Differential Contributions of Prefrontal, Medial Temporal, and Sensory-Perceptual Regions to True and False Memory Formation

Hongkeun Kim[1,2] 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, NC 27708-0999, USA

The neural correlates of true memory formation (TMF) and false memory formation (FMF) were investigated using functional magnetic resonance imaging (fMRI). Using a parametric subsequent memory paradigm, encoding activity was analyzed as a function of whether it predicted subsequent hits to targets (TMF activity) or subsequent false alarms to critical lures (FMF activity). The fMRI analyses yielded 3 main findings. First, the left prefrontal cortex (PFC) was involved in both TMF and FMF activities. This finding is consistent with the evidence that semantic elaboration, which has been associated with left PFC, tends to enhance both true and false remembering. Second, the left posterior medial temporal lobes (MTLs) contributed to TMF but not to FMF activity. This finding is consistent with the notion that MTL is involved in the storage of a consciously, but not unconsciously, processed event. Third, late visual regions were engaged in both TMF and FMF activities, whereas early visual areas were involved primarily in TMF activity. This dissociation indicates that elaborative perceptual processing, but not basic sensory processing, contributes to false remembering. Taken together, the results suggest that FMF is an unintended consequence, or by-product, of elaborative semantic and visual encoding processes.

Keywords: false memory, fMRI, human memory, medial temporal lobe, prefrontal cortex, subsequent memory

Introduction

When our memory fails, we usually forget events that happened in the past. Sometimes, however, something more surprising occurs: We falsely remember events that never happened. Until very recently, functional neuroimaging studies of false memory (Schacter, Reiman, et al. 1996; Cabeza et al. 2001; von Zerssen et al. 2001; Okado and Stark 2003) have exclusively focused on the retrieval stage, with no explicit consideration of the encoding stage. However, behavioral evidence indicates that the encoding process is also critical in generating false remembering (Rhodes and Anastasi 2000; Gallo et al. 2001). Consistent with this evidence, a few recent functional neuroimaging studies (Gonsalves et al. 2004; Okado and Stark 2005) have successfully isolated neural correlates of certain types of false memory encoding. For example, in the study of Gonsalves et al. (2004), subjects sometimes claimed to have seen photos of objects they had imagined but not actually seen. Gonsalves et al. reported that this type of false remembering is related to activations in precuneus and inferior parietal regions during encoding. To the extent that activation of these regions increased the likelihood of later false remembering, these regions were engaged in "false memory formation" (FMF).

The goal of the present functional magnetic resonance imaging (fMRI) study was to compare brain regions involved in "true memory formation" (TMF) versus "false memory formation" (FMF). In particular, we investigated the question of whether and to what extent brain regions involved in TMF versus FMF overlap or separate from each other. In the few prior studies of false memory encoding (Gonsalves et al. 2004; Okado and Stark 2005), different encoding tasks were used to infer brain regions involved in TMF versus FMF, making it impossible or at best not highly meaningful to compare brain regions involved in TMF versus FMF. For example, in the study of Gonsalves et al. (2004), TMF was assessed by subsequent responses to word-plus-photo trials, whereas FMF was assessed by subsequent responses to word-only trials. Therefore, any difference in brain regions engaged in TMF versus FMF could simply reflect different encoding tasks. To address this issue, we investigated the neural activities reflecting TMF and FMF during the processing of the identical stimuli and, then, compared the 2 types of activities directly with each other. Moreover, whereas false remembering in prior studies reflected failure to remember the sources of studies items, we were interested in false memory for the items themselves, that is, in false remembering items that were never encountered.

The present encoding task was an adaptation of the Deese-Roediger-McDermott (DRM) paradigm (Roediger and McDermott 1995), which is a straightforward word-learning task. In each encoding trial of the present study, subjects studied a "mini word list" comprising 4 instances (e.g., horse, chicken, sheep, goat) of a semantic category (e.g., farm animal). At test, they performed an old/new recognition test with confidence ratings that included studied words (e.g., horse, chicken) as well as nonstudied words from studied categories (e.g., cow, pig). Using the subsequent memory procedure, we calculated 2 different measures for each encoding trial: 1) how many (and how confidently) studied words were later remembered (subsequent hit rate: a measure of TMF) and 2) how many (and how confidently) nonstudied semantic associates were later falsely remembered (subsequent false alarm rate: a measure of FMF). Based on these 2 measures, we conducted a parametric study of fMRI signals (Friston 1997) at the encoding phase. These analyses informed us, based on identical encoding trials, which brain regions show "subsequent true memory effects" and which brain regions show "subsequent false memory effects."

We focused on 3 regions of interest (ROIs): the left ventrolateral prefrontal cortex (PFC), the medial temporal lobes (MTLs), and the sensory-perceptual cortex. TMF activity in these regions has been frequently reported in prior fMRI studies using the subsequent memory paradigm (e.g., Wagner et al. 1998; Kirchhoff et al. 2000; Prince et al. 2005). First, left ventrolateral PFC is involved in controlled semantic processing (e.g., Gabrieli et al. 1998; Buckner et al. 1999), and, that, in

associative false memory paradigms, controlled semantic processing enhances not only true memory for studied words but also false memory for nonstudied semantic associates (e.g., Rhodes and Anastasi 2000; Gallo et al. 2001). Thus, we predicted that left ventrolateral PFC would contribute to both TMF and FMF activities. Second, MTL region is involved in the storage of a consciously, but not unconsciously, processed event (Moscovitch 1992; Moscovitch and Winocur 2002), and hence, its activity is more likely to vary with memory for items that were actually experienced during the encoding episode. Thus, we predicted that MTL would contribute to mainly TMF activity. Third, turning to the sensory-perceptual cortex, we predicted a functional dissociation between early and late visual areas. Early visual areas are sensitive to sensory properties of individual stimuli (e.g., Van Essen and Deyoe 1995), and hence, they are likely to show mainly TMF activity. In contrast, late visual areas are also sensitive to the semantic/associative nature of stimuli (e.g., Büchel et al. 1998), and hence, they are likely to show not only TMF but also FMF activity.

To our interest, a recent study (Kubota et al. 2006) using 2-channel near-infrared spectroscopy (NIRS) system in combination with the standard DRM procedure investigated the role of PFC activity in false memory encoding. This study reported that left PFC hemodynamic increase during encoding is associated with later false remembering, consistent with PFC contribution to FMF activity. However, the functional significance of this finding in relation to other brain regions remains to be clarified because this study monitored only PFC activity. Moreover, NIRS technology, while good in temporal resolution, is limited in spatial resolution and measurement stability. Thus, the present approach based on the whole-brain fMRI procedure should provide new and more comprehensive insights into the question of whether and to what extent brain regions involved in TMF versus FMF overlap or separate from each other.

Materials and Methods

Subjects

Sixteen young adults (9 females; age range 18–31) 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.

Behavioral Paradigm

The present encoding task was an adaptation of the DRM paradigm (Roediger and McDermott 1995). The materials were 72 categorical 6-word lists selected from category norms (Battig and Montague 1969; Yoon et al. 2004). Each list consisted of the 6 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 "4-legged animal" was not included (for full listing of stimulus words, see Supplementary Material available online). To make sure minimal associative overlap between the categories, we have examined the probability that the critical lures would be generated as an associative response to the other categories (e.g., the probability that "cow" would be generated as an associative response to "wild animal"). The associative response probability was less than 1% in 10211 out of

10224 (72 × 2 × 71) examined and less than 5% in the remaining 13. Some semantic overlap between the categories at the superordinate level (e.g., animal) was inevitable for obtaining a sufficient number of categories. However, the overlap may not be a critical problem given minimal overlap at the ordinate or subordinate level.

The study phase was a single scan consisting of 72 trials/lists. Each encoding screen simultaneously showed a category name at the top and its 4 instances below the category name in column format (see Fig. 1). In an additional 10 "catch" trials, only 3 of the 4 instances belonged to the category. Each encoding screen was presented for 4 s, followed by a fixation cross for 2 s. The subjects' task was to decide whether all 4 or only 3 instances belonged to the category. They responded by pressing one of the 2 keys in a response box using their right hand. A fixation period, ranging from 1.5 to 4.5 s, was interspersed across trials to "jitter" the onset times of trials and allow event-related fMRI analyses. The words were displayed in colors to promote the encoding of sensory/perceptual information (Cabeza et al. 2001). In a given trial, all 4 words were displayed in the same color, but 5 different colors were alternatively used across trials.

The test phase, which started approximately 10 min after completion of the study phase, consisted of 6 scans. The fMRI data from these scans are not reported here and will be the focus of a separate publication. 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. All words in the test phase were displayed in white color against black background. Subjects responded by pressing one of 4 keys according to whether the word was judged to be "sure old," "unsure old," "unsure new," or "sure new."

Using the subsequent memory procedure, we calculated 2 different measures for each encoding trial: 1) how many (and how confidently) studied words were later remembered (subsequent hit rate: a measure of TMF) and 2) how many (and how confidently) nonstudied semantic associates were later falsely remembered (subsequent false alarm rate: a measure of FMF). Each high-confidence hit response was assigned 1 point and each low-confidence hit was assigned 0.5 point, yielding a 0–4 range for the subsequent hit measure. Each high-confidence false alarm was assigned 2 points and each low-confidence false alarm was assigned 1 point, also yielding a 0–4 range for the subsequent false alarm measure. This scoring scheme reflected our reasoning that strength of TMF (or FMF) for the encoding lists is reflected in the number of hits (or false alarms) as well as in degree of confidence associated with hits (or false alarms). Based on subsequent hit and subsequent false alarm measures, we conducted a parametric study of fMRI signals at the encoding phase (see below). It should be noted that parametric analyses as implemented in SPM2 are scale invariant in so far as the scales are linearly related. Thus, assignment of [1, 2] to low- versus high-confidence false alarms yields numerically identical results to assignment of [0.5, 1] to low- versus high-confidence false alarms.

The fMRI Methods

Magnetic resonance image 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 T_2-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 T_1-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.

The fMRI analyses focused on data from the encoding phase. Image processing and analyses were performed using SPM2 software (www. fil.ion.ucl.ac.uk/spm/). After discarding the first 6 volumes, the functional images were slice-timing corrected and motion corrected and then spatially normalized to the Montreal Neurological Institute templates
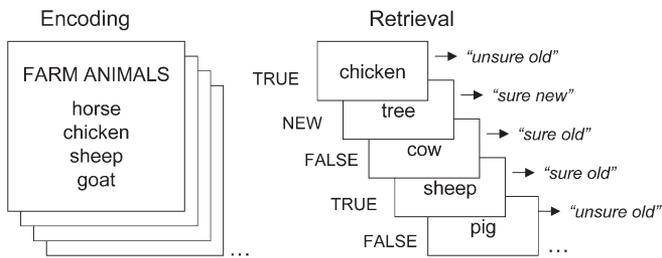
Figure 1. Behavioral paradigm. The encoding task was a category judgment task. The retrieval task was an old–new recognition test with confidence ratings that included studied words (True words) and nonstudied words from studied categories (False words) and nonstudied, unrelated words (New words).

implemented in SPM2. The coordinates were later converted to Talairach and Tournoux's (1988) space. Subsequently, the functional images were spatially smoothed using an 8-mm isotropic Gaussian kernel and resliced to a resolution of 3.75 mm³ 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). Two separate parametric analyses were performed. In one analysis, the height of the modeled HRF was parametrically modulated by the subsequent hit measure (i.e., the HRF was multiplied by a linear increase function of subsequent hit measure) and in the other analysis parametrically modulated by the subsequent false alarm measure. The general linear model, as implemented in SPM2, was used to model the effects of the parametric regressors and other confounding effects (e.g., head movement and magnetic field drift). Filler trials and trials on which encoding responses were "incorrect" were modeled by a separate regressor, but not considered in the analyses. For each participant, statistical parametric maps pertaining to the parametric regressors were identified and subsequently integrated across subjects using a random-effects model. The resulting t-map images were examined with a statistical threshold set at $P < 0.001$, uncorrected, and a minimum cluster size of at least 10 contiguous voxels. These analyses informed us, based on identical encoding trials, which brain regions show TMF activity (i.e., positive covariation between encoding trial activations and later hit rate for words from those trials) and which brain regions show FMF activity (i.e., positive covariation between encoding trial activations and later false alarm rate for semantic associates of words from those trials).

In order to identify similarities and differences between TMF and FMF activities, we performed 2 different analyses. First, to identify regions showing both TMF and FMF activations, the 2 t maps were inclusively masked, each with a statistical threshold of $P < 0.01$, uncorrected, (joint probability = $0.01 \times 0.01 = 0.0001$). This procedure yielded an activation map containing only those voxels that showed both TMF and FMF activities. The voxels that showed significant differences between the 2 t maps (see below) were subsequently eliminated from the activation map. Second, to identify regions showing differences between TMF and FMF activities, these activations were directly compared with each other (i.e., TMF > FMF and FMF > TMF). In these standard pairwise contrasts, the significance threshold was set at $P < 0.001$, uncorrected. In both contrasts, to ensure that activation differences reflected activations in the target condition rather than deactivations in the control condition, the contrast was inclusively masked with activity in the target condition at $P < 0.01$. In both inclusive masking and direct-contrast analyses, to further reduce the risk of false positive activations, a spatial extent threshold of 10 voxels was also employed.

Our scoring scheme assumes that 2 low-confidence responses are equivalent to one high-confidence response. This assumption is heuristic and may not be strictly valid. To address this issue, we performed an additional parametric subsequent analysis that is based on only high-confidence subsequent hits or false alarms. In this analysis, the height of the modeled HRF was parametrically modulated separately by the number of high-confidence subsequent hits and by the number of high-confidence subsequent false alarms. This analysis involved 11 participants excluding the 5 participants with sparse number (<15) of high-confidence false alarms. The resulting t-map images were exam-

ined with a significance threshold set at $P < 0.005$, uncorrected, and a spatial extent threshold of 10 voxels. Significance and spatial thresholds used for follow-up inclusive masking and direct-contrast analyses were identical to the comparable analyses described above, except that the significance threshold for direct contrasts was set at $P < 0.005$. A lower significance threshold compared with the original analysis was used to adjust for lowered statistical power.

Results

Behavioral Performance

Category judgment at the study phase was highly accurate (mean, 95% correct). The mean reaction time (RT) for correct trials was 2588 ms (standard deviation = 285). Behavioral performance at the test phase is shown in Figure 2. Combined across high- and low-confidence responses, the proportion of hits for True words (73%) was significantly greater than the proportion of false alarms for False words (49%; $t_{15} = 7.84$, $P < 0.001$), which in turn was significantly greater than the proportion of false alarms for New words (19%; $t_{15} = 10.25$, $P < 0.001$). Thus, False words elicited a more robust false recognition relative to New words. There was no significant correlation between RT and subsequent hit rate across trials (computed within each subject and then averaged across subjects; Mr [Mean r] = 0.084, $t_{15} = 2.10$, $P > 0.05$) or between RT and subsequent false alarm rate (Mr = 0.011, $t_{15} = 0.66$, $P > 0.50$). Thus, TMF and FMF activations and the differences between them cannot be attributed to differences in processing time during the study phase. There was a modest, but significant, correlation between subsequent hit and false alarm rates across trials (Mr = 0.20, $t_{15} = 3.76$, $P < 0.01$). This correlation provides behavioral evidence that a common factor, namely, controlled elaborative processing, contributes to effective TMF as well as effective false memory formation.

The fMRI Findings

Table 1 separately lists brain regions showing FMF and TMF activations, as revealed by the parametric analysis of encoding trials. FMF activity was found in several regions, including ventrolateral PFC and late visual cortex (occipitotemporal and occipitoparietal cortices). TMF activity was found in similar regions and additionally in MTL (posterior parahippocampal cortex) and early visual cortex (occipital pole). To more systematically compare TMF and FMF activities, inclusive masking and direct-contrast analyses were performed (see Materials and Methods). The results of these analyses are shown in Table 2. There were 3 main findings.

First, as predicted, left ventrolateral PFC showed both TMF and FMF activities. As illustrated by the left panel of Figure 3, left ventrolateral PFC activity predicted both subsequent hit rate and subsequent false alarm rate. A dorsomedial PFC region also showed both TMF and FMF activities. There was no PFC region that showed significant difference between TMF and FMF activities.

Second, also as predicted, MTL showed only TMF activity. Within left posterior MTL (parahippocampal cortex), TMF activity was greater than FMF activity, and the latter was not significantly different from zero ($t_{15} = 1.13$, $P > 0.25$). As illustrated by the right panel of Figure 3, left parahippocampal activity predicted subsequent hit rate but not subsequent false alarms. No MTL region showed a sign of FMF activity even when a more lenient threshold ($P < 0.01$) was employed. To test for
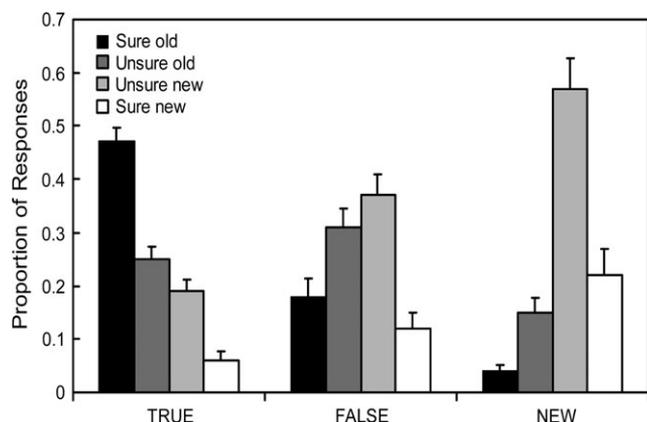
Figure 2. Behavioral performance at the retrieval phase. Error bars represent 1 standard error.

Table 1

Brain regions showing activations associated with TMF and FMF

Contrast	H	BA	Talairach			Voxels	T value
Region			x	y	z		
FMF activity							
Ventrolateral PFC	L	45/44	−46	15	10	15	4.05
	L	44/6	−46	2	31	19	4.82
Occipitotemporal cortex	L	37/19	−38	−60	−6	28	5.32
	L	19	−23	−70	−6	12	4.15
	R	37/19	27	−56	−10	54	5.89
Occipitoparietal cortex	L	39/19	−30	−76	25	31	5.33
	R	39/19	34	−69	28	11	5.38
	R	18	19	−84	12	10	4.31
Parietal cortex	L	40	−60	−43	27	11	4.45
TMF activity							
MTL: posterior parahippocampal cortex	L	27/30	−8	−33	−2	21	5.29
Occipital pole	L	18/17	−19	−96	5	80	5.81
	R	18/17	19	−99	5	37	4.75
Ventrolateral PFC	L	45/44	−46	30	9	61	5.54
	L	44/6	−42	1	28	46	6.14
Dorsomedial PFC	L	6/8	−4	11	56	13	5.68
Occipitotemporal cortex	L	37/19	−38	−60	−13	67	5.00
	R	19/18	11	−83	−19	83	5.89
Occipitoparietal cortex	L	39/19	−30	−72	31	20	4.87
	L	19	−34	−84	14	29	5.83
Thalamus	L	—	−4	−4	8	28	5.89

Note: H, hemisphere; L, left; R, right.

Table 2

Brain regions showing similarities and differences between TMF and FMF activations

Contrast	H	BA	Talairach			Voxels	T value
Region			x	y	z		
Both TMF and FMF activity							
Ventrolateral PFC	L	45/44	−42	30	9	40	4.91
	L	44/6	−42	1	28	63	6.14
Dorsomedial PFC	L	6/8	−4	21	48	23	4.23
Occipitotemporal cortex	L	37/19	−38	−60	−13	119	5.00
	R	37/19	30	−52	−10	111	5.93
Occipitoparietal cortex	L	39/19	−30	−72	31	55	4.87
	R	19	30	−73	25	11	3.70
TMF activity > FMF activity							
MTL: posterior parahippocampal cortex	L	27/30	−8	−33	−2	18	5.21
Occipital pole	L	18/17	−23	−100	5	11	4.37
	R	18/17	19	−96	5	28	5.50
Retrosplenial cortex	L	30	−15	−55	6	13	4.40
Inferior temporal cortex	L	20	−52	−45	−10	14	4.43
Thalamus	L	—	−8	−4	8	18	5.63
FMF activity > TMF activity							
No reliable activations							

Note: For abbreviations, see Table 1.

a dissociation of TMF and FMF activities in left ventrolateral PFC versus MTL, ROI analyses were performed. ROIs were functionally defined as the significant clusters in the group analyses. Hence, the sizes of the ROIs corresponded to the sizes of clusters in Table 2. The mean parameter estimate across all significant voxels was extracted for TMF and FMF activities, respectively, from each subject and ROI. Confirming the dissociation, a 2 (PFC vs. MTL) by 2 (TMF vs. FMF) analysis of variance (ANOVA) yielded a significant interaction ($F_{1,15} = 4.73$, $P < 0.05$).

Third, confirming our prediction regarding visual cortex, we found dissociation between late and early visual regions. Whereas late visual regions (bilateral occipitotemporal and occipitoparietal; Brodmann area [BA] 37/31/19) showed both TMF and FMF activities, early visual regions (bilateral occipital pole; BA 18/17) showed only TMF activity. In the early visual regions, TMF activity was greater than FMF activity, and the latter was not significantly different from zero (left occipital pole, $t_{15} = 0.24$, $P > 0.80$; right occipital pole, $t_{15} = 1.26$, $P > 0.20$). As

illustrated by Figure 4, in late visual regions, encoding activity predicted subsequent hit rate as well as subsequent false alarm rate, whereas in early visual regions, encoding activity predicted only subsequent hit rate. To test for a dissociation of TMF and FMF activities in early versus late visual cortex, we conducted a 2 (early vs. late) by 2 (TMF vs. FMF) ANOVA and averaged across hemispheres. Confirming the occipital pole versus occipitotemporal dissociation, a 2 (occipital pole vs. occipitotemporal) by 2 (TMF vs. FMF) ANOVA yielded a significant interaction ($F_{1,15} = 6.28$, $P < 0.05$). Confirming the occipital pole versus occipitoparietal dissociation, a 2 (occipital pole vs. occipitoparietal) by 2 (TMF vs. FMF) ANOVA also yielded a significant interaction ($F_{1,15} = 5.65$, $P < 0.05$).

Beyond our predictions, a few regions (left inferior temporal cortex, retrosplenial, thalamus) showed greater TMF than FMF activity. FMF activity was significantly greater than zero in the inferior temporal cortex ($t_{15} = 2.25$, $P < 0.05$), but not in retrosplenial cortex ($t_{15} = 0.97$, $P > 0.30$) or the thalamus ($t_{15} = 1.11$, $P > 0.20$). Importantly, no brain region showed greater FMF than TMF activity. Thus, brain regions showing significant FMF activity were a subset of those showing TMF activity, namely, PFC and late visual areas.

Our 3 predictions were also confirmed in the parametric analysis based on only high-confidence responses. Results of this analysis are listed in Table 3 (for separate listing of brain regions showing TMF and FMF activations, see Supplementary Material available online). As may be seen in comparison of Tables 2 and 3, the results were largely consistent with the previous analysis based on both high- and low-confidence responses. First, left ventrolateral PFC (BA 45/44) showed both TMF and FMF activities. Second, MTL (left hippocampus) showed TMF activity but not FMF activity. Finally, late visual regions (left occipitotemporal and occipitoparietal cortices) showed both TMF and FMF activities, whereas early visual areas (left occipital pole) showed mainly TMF activity. In this analysis, one small region, located in right inferior parietal cortex (BA 40), showed greater FMF than TMF activity. In this region, FMF activity was significantly positive ($t_{10} = 4.39$, $P < 0.01$), whereas TMF activity was significantly negative ($t_{10} = -3.51$, $P < 0.01$). Thus, higher
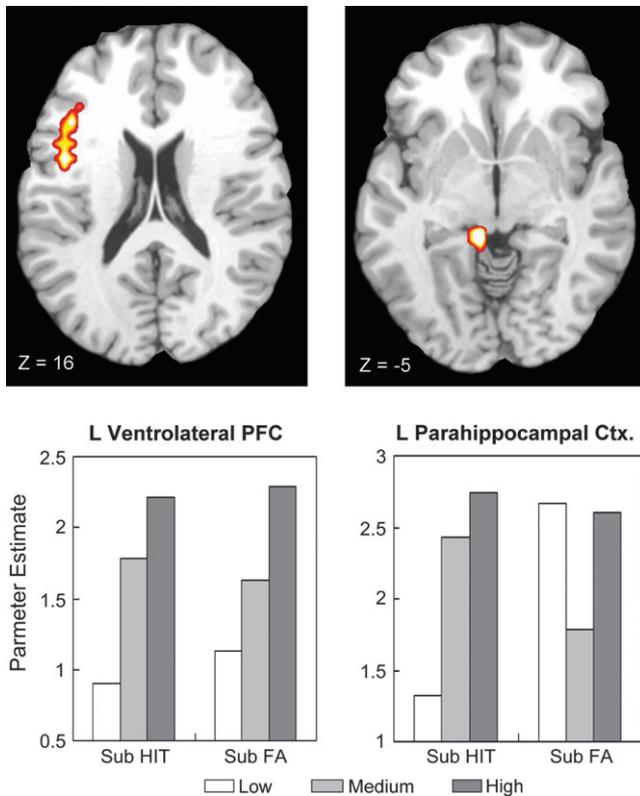
Figure 3. Differential contributions of PFC and MTL regions to TMF and FMF. The encoding activity in the left ventrolateral PFC region (BA 44; $x = -45$, $y = 12$, $z = 17$) predicted both subsequent hit rate and subsequent false alarm rate. By contrast, the encoding activity in the left parahippocampal region (BA 30; $x = -8$, $y = -37$, $z = -1$) predicted subsequent hit rate but not subsequent false alarm rate. For illustration of parametric effects, encoding trials were classified into high ($x \geq 3.5$), medium ($3 \geq x \geq 2.5$), and low subsequent hits and high ($x \geq 3$), medium ($2 \geq x \geq 1$), and low subsequent false alarms in a separate analysis. Bars in the graphs represent the parameter estimates of these trial types. Note that separate analyses of these trial types were only used for illustration purposes but not for the statistical effects reported in the text. Ctx., cortex; Sub, subsequent.

activation of this region was associated with subsequent higher false alarm rate as well as subsequent lower hit rate.

Discussion

Clarifying the neural bases of memory errors and distortions is critical for understanding the mechanisms of normal memory function, as well as for its clinical and legal implications. The present study investigated the question of whether and to what extent brain regions involved in TMF versus FMF activity overlap or are different from each other. Though a few prior fMRI studies (Gonsalves et al. 2004; Okado and Stark 2005) investigated false memory encoding, they did not directly compare true and false memory encoding and did not investigate item false memory but only source memory. A recent study using NIRS technology (Kubota et al. 2006) investigated item false memory but monitored only PFC activity. To address these issues, we directly compared TMF and FMF activities during the processing of the same stimuli using the whole-brain fMRI procedure and investigated false memory for items that were never encountered.

The present study yielded 3 main findings. First, left ventrolateral PFC regions were involved in both TMF and FMF
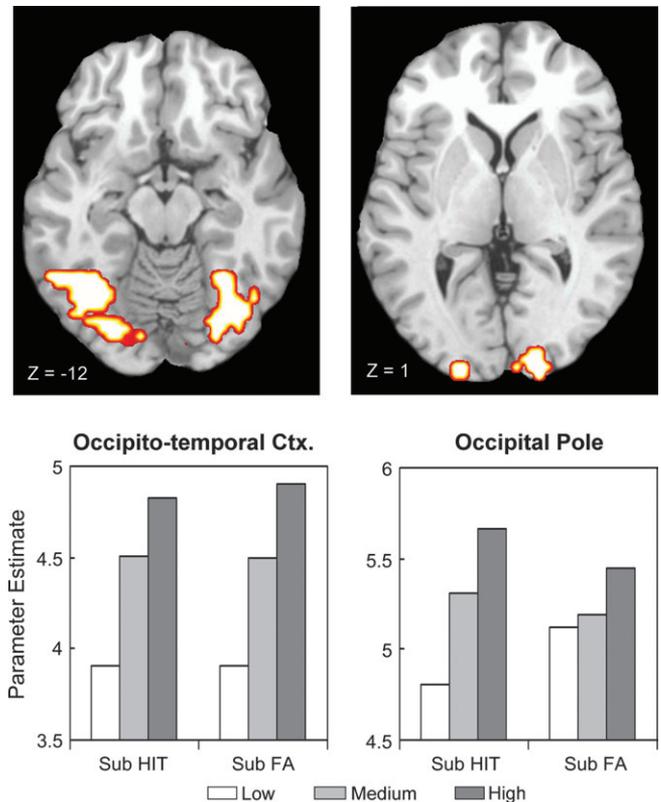

Figure 4. Differential contributions of late and early visual regions to TMF and FMF. The encoding activity in the bilateral occipitotemporal regions (BA 37; $x = -34$, $y = -60$, $z = -10$) predicted both subsequent hit rate and subsequent false alarm rate. By contrast, the encoding activity in the bilateral occipital pole regions (BA 18; $x = -19$, $y = -96$, $z = 8$) predicted subsequent hit rate but not subsequent false alarm rate. For more explanation, see Figure 3.

activities. Second, left MTL regions were engaged only in TMF activity. Third, late visual areas contributed to both TMF and FMF activities, whereas early visual areas mediated only TMF activity. These 3 findings were confirmed in analyses involving both high- and low-confidence responses as well as in analyses involving only high-confidence responses, suggesting that they do not reflect scoring "artifacts." These 3 findings are discussed in separate sections below.

Left Ventrolateral PFC Showed Both TMF and FMF Activities

Behavioral studies of false memory have found that semantic elaboration increases not only the likelihood of correctly remembering studied items but also the probability of falsely remembering nonstudied associates (e.g., Rhodes and Anastasi 2000; Gallo et al. 2001). Functional neuroimaging studies have linked semantic elaboration to activations in left ventrolateral PFC (Dolan and Fletcher 1997; Gabrieli et al. 1998; Buckner et al. 1999; Reber et al. 2002). Combining these 2 lines of evidence, we predicted that left ventrolateral PFC would show both TMF and FMF activities, and our prediction was confirmed. The finding of TMF activity in left ventrolateral PFC (BAs 44, 45) is consistent with several fMRI studies of subsequent true memory (e.g., Wagner et al. 1998; Kirchhoff et al. 2000; Prince et al. 2005). The present study demonstrates that left ventrolateral PFC contributes not only to subsequent true but also to subsequent false memory. A similar finding has been recently reported by Kubota

Table 3

Brain regions showing similarities and differences between TMF and FMF activations based on only high-confidence hits and false alarms

Contrast	H	BA	Talairach			Voxels	T value
Region			x	y	z		
Both TMF and FMF activity							
Ventrolateral PFC	L	45/44	−42	27	16	61	4.06
	R	45	42	27	16	14	4.44
Occipitotemporal cortex	L	37	−38	−60	−13	22	4.11
Occipitoparietal cortex	L	39/19	−27	−80	39	31	5.24
TMF activity > FMF activity							
Hippocampus	L	—	−26	−15	−9	18	3.27
Occipital pole	L	17/18	−19	−96	5	47	5.65
	R	17/18	19	−100	9	15	5.73
Occipital cortex	L	18	−34	−88	1	27	4.31
Inferior temporal cortex	L	20	−52	−45	−10	16	6.30
Thalamus	L	—	−8	−14	8	13	4.32
Precentral cortex	L	6	−45	1	28	12	3.61
Cerebellum	R	—	8	−83	−19	19	5.72
FMF activity > TMF activity							
Inferior parietal cortex	R	40	56	−42	47	15	5.78

Note: For abbreviations, see Table 1.

et al. (2006), based on NIRS technology. Activation in left ventrolateral PFC in the present study extended to posterior part (BA 6). Activation in this region may reflect working memory component of the present encoding task such as sequential processing of category name and its 4 instances (Cabeza et al. 2002). More generally, although we attribute the TMF/FMF overlap in left ventrolateral PFC to semantic elaboration, there is no direct measure of semantic elaboration at the time of encoding in the present study. Thus, further research is required to specify the nature of the shared process.

The essential feature of semantic elaboration is the integration of incoming information with preexistent semantic knowledge. This integration process, while strengthening the formation of "veridical memory traces," may also contribute to the creation of "illusory memory traces." This idea fits well with 2 popular theoretical accounts of false memory encoding. According to the "spreading activation view" (e.g., Underwood 1965; Roediger et al. 2001), in the semantic memory network, activation flows from list items to critical lures, and associative activation of the latter is subsequently misattributed to actual encounters. According to the "fuzzy trace view" (e.g., Brainerd and Reyna 1990; Schacter, Verfaellie et al. 1996), studying a list of associates leads to formation of "verbatim traces," which contain item-specific information, as well as "gist traces," containing the general topic of the list. Both theories predict that the formation of illusory memory traces (spreading activation, gist traces) should be enhanced by semantic elaboration. The present results suggest that this process is mediated by same left PFC regions that mediate the formation of veridical memory traces.

MTL Showed TMF Activity but Not FMF Activity

There is abundant evidence that MTL is critical for declarative (conscious) memory but not for nondeclarative (nonconscious) memory (for a review, see Squire et al. 2004). Accordingly, MTL has been described as a memory module whose specific domain is consciously apprehended information (Moscovitch 1992). In the present study, each list item was individually attended and consciously experienced, whereas critical lures were not. Although it is conceivable that critical lures are sometimes spontaneously generated while freely processing lists of 15

converging associates, it is very unlikely that this would happen when a task is performed on mini lists of 4 category instances. Thus, assuming that participants are consciously aware of list items but not of critical lures, we predicted that MTL would show TMF activity but not FMF activity. Confirming this prediction, a left posterior parahippocampal region showed significant TMF activity. Finding TMF activity in posterior MTL is consistent with the results of several subsequent memory fMRI studies (e.g., Wagner et al. 1998; Kirchhoff et al. 2000; Reber et al. 2002). Analyses involving only high-confidence hits, but not analyses involving both high- and low-confidence hits, found TMF activity in hippocampus. This finding is consistent with the view that hippocampal activity is more related to subsequent recollection (i.e., high-confidence hit) than to subsequent familiarity (Davachi et al. 2003; Ranganath et al. 2004).

No MTL region, whether hippocampal or parahippocampal, was found that significantly contributed to FMF activity. Contribution of PFC, but not MTL regions, to FMF activity fits well with the notion that PFC is involved in "working with memory" traces, whereas MTL is engaged in creating "raw memory traces" (Moscovitch 1992; Moscovitch and Winocur 2002). Nevertheless, whether and to what extent MTL is engaged in FMF activity may be determined in part by specific parameters of the DRM paradigm. Unlike the standard DRM paradigm in which 15 words of converging associates are presented (e.g., Roediger and McDermott 1995), in the present paradigm, a categorical list composed of only 4 words were presented. This adaptation of the standard paradigm was critical for obtaining a sufficient number of encoding trials for fMRI analysis and to identify both TMF and FMF activities during the processing of the same stimuli. Critical lures are likely to be more consciously processed in the standard DRM paradigm with 15 words of converging associates than in the present paradigm. For example, overt or conscious generation of critical lures may be more frequent in the standard DRM paradigm. Thus, with the standard DRM paradigm, MTL may significantly contribute to FMF activity though it still contributes more to TMF than to FMF activity.

Dissociation between Early and Late Visual Regions

Prior subsequent memory studies have shown the involvement of sensory-perceptual regions in TMF (e.g., Wagner et al. 1998; Kirchhoff et al. 2000; Otten et al. 2002). In visual processing areas, posterior–anterior progression of more elaborative processing is well established (e.g., Van Essen and Deyoe 1995; Büchel et al. 1998). Early visual processing is sensitive to sensory properties of individual stimuli, and hence, they are likely to show mainly TMF activity. In contrast, late higher order visual processing, like elaborative semantic processing, may contribute to TMF as well as FMF activities. Based on these considerations, we predicted that late visual areas would be engaged in both TMF and FMF, whereas early visual areas would be engaged mainly in TMF. Consistent with this prediction, we found a functional dissociation between late visual regions (bilateral occipitotemporal and occipitoparietal), which showed both TMF and FMF activities, and early visual regions (bilateral occipital pole), which showed only TMF.

Though we have emphasized bottom-up processing of stimulus information, subsequent memory effects in sensory-perceptual regions might also reflect top–down modulation effects from PFC regions (Fernández and Tendolkar 2001; Paller and Wagner 2002). The top–down mechanism might reflect general attentional effects, modulating levels of attention given

to stimulus processing (Hopfinger et al. 2000). A top–down mechanism initiated by PFC may also modulate posterior perceptual/mnestic representations, such as integration with existing representations (Kirchhoff et al. 2000). Given the current evidence that PFC processing contributes to both TMF and FMF, the top–down modulation of posterior perceptual/mnestic representations might promote TMF as well as FMF. Interestingly, a recent fMRI study comparing true versus false memory retrieval activity using abstract shape stimuli (Slotnick and Schacter 2004) reported a highly compatible dissociation between late versus early visual areas as the present study. This study found comparable levels of true and false memory retrieval activity in late visual areas but only true memory retrieval activity in early visual areas. The authors suggested that this dissociation during retrieval might reflect the reactivation of true and false memory traces formed in early versus late visual areas during encoding. Our data provide direct support for this hypothesis.

Conclusion

Our study provides the first direct whole-brain contrast between neural activity associated with TMF versus FMF. The results indicate that encoding activity in left ventrolateral PFC and late visual areas contributes to both TMF and FMF. By contrast, encoding activity in a left posterior MTL and early visual areas contributes mainly to TMF. The finding of brain regions associated with FMF activity provides direct support for the view that false memory is due, at least in part, to processes that take place at the time of encoding, such as the spreading activation view (e.g., Underwood 1965; Roediger et al. 2001) and the fuzzy trace view (e.g., Brainerd and Reyna 1990; Schacter, Verfaellie, et al. 1996). The finding could be also accommodated by false memory theories postulating changes at retrieval, to the extent that these changes are, at least in part, a consequence of changes during encoding. The analyses based on only high-confidence responses revealed a right posterior parietal region associated with greater FMF than TMF activity. Except this small region, false memory formation was mediated largely by a subset of the brain regions involved in TMF that support elaborative encoding, namely, PFC and late visual areas. This finding should be taken cautiously because FMF effect may have less statistical power than TMF effect. However, taken together, the present findings suggest that FMF is largely an unintended consequence, or by-product, of elaborative semantic and visual encoding processes Schacter (2001).

Notes

Address correspondence to Prof. Hongkeun Kim, Department of Rehabilitation Psychology, Daegu University, Daemyung3-Dong 2288, Nam-Gu, Daegu 705-714, South Korea. Email: hongkn@daegu.ac.kr.

References

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.

Büchel C, Price C, Friston K. 1998. A multimodal language region in the ventral visual pathway. Nature. 394:274–277.

Buckner RL, Kelley WM, Petersen SE. 1999. Frontal cortex contributes to human memory formation. Nat Neurosci. 2:311–314.

Cabeza R, Dolcos F, Graham R, Nyberg L. 2002. Similarities and differences in the neural correlates of episodic memory retrieval and working memory. Neuroimage. 16:317–330.

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.

Davachi L, Mitchell JP, Wagner AD. 2003. Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. Proc Natl Acad Sci USA. 100:2157–2162.

Dolan RJ, Fletcher PC. 1997. Dissociating prefrontal and hippocampal function in episodic memory encoding. Nature. 388:582–585.

Fernández G, Tendolkar I. 2001. Integrated brain activity in medial temporal and prefrontal area predicts subsequent memory performance: human declarative memory formation at the system level. Brain Res Bull. 55:1–9.

Friston KJ. 1997. Imaging cognitive anatomy. Trends Cogn Sci. 1:21–27.

Gabrieli JD, Poldrack RA, Desmond JE. 1998. The role of left prefrontal cortex in language and memory. Proc Natl Acad Sci USA. 95:906–913.

Gallo DA, Roediger HL, McDermott KB. 2001. Associative false recognition occurs without strategic criterion shifts. Psychon Bull Rev. 8:579–586.

Gonsalves B, Reber PJ, Gitelman DR, Parrish TB, Mesulam M-M, Paller KA. 2004. Neural evidence that vivid imagining can lead to false remembering. Psychol Sci. 15:655–660.

Hopfinger JB, Buonocore MH, Mangun GR. 2000. The neural mechanisms of top-down attentional control. Nat Neurosci. 3:284–291.

Kirchhoff BA, Wagner AD, Maril A, Stern CE. 2000. Prefrontal-temporal circuitry for episodic encoding and subsequent memory. J Neurosci. 20:6173–6180.

Kubota Y, Toichi M, Shimizu M, Mason RA, Findling RL, Yamamoto K, Calabrese JR. 2006. Prefrontal hemodynamic activity predicts false memory—a near-infrared spectroscopy study. Neuroimage. 31:1783–1789.

Moscovitch M. 1992. Memory and working-with-memory: a component process model based on modules and central systems. J Cogn Neurosci. 4:257–267.

Moscovitch M, Winocur G. 2002. The frontal cortex and working with memory. In: Stuss DT, Knight RT, editors. Principles of frontal lobe function. New York: Oxford Univesity Press. p. 188–209.

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

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

Otten LJ, Henson RNA, Rugg MD. 2002. State-related and item-related neural correlates of successful memory encoding. Nat Neurosci. 5:1339–1344.

Paller KA, Wagner AD. 2002. Observing the transformation of experience into memory. Trends Cogn Sci. 6:93–102.

Prince SE, Daselaar SM, Cabeza R. 2005. Neural correlates of relational memory: successful encoding and retrieval of semantic and perceptual associations. J Neurosci. 25:1203–1210.

Reber PJ, Siwiec RM, Gitelman DR, Parrish TB, Mesulam M-M, Paller KA. 2002. Neural correlates of successful encoding identified using functional magnetic resonance imaging. J Neurosci. 22:9541–9548.

Ranganath C, Yonelinas AP, Cohen MX, Dy CJ, Tom SM, D'Esposito M. 2004. Dissociable correlates of recollection and familiarity within the medial temporal lobes. Neuropsychologia. 42:2–13.

Rhodes MG, Anastasi JS. 2000. The effects of a levels-of-processing manipulation on false recall. Psychon Bull Rev. 7:158–162.

Roediger HL, Balota DA, Watson JM. 2001. Spreading activation and the arousal of false memories. In: Roediger HL, Nairne JS, Neath I, Surprenant AM, editors. The nature of remembering: essays in honor

of Robert G. Crowder. Washington (DC): American Psychological Association. p. 95-115.

Roediger HL, McDermott KB. 1995. Creating false memories: remembering words not presented in lists. J Exp Psychol Learn Mem Cogn. 21:803-814.

Schacter DL. 2001. The seven sins of memory. New York: Houghton Mifflin.

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, Verfaellie M, Pradere D. 1996. The neuropsychology of memory illusions: false recall and recognition in amnestic patients. J Mem Lang. 35:319-334.

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

Squire LR, Stark CE, Clark RE. 2004. The medial temporal lobe. Annu Rev Neurosci. 27:279-306.

Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain. Stuttgart (Germany): Thieme Verlag.

Underwood BJ. 1965. False recognition produced by implicit verbal responses. J Exp Psychol. 70:122-129.

Van Essen DC, Deyoe EA. 1995. Concurrent processing in the primate visual cortex. In: Gazzaniga MS, editor. The cognitive neurosciences. Cambridge (MA): MIT Press. p. 383-400.

von Zerssen GC, Mecklinger A, Opitz B, von Cramon. 2001. Conscious recollection and illusory recognition: an even-related fMRI study. Eur J Neurosci. 13:2148-2156.

Wagner AD, Schacter DL, Rotte M, Koutstaal W, Maril A, Dale AM, Rosen BR, Buckner RL. 1998. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science. 281: 1188-1191.

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 American and Chinese adults. Psychol Aging. 19:379-393.