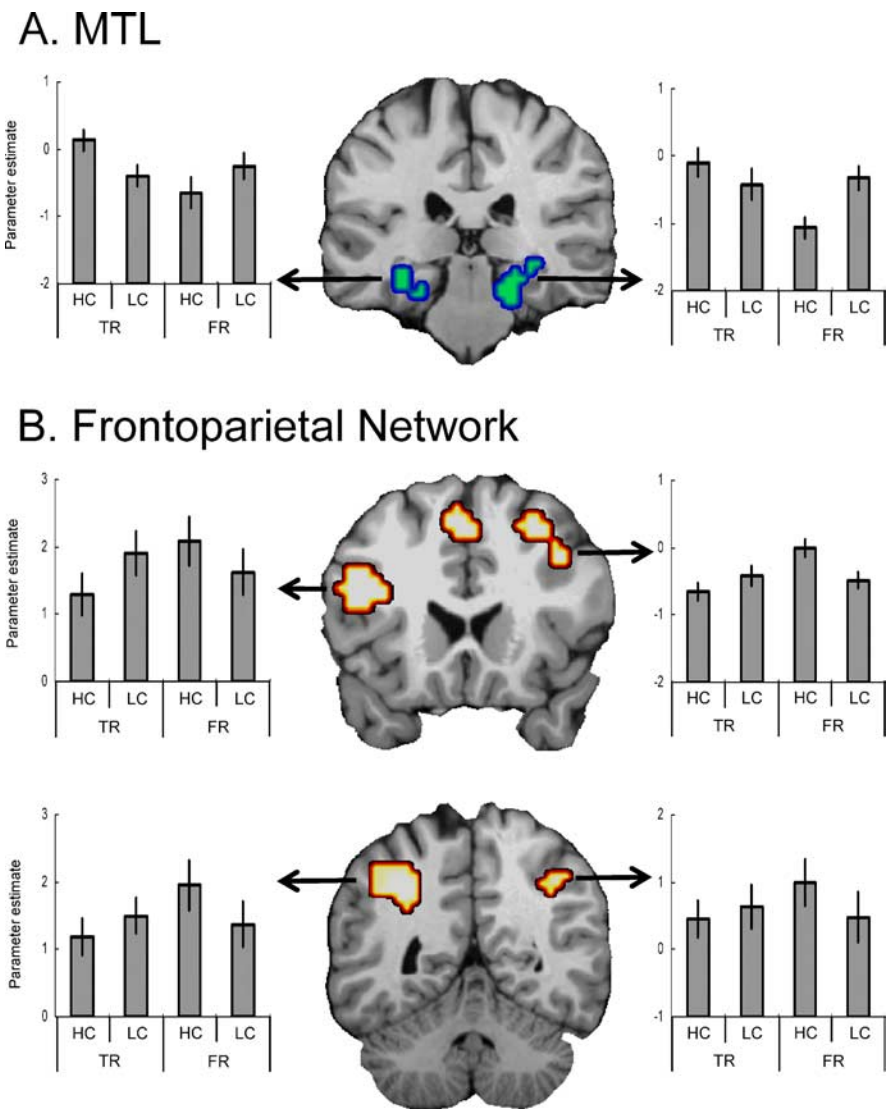

Figure 4. Activity in bilateral posterior hippocampal and parahippocampal area (***A***) was greater for high- than low-confidence true recognition, but greater for low- than high-confidence false recognition. Activity within a frontoparietal network (***B***) in both hemispheres was greater for low- than high-confidence true recognition, but greater for high- than low-confidence false recognition. See Figure 2 legend for explanation of the bar graphs.

Table 4. Brain regions showing significant differences between high-confidence true recognition (HC-true recognition) versus high-confidence false recognition (HC-false recognition) activity

			Talairach				
	H	BA	*x*	*y*	*z*	Voxels	*t*
HC-true recognition activity > HC-false recognition activity							
MTL: hippocampus/PHC	L	35	−19	−30	−18	39	4.27
	R	35	19	−30	−18	33	6.07
	R	34	23	−1	−13	17	4.26
HC-false recognition activity > HC-true recognition activity							
PFC							
Anterior	L	10/11	−23	51	−9	21	4.00
Ventrolateral	L	45/9	−38	9	31	38	4.27
Medial	B	8/32	0	21	44	78	4.55
Lateral PPC	L	40	−30	−43	37	30	4.66
	L	19	−30	−68	38	21	4.31
	R	39	34	−76	39	19	4.47

For abbreviations, see Table 2.

2002), and for high- than low-confidence responses (Chua et al., 2006; Moritz et al., 2006).

In the case of false recognition, MTL activity was greater for low- than high-confidence responses (Figs. 3*B*, 4*A*). A possible interpretation of this finding is that successful recollection of episodic memory details challenges the validity of false memories, making them less trustworthy. Consistent with this interpretation, participants can reduce false recognition by recalling information that is inconsistent with the occurrence of an illusory event (Brainerd et al., 1995; Rotello et al., 2000). Thus, the same successful recovery MTL mechanism that increases confidence in true recognition may reduce confidence in false recognition. Alternatively, the finding may reflect novelty detection (Knight, 1996; Tulving et al., 1996) or encoding operations during retrieval (Stark and Okado, 2003). If despite their conceptual familiarity critical lures appear novel in some way (e.g., perceptual appearance), this novelty could reduce confidence in recognizing them as old (supplemental analysis, available at www.jneurosci.org as supplemental material).

We interpret frontoparietal activity during high-confidence false recognition as reflecting mainly familiarity. This assumption is supported by behavioral evidence that false recognition lacks sensory and perceptual detail (Norman and Schacter, 1997; Hicks and Starns, 2005) and is reduced by manipulations that make exemplars more distinct [e.g., repetition (Arndt and Hirshman, 1998)]. The link between frontoparietal regions and familiarity is supported by lesion (Duarte et al., 2005) and functional neuroimaging studies. For example, fMRI studies have found that activity in several PFC and parietal regions is greater for know than remember responses (Henson et al., 1999), is greater for incorrect than correct source judgments (Cansino et al., 2002), is associated with a "feeling of knowing" (Maril et al., 2003), and increases linearly with perceived oldness (Daselaar et al., 2006; Montaldi et al., 2006). These familiarity-related activations were often very close to the ones associated with high-confidence false recognition in the present study. For example, Yonelinas et al. (2005) found familiarity-related activations less than 1 cm away than the ones we found in left BA 45 [Yonelinas et al. (2005): −48, 24, 21; present study: −42, 19, 20] and right BA 40 [Yonelinas et al. (2005): 39, −51, 36; present study: 42, −45, 45]. Thus, the present finding that frontoparietal activity mediated high-confidence false recognition in the present study is not incompatible with evidence that these regions contribute also to true recognition, as the latter reflects not only recollection but also familiarity.

The finding that high PFC activity was associated with high-confidence false recognition may appear inconsistent with evi-

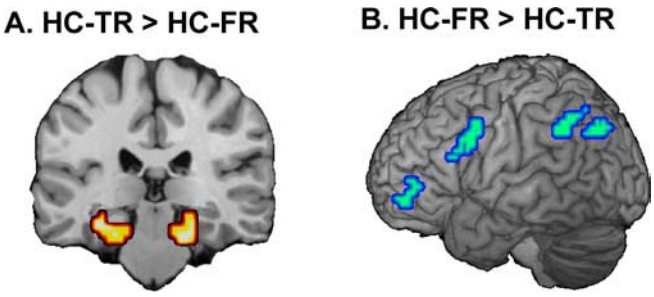

Figure 5. Activity within medial temporal lobes (***A***) was greater for high-confidence true recognition (HC-TR) than for high-confidence false recognition (HC-FR). Activity within a frontoparietal network (***B***) was greater for high-confidence false recognition than for high-confidence true recognition.

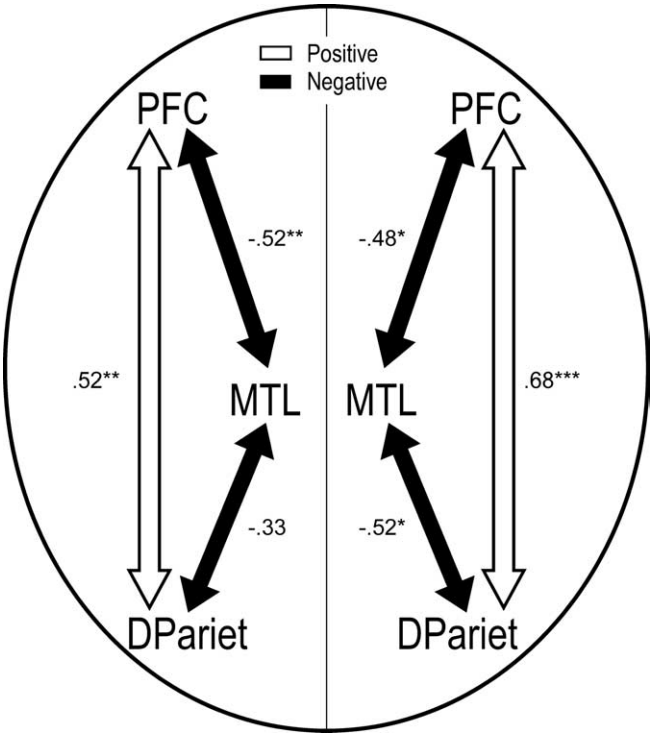

Figure 6. Schematic illustration of results from the correlation analyses among six ROIs. From each subject and a pair of ROIs, a within-subject Pearson correlation was computed across the mean parameter estimates of the four critical conditions. Numeric values are across-subjects means of these correlations. DPariet, Dorsal parietal cortex. Asterisks represent different levels of statistical significance (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

dence that frontal lesions sometimes increase false recognition (Melo et al., 1999; Verfaellie et al., 2004) (for review, see Gallo, 2006). However, this inconsistency is only apparent because we are not claiming that all PFC regions are associated with high-confidence false recognition. In fact, we found that activity in right dorsolateral/anterior PFC region (BA 46/10) was associated with low-confidence responses for both true and false recognition (Fig. 2*A*). fMRI studies have previously associated this region with "post-retrieval monitoring" (Henson et al., 2000) or general monitoring operations (Fleck et al., 2006), and there is evidence that damage to this region leads to increased false recognition (Schacter et al., 1996a). Thus, even if activity in some PFC regions contributes to false recognition, when regions involved in rejecting false memories are damaged, the net result may be an increase in false memories.

More generally, rather than top-down monitoring or control processes, we interpret the role of frontoparietal regions in high-confidence false recognition as reflecting the processing of bottom-up familiarity signals emanating from other brain regions, which were not detected in the present study. These regions include perirhinal regions that electrophysiological (for a review, see Brown and Aggleton, 2001) and fMRI (Gonsalves et al., 2005; Daselaar et al., 2006; Montaldi et al., 2006) studies have associated with familiarity, as well as anterior and posterior cortical regions that functional neuroimaging studies have associated with various forms of priming (for a review, see Henson, 2003). Most likely, different frontal and parietal regions are involved in processing different aspects of the familiarity signal. For example, whereas parietal regions may reflect attentional shifts toward the familiarity signal, left ventrolateral PFC (BA 45) may more specifically reflect processing and evaluation of the familiarity signal elicited by the critical lure.

Given the hypothesis that confidence in true and false memories depends on different neurocognitive mechanisms, an interesting question is how these mechanisms interact. Correlation analyses yielded negative coupling between MTL versus frontoparietal activity. A possible interpretation of this finding is that input from MTL is an important dimension regulating frontoparietal involvement in episodic memory retrieval (Moscovitch, 1992; Buckner and Wheeler, 2001; Rudy et al., 2005). When MTL yields abundant raw memory materials and recollection, there may be less need for familiarity processes mediated by the frontoparietal network. Thus, MTL and the frontoparietal network may play complementary roles during episodic retrieval.

In summary, the present study shows that despite overlaps previously reported between true and false recognition activations (Cabeza et al., 2001; Schacter et al., 1996b), when one focuses exclusively on high-confidence responses, the neural correlates of true and false memory are clearly different. High-confidence true recognition was associated with high MTL activity, whereas high-confidence false recognition was associated with high frontoparietal activity. These results indicate that confidence in true recognition is mediated primarily by a recollection-related MTL mechanism whereas confidence in false recognition reflects mainly a familiarity-related frontoparietal mechanism.

References

Arndt J, Hirshman E (1998) True and false recognition in MINERVA2: explanations from a global matching perspective. J Mem Lang 39:371–391.

Battig WF, Montague WE (1969) Category norms for verbal items in 56 categories: a replication and extension of the Connecticut norms. J Exp Psychol 80:1–46.

Brainerd CJ, Reyna VF (1990) Gist is the gist: the fuzzy-trace theory and new intuitionism. Dev Rev 10:3–47.

Brainerd CJ, Reyna VF (2002) Fuzzy-trace theory and false memory. Curr Dir Psychol Sci 11:164–169.

Brainerd CJ, Reyna VF, Kneer R (1995) False-recognition reversal: when similarity is distinctive. J Mem Lang 34:371–391.

Brown MW, Aggleton JP (2001) Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Nat Rev Neurosci 2:51–61.

Buckner RL, Wheeler ME (2001) The cognitive neuroscience of remembering. Nat Rev Neurosci 2:624–634.

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 USA 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.

Chua EF, Schacter DL, Rand-Giovannetti E, Sperling RA (2006) Understanding metamemory: neural correlates of the cognitive process and

subject level of confidence in recognition memory. NeuroImage 29:1150–1160.
Daselaar SM, Fleck MS, Cabeza R (2006) Triple dissociations in the medial temporal lobes: recollection, familiarity, and novelty. J Neurophysiol 96:1902–1911.
Deese J (1959) On the prediction of occurrence of particular verbal intrusions in immediate recall. J Exp Psychol 58:17–22.
Duarte A, Ranganath C, Knight RT (2005) Effects of unilateral prefrontal lesions on familiarity, recollection, and source memory. J Neurosci 25:8333–8337.
Eichenbaum H, Yonelinas AP, Ranganath C (2007) The medial temporal lobe and recognition memory. Annu Rev Neurosci 30:123–152.
Eldridge LL, Knowlton BJ, Furmanski CS, Bookheimer SY, Engel SA (2000) Remembering episodes: a selective role for the hippocampus during retrieval. Nat Neurosci 3:1149–1152.
Fleck MS, Daselaar SM, Dobbins IG, Cabeza R (2006) Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. Cereb Cortex 16:1623–1630.
Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 33:636–647.
Fortin NJ, Wright SP, Eichenbaum H (2004) Recollection-like memory retrieval in rats is dependent on the hippocampus. Nature 431:188–191.
Gallo DA (2006) Associative illusions of memory: false memory research in DRM and related tasks. New York: Psychology.
Gonsalves BD, Kahn I, Curran T, Norman KA, Wagner AD (2005) Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition. Neuron 47:751–761.
Henson RNA (2003) Neuroimaging studies of priming. Prog Neurobiol 70:53–81.
Henson RNA, Rugg MD, Shallice T, Josephs O, Dolan RJ (1999) Recollection and familiarity in recognition memory: an event-related functional magnetic resonance imaging study. J Neurosci 19:3962–3972.
Henson RNA, Rugg MD, Shallice T, Dolan RJ (2000) Confidence in recognition memory for words: dissociating right prefrontal roles in episodic retrieval. J Cogn Neurosci 12:913–923.
Hicks JL, Starns JJ (2005) False memories lack perceptual detail: evidence from implicit word-stem completion and perceptual identification tests. J Mem Lang 52:309–321.
Kim H, Cabeza R (2007) Differential contributions of prefrontal, medial temporal, and sensory-perceptual regions to true and false memory formation. Cereb Cortex 17:2143–2150.
Knight RT (1996) Contribution of human hippocampal region to novelty detection. Nature 383:256–259.
Lampinen JM, Meier CR, Arnal JD, Leding JK (2005) Compelling untruths: content borrowing and vivid false memories. J Exp Psychol Learn Mem Cogn 31:954–963.
Lindsay DS, Read JD, Sharma K (1998) Accuracy and confidence in person identification: the relationship is strong when witnessing conditions vary widely. Psychol Sci 9:215–218.
Maril A, Simons JS, Mitchell JP, Schwartz BL, Schacter DL (2003) Feeling-of-knowing in episodic memory: an event-related fMRI study. NeuroImage 18:827–836.
Melo B, Winocur G, Moscovitch M (1999) False recall and false recognition: an example of the effects of selective and combined lesions to the medial temporal lobe/diencephalon and frontal lobe structures. Cogn Neuropsychol 16:343–359.
Montaldi D, Spencer TJ, Roberts N, Mayes AR (2006) The neural system that mediates familiarity memory. Hippocampus 16:504–520.
Moritz S, Gläscher J, Sommer T, Büchel C, Braus DF (2006) Neural correlates of memory confidence. NeuroImage 33:1188–1193.
Moscovitch M (1992) Memory and working-with-memory: a component process model based on modules and central systems. J Cogn Neurosci 4:257–267.
Norman KA, Schacter DL (1997) False recognition in younger and older adults: exploring the characteristics of illusory memories. Mem Cognit 25:838–848.
Reyna VF, Brainerd CJ (1995) Fuzzy-trace theory: an interim synthesis. Learn Individ Differ 7:1–75.
Roediger HL, McDermott KB (1995) Creating false memories: remembering words not presented in lists. J Exp Psychol Learn Mem Cog 21:803–814.
Rotello CM, Macmillan NA, Tassel VA (2000) Recall-to-reject in recognition: evidence from ROC curves. J Mem Lang 43:67–88.
Rudy JW, Biedenkapp JC, O'Reilly RC (2005) Prefronal cortex and the organization of recent and remote memories: an alternative view. Learn Mem 12:445–446.
Schacter DL (2001) The seven sins of memory: how the mind forgets and remembers. Boston: Houghton Mifflin.
Schacter DL, Curran T, Galluccio L, Milberg WP, Bates JF (1996a) False recognition and the right frontal lobe: a case study. Neuropsychologia 34:793–808.
Schacter DL, Reiman E, Curran T, Yun LS, Bandy D, McDermott KB, Roediger III HL (1996b) Neuroanatomical correlates of veridical and illusory recognition memory: evidence from position emission tomography. Neuron 17:267–274.
Schacter DL, Verfaellie M, Pradere D (1996c) The neuropsychology of memory illusions: false recall and recognition in amnesic patients. J Mem Lang 35:319–334.
Stark CE, Okado Y (2003) Making memories without trying: medial temporal lobe activity associated with incidental memory formation during recognition. J Neurosci 17:6748–6753.
Talairach J, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. Stuttgart, Germany: Thieme.
Tulving E, Markowitsch HJ, Craik FIM, Habib R, Houle S (1996) Novelty and familiarity activations in PET studies of memory encoding and retrieval. Cereb Cortex 6:71–79.
Verfaellie M, Rapcsak SZ, Keane MM, Alexander MP (2004) Elevated false recognition in patients with frontal lobe damage is neither a general nor a unitary phenomenon. Neuropsychology 18:94–103.
Wagner AD, Shannon BJ, Kahn I, Buckner R (2005) Parietal lobe contributions to episodic memory retrieval. Trends Cogn Sci 9:445–452.
Wheeler ME, Buckner RL (2004) Functional-anatomic correlates of remembering and knowing. NeuroImage 21:1337–1349.
Yonelinas AP (2001) Consciousness, control and confidence: the three Cs of recognition memory. J Exp Psychol Gen 130:361–379.
Yonelinas AP, Kroll NE, Quamme JR, Lazzara MM, Sauve MJ, Widaman KF, Knight RT (2002) Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. Nat Neurosci 5:1236–1241.
Yonelinas AP, Otten LJ, Shaw KN, Rugg MD (2005) Separating the brain regions involved in recollection and familiarity in recognition memory. J Neurosci 25:3002–3008.
Yoon C, Feinberg F, Hu P, Gutchess AH, Hedden T, Chen H, Jing Q, Cui Y, Park DC (2004) Category norms as a function of culture and age: comparisons of item responses to 105 categories by Am and Chinese adults. Psychol Aging 19:379–393.