In Search of Recollection and Familiarity Signals in the Hippocampus

Peter E. Wais[1], Larry R. Squire[1,2], and John T. Wixted[1]

Abstract

■ fMRI studies of recognition memory have often been interpreted to mean that the hippocampus selectively subserves recollection and that adjacent regions selectively subserve familiarity. Yet, many of these studies have confounded recollection and familiarity with strong and weak memories. In a source memory experiment, we compared correct source judgments (which reflect recollection) and incorrect source judgments (often thought to reflect familiarity) while equating for old–new memory strength by including only high-confidence hits in the analysis. Hippocampal activity associated with both correct source judgments and incorrect source judgments exceeded the activity associated with forgotten items and did so to a similar extent. Further, hippocampal activity was greater for high-confidence old decisions relative to forgotten items even when source decisions were at chance. These results identify a recollection signal in the hippocampus and may identify a familiarity signal as well. Similar results were obtained in the parahippocampal gyrus. Unlike in the medial temporal lobe, activation in prefrontal cortex increased differentially in association with source recollection. ■

INTRODUCTION

Dual-process theories of recognition memory hold that two distinct memory processes, recollection and familiarity, underlie one's ability to recognize an item as having been previously encountered (Mandler, 1980). Recollection involves the retrieval of contextual detail associated with the test item, whereas familiarity involves simply knowing that the item was encountered before. According to one view (the high-threshold/signal-detection model), high confidence in a recognition decision strongly implies that the decision was based on recollection, whereas lower confidence necessarily implies that the decision was based on familiarity (Yonelinas, 1994). A closely related view holds that decisions accompanied by a "remember" response are based on recollection, whereas decisions accompanied by a "know" response are based on familiarity. A common finding in the functional magnetic resonance imaging (fMRI) literature is that hippocampal activity associated with highly confident responses (or remember responses) exceeds the activity associated with forgotten items, whereas the activity associated with less confident responses (or know responses) does not (Ranganath et al., 2004; Eldridge, Knowlton, Furmanski, Bookheimer, & Engel, 2000). This pattern of findings has often been interpreted to mean that the hippocampus selectively subserves recollection.

An alternative view holds that the level of confidence associated with a recognition decision is related to memory strength and that memory strength reflects varying degrees of both recollection and familiarity (Wixted, 2007). According to this view, recognition decisions made with low confidence reflect small contributions of recollection and familiarity, whereas decisions made with high confidence reflect large contributions of recollection and familiarity. On average, recollection will typically be associated with higher confidence than familiarity. Nevertheless, the confidence distributions associated with recollection and familiarity extend across the full range of memory strength. Accordingly, confidence ratings per se cannot be used to disentangle the two processes. This view further holds that remember–know judgments are tantamount to confidence ratings (Dunn, 2004, 2008). If so, then know judgments do not reflect strong, familiarity-based memories that are devoid of recollection. Instead, compared to remember judgments, they reflect weaker memories that are associated with lesser degrees of confidence and lesser degrees of recollection (Wais, Mickes, & Wixted, 2008). According to this alternative view, fMRI studies that have relied on confidence ratings or the remember–know procedure to identify the neural correlates of recollection and familiarity involve a memory strength confound. That is, it is not clear whether the reported effects in these studies were due to differences between recollection and familiarity or due to differences in memory strength (Wais, 2008; Squire, Wixted, & Clark, 2007).

A different approach to separating recollection and familiarity involves the source memory procedure. Correctly identifying an item as old and with correct source

---

[1]University of California, San Diego, La Jolla, CA, [2]Veterans Affairs Medical Center, San Diego, CA

information is thought to identify a recollection-based decision, whereas correctly identifying an item as old but with incorrect source information is thought to identify a familiarity-based decision (Ranganath et al., 2004). A typical finding from fMRI studies is that hippocampal activity associated with source-correct responses exceeds the activity associated with forgotten items, whereas the activity associated with source-incorrect responses does not (Kensinger & Schacter, 2006; Weis et al., 2004). This pattern has also been interpreted to mean that the hippocampus selectively subserves recollection (Eichenbaum, Yonelinas, & Ranganath, 2007).

Yet, source memory studies analyzed in this way also confound the strength of memory (as indicated by confidence in the old–new decision) with source-correct and source-incorrect decisions. Specifically, confidence is typically higher for old–new decisions that are subsequently associated with correct source judgments than for old–new decisions that are subsequently associated with incorrect source judgments (Gold et al., 2006; Slotnick & Dodson, 2005). This difference in old–new confidence suggests that comparisons between source-correct and source-incorrect decisions involve relatively strong, recollection-based decisions and relatively weak, familiarity-based decisions. This strength confound is problematic because the distinction between recollection and familiarity is independent of memory strength. Moreover, the relationship between the BOLD signal in the hippocampus and the neural activity that underlies memory strength is unknown and may be nonlinear (Johnson, Muftuler, & Rugg, 2008; Squire et al., 2007). If so, then differences in activity between conditions associated with strong versus intermediate-strength memories (e.g., source correct vs. source incorrect, or remember vs. know) may be more detectable in the hippocampus than differences in activity between conditions associated with intermediate-strength versus very weak memories (e.g., source incorrect vs. misses, or know vs. misses). In light of these considerations, it is important to equate for memory strength if the objective is to compare activity in the hippocampus associated with recollection and familiarity (Wais, 2008; Squire et al., 2007; Wixted, 2007).

To equate memory strength across conditions, we made use of confidence ratings. Evidence that confidence ratings provide a valid measure of memory strength is provided by past work showing that the confidence expressed in an old–new decision is strongly related to the accuracy of that decision (e.g., Mickes, Wixted, & Wais, 2007). In our fMRI study, we equated memory strength for source-correct versus source-incorrect judgments by using only old decisions that were made with high confidence, and the question of interest was whether hippocampal activity measured at retrieval would be differentially elevated for source-correct judgments even under these conditions. If the typical pattern is found even after controlling for memory strength (i.e., if differentially elevated hippocampal activity is detected for

the source-correct condition), then it would clearly weigh against the suggestion that prior fMRI studies have been compromised by a memory strength confound (Wais, 2008; Squire et al., 2007) and would support the notion that the hippocampus does not play a role in familiarity-based decisions. However, if hippocampal activity associated with source-correct and source-incorrect decisions are both elevated once they are equated for memory strength, then the strength-confound hypothesis—and the idea that the hippocampus does play a role in familiarity-based decisions—would remain viable. Such an outcome would not definitively identify a familiarity signal in the hippocampus because one might suppose that strong memories accompanied by incorrect source decisions do not reflect familiarity but instead reflect the undetected recollection of details unrelated to the source question. Nevertheless, if equally strong source-correct and source-incorrect decisions are both associated with elevated hippocampal activity, it would indicate that prior arguments against a familiarity signal in the hippocampus may be less definitive then they have been taken to be, and it would underscore the importance of controlling for memory strength in future investigations into the neuroanatomical basis of recollection and familiarity.

METHODS

Participants

Informed consent was obtained from 18 students (6 women) at the University of California, San Diego. All participants were right-handed. Two participants who did not score above chance levels for their source memory judgments were excluded from further analysis.

Stimuli

Two hundred forty English nouns were selected from the MRC Psycholinguistic Database with the following constraints: word frequency of 50 to 300, length of 5 to 12 letters, and two to four phonemes. The words were randomly divided into one list of 192 targets and another list of 48 foils.

Behavioral Procedure and Data Analysis

Participants studied a list of words presented on a desktop computer. Words were presented in six blocks of 32 words each. Words in each block were randomly ordered for each participant. During each 2.5-sec trial in the study session, participants responded to one of two contextual-cue questions posed for each word, the *common* question or the *discuss* question (common: does the word describe something you expect to encounter in a typical week?; or discuss: does the word describe something you would discuss with a close friend?). The contextual-cue question was the same for

all trials within a block so that participants studied the target words as alternating blocks with either the "common" or "discuss" cues. Participants were instructed to read each word, enter a yes or no answer to the cued question, and to remember the word, including its contextual cue, for a memory test during their subsequent scanning session.

The memory test for each participant was conducted in the MRI scanner approximately 3 hr after the study session. Participants saw test items in six blocks of 40 words (5.0 sec per word). Each test block included 32 targets from the study session, plus eight foils. The test items were presented in a random order to each participant and intermixed with trials from an odd–even digit task described below (fMRI Scanning Parameters, Procedure, and Data Analysis). For each word presented in the test phase, participants first gave a confidence judgment as to whether the word was old or new (1 = definitely new, 2 = probably new, 3 = maybe new, 4 = maybe old, 5 = probably old, 6 = definitely old) and then gave a confidence judgment for their source decision (1 = definitely discuss, 2 = probably discuss, 3 = maybe discuss, 4 = maybe common, 5 = probably common, 6 = definitely common). For clarity, we will refer to old–new confidence ratings in terms of the 1-through-6 numerical scale, bu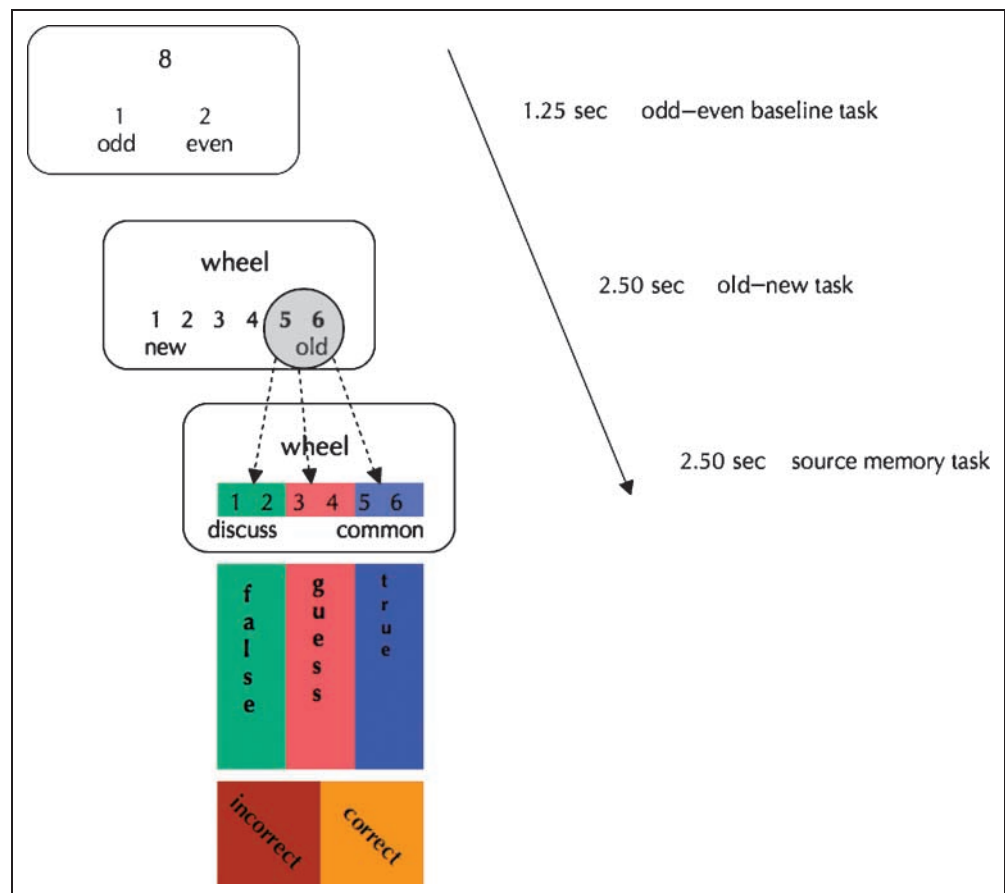t we will refer to source confidence ratings in terms of correct and incorrect decisions that were made with low, medium, or high confidence (e.g., correct source confidence ratings of 1 for "discuss" items and 6 for "common" items will both be referred to as correct source decisions made with high confidence).

Because we were concerned that our task instructions should be received as logical, participants were instructed not to enter a source judgment for words endorsed as new. Nevertheless, the task demand remained constant for the old–new test in each trial, and our analysis of the fMRI results was based on the data collected during the old–new test. The old–new scale and the source decision scale were each presented for 2.5 sec beneath the test word on each trial (Figure 1). In order to facilitate fMRI analysis (see below), participants also performed an odd–even classification task (Stark & Squire, 2001) on trials randomly intermixed with the memory task. For this baseline task, the digits 1 to 9 were presented for 1.25 sec each in blocks of 2, 4, 6, or 12.

fMRI Scanning Parameters, Procedure, and Data Analysis

Imaging was carried out in a GE Signa Excite 3-T scanner at the Center for Functional MRI (University of California, San Diego). Functional images were acquired using a

Figure 1. Example of test procedure and behavioral analysis: The stimulus "wheel" was studied under the *common* contextual-cue condition and endorsed as a high-confidence hit (5 or 6) in the old–new recognition task. For the subsequent source memory task, a response of 1 or 2 indicates false source memory, a response of 3 or 4 indicates source guesses, and a response of 5 or 6 indicates true source memory (source responses 1, 2, and 3 are incorrect and source responses 4, 5, and 6 are correct).

gradient-echo, echo-planar, T2*-weighted pulse sequence (TR = 2.5 sec, TE = 30, 90° flip angle, bandwidth = 250 MHz, FOV = 22 cm). Forty-two slices covering the whole brain were acquired perpendicular to the long axis of the hippocampus (matrix size = 64 × 64, slice thickness = 5 mm). Following six functional runs, high-resolution structural images were acquired using a T1-weighted, fast spoiled gradient-echo (FSPGR) pulse sequence (TE = 3.1, 12° flip angle, FOV = 25 cm, 172 slices, 1 mm slice thickness, matrix size = 256 × 256).

Between word presentations, participants were given 0, 2, 4, 6, or 12 trials of the 1.25-sec baseline task that served to jitter the MR signal acquired for subsequent deconvolution of the hemodynamic response function (hrf). For each participant, the fMRI data were partitioned into 10 categories (see Results for an explanation of the trials in each category). The first seven categories were based on the old–new confidence ratings provided on each trial: (a) correct old responses to targets (i.e., hits) that were rated 6; (b) hits that were rated 5; (c) high-confidence hits (rated 5 or 6) that were subsequently associated with correct source decisions; (d) high-confidence hits (5s or 6s) that were subsequently associated with incorrect source decisions; (e) misses (targets rated 1, 2, 3, or in some cases, 4); (f) false alarms (foils rated 4, 5, or 6); and (g) correct rejections (foils rated 1, 2, or 3). Additionally, the high-confidence hits (those associated with an old–new confidence rating of 5 or 6) were subdivided into three additional categories based on the confidence ratings for the source decision: (h) "true source" decisions (correct source decisions made with medium or high confidence on the source confidence scale); (i) "source guesses" (correct and incorrect source decisions made with low confidence); and (j) "false source" decisions (incorrect source decisions made with medium or high confidence).

For each of the 10 categories, a hemodynamic response (relative to the baseline condition) was estimated for the 25 sec following the presentation of the word by using signal deconvolution with the AFNI suite of programs (Cox, 1996). Data analysis was then based on the area under the hrf from 0 to 15 sec following the presentation of the word (at about 15 sec, the hrf returned to baseline). The anatomical scans and the fMRI data were normalized to the template of the Talairach brain (Talairach & Tournoux, 1988). Functional data were resampled to 2 × 2 × 2 mm and blurred with a 4-mm FWHM Gaussian kernel. These data were used for the whole-brain analysis. For the analysis of medial temporal lobe (MTL) activity, the ROI–large deformation diffeomorphic metric mapping (ROI–LDDMM) alignment method (Miller, Beg, Ceritoglu, & Stark, 2005) was used to improve cross-participant alignment and increase statistical reliability (Kirwan, Jones, Miller, & Stark, 2007).

Voxel-based t tests (threshold of $p < .001$, two-tailed) were then carried out as group analyses across all 16 participants for both the whole brain and MTL analyses

based on the area under the hrf for contrasts of interest (described below). Monte Carlo simulations were then used to correct for multiple comparisons and to determine how large a cluster of voxels was needed in order to be statistically significant ($p < .05$). The coordinates of all of the regions of activity we identified that were statistically significant (p-corrected < .05) are listed in Table S1 as supplementary information.

RESULTS

Behavioral Results

The participants demonstrated good old–new recognition memory for the target words ($d' = 1.93 \pm 0.15$; and $79 \pm 2\%$ correct for the old–new response). Words studied in the "common" cue condition were recognized as readily as words studied in the "discuss" cue condition ($d' = 1.98 \pm 0.17$ vs. $d' = 1.89 \pm 0.16$). The distribution of responses for hits and false alarms across the six-level old–new confidence scale revealed a bias to respond "old" (Figure 2). As a result, responses to targets given a confidence level 4 (maybe old) were at chance accuracy overall ($56 \pm 6\%$ correct), whereas responses to foils given a confidence level 3 (maybe new) were much more accurate ($80 \pm 4\%$ correct). The accuracy of responses to targets given a confidence rating of 5 or 6 was also high ($71 \pm 6\%$ correct and $87 \pm 3\%$ correct, respectively). The old–new accuracy calculations for each rating included scores of participants who had at least five observations for that rating (e.g., at least 5 ratings of 3).

Overall source accuracy was $64 \pm 2\%$ correct, and source d' was 0.71 ± 0.10. As has been observed in prior studies, source accuracy varied as a function of confidence in the old–new decision for targets. Source accuracy was highest for items that received an old–new
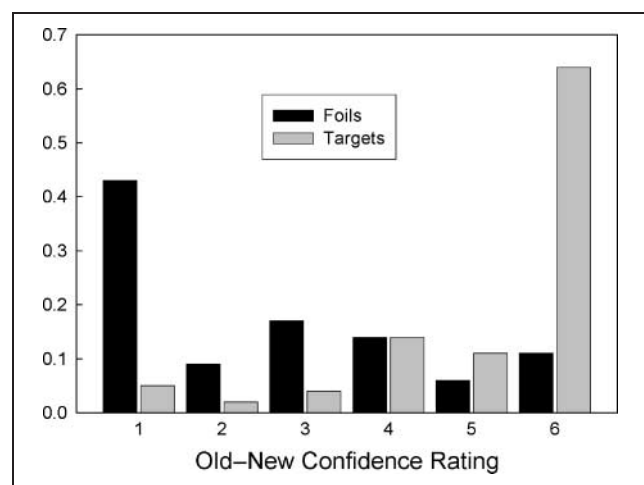
Figure 2. Proportion of responses to targets and foils for each confidence level in the old–new task, $n = 16$, mean recognition $d' = 1.93$ (0.15).

confidence rating of 6 (67% correct, significantly above chance, $p < .05$), next highest for items that received an old–new confidence rating of 5 (63% correct, significantly above chance, $p < .05$), and lowest for items that received an old–new confidence rating of 4 (50% correct).

For target items that were correctly declared to be old (i.e., targets that received a rating of 4, 5, or 6), we computed the mean old–new confidence rating separately depending on whether the subsequent source decision was correct or incorrect. The mean old–new confidence associated with source-correct decisions (Table 1) was significantly higher than the mean old–new confidence associated with source-incorrect decisions ($p < .02$). Thus, our results exhibit the typical memory strength confound that has been observed in prior source memory studies (Gold et al., 2006; Slotnick & Dodson, 2005).

To eliminate this confound, we combined old–new hits that had been given confidence ratings of 5 or 6 and then divided those responses into source-correct and source-incorrect categories (decisions made with a confidence rating of 6 yielded too few observations to detect MTL activity reliably). A possible difficulty with this approach is that it could introduce the strength confound we sought to avoid (i.e., the old–new confidence ratings for incorrect source judgments could be lower, on average, than the old–new confidence ratings for source-correct judgments). However, in our study, this did not occur. As shown in Table 1, for old–new decisions made with a confidence rating of 5 or 6, the mean old–new confidence for source-incorrect decisions (5.82) was virtually identical to the mean old–new confidence for source-correct decisions (5.84), and the small difference between them did not approach significance. Thus, a comparison of activity associated with these source-correct and source-incorrect decisions is not confounded with memory strength (as measured by confidence in the old–new decision).

To allow for a further analysis of neural activity associated with source recollection, we also partitioned correct old decisions in a more fine-grained manner. Specifically, instead of separating them into two categories (i.e., source-correct and source-incorrect), we separated them into three categories: true source judgments (i.e., correct source judgments made with medium or high confidence), source guesses (source judgments

made with low confidence, whether correct or incorrect), and false source judgments (incorrect source judgments made with medium or high confidence). Table 1 shows that there is an old–new strength confound if all of the correct old decisions are used, so we included only old decisions made with a confidence rating of 5 or 6 (which eliminated the confound). For these high-confidence old decisions, 44% were followed by true source judgments, 37% were followed by source guesses, and 19% were followed by false source judgments. The source accuracy of the guesses was $55 \pm 2\%$, which did not differ significantly from chance ($p = .12$). Thus, source information was, in fact, absent when items were recognized as old with high confidence and the source judgment was a guess.

In the fMRI analyses described below, we compared the activity associated with either two or three categories of source judgments (equated for old–new confidence) against the activity associated with forgotten items. Typically, a target item is considered to be forgotten if it is incorrectly declared to be new (i.e., if it receives a confidence rating of 1, 2, or 3). However, as indicated above, many of the participants in our experiment exhibited a liberal response bias such that ratings of 4 were as likely to be given to targets as to foils (Figure 2). In that case, a confidence rating of 4 for a target indicates a forgotten item as well. In all of the analyses described below, we considered old–new confidence ratings of 1, 2, 3, or 4 to reflect forgotten words for the 11 of 16 participants who exhibited no better than chance accuracy when responding 4 (maybe old), and we considered responses of 1, 2, or 3 to denote forgotten words for the remaining five participants whose old–new confidence ratings of 4 were associated with greater than chance accuracy.

fMRI Results

Our objective was to measure activity associated with correct old decisions accompanied by source recollection and correct old decisions that were not accompanied by source recollection after eliminating the typical memory strength confound. To that end, we first measured activity associated with source-correct and source-incorrect judgments using only those old decisions that were made with relatively high confidence (i.e., 5 or 6).

Table 1. Mean Old–New Confidence (*SEM*) for Different Source Accuracy Conditions and Different Levels of Source Confidence When All Hits Were Included (Old–New Confidence Ratings of 4, 5, or 6) and When Only Relatively High-confidence Hits Were Included (Old–New Confidence Ratings of 5 or 6)

	By Two Source Categories		By Three Source Categories		
Old–New Ratings Partitioned	*Source Correct*	*Source Incorrect*	*True Source*	*Source Guess*	*False Source*
4, 5, 6	5.62 (0.08)*	5.44 (0.09)*	5.73 (0.07)†	5.37 (0.10)†,‡	5.58 (0.07)‡,†
5, 6	5.84 (0.04)	5.82 (0.04)	5.85 (0.05)	5.81 (0.04)	5.81 (0.04)

Values that share symbols (e.g., *) differ significantly from each other ($p < .05$).

In one voxel-based *t* test (thresholded at $p < .001$), we found that, in the left hippocampus, activity associated with source-correct decisions was significantly greater than the activity associated with forgotten items (Figure 3). In the same region of the left hippocampus, a second, independent voxel-based *t* test revealed that activity associated with source-incorrect decisions was also significantly greater than activity associated with forgotten items (Figure 3). We also directly contrasted source-correct decisions versus source-incorrect decisions, but no statistically significant regions ($p < .001$) were identified in the MTL. These results suggest that increased activation in the left hippocampus is associated with increased strength of memory (i.e., high-confidence hits versus forgotten items) and that activity does not differ whether or not the decision is accompanied by successful source recollection.

The analysis summarized in Figure 3 assumes that source-correct decisions were associated with recollection and that source-incorrect decisions were associated with the absence of recollection (and may have been based on familiarity). However, both of these categories included source memory judgments made with low, medium, and high confidence (i.e., "maybe," "probably,"
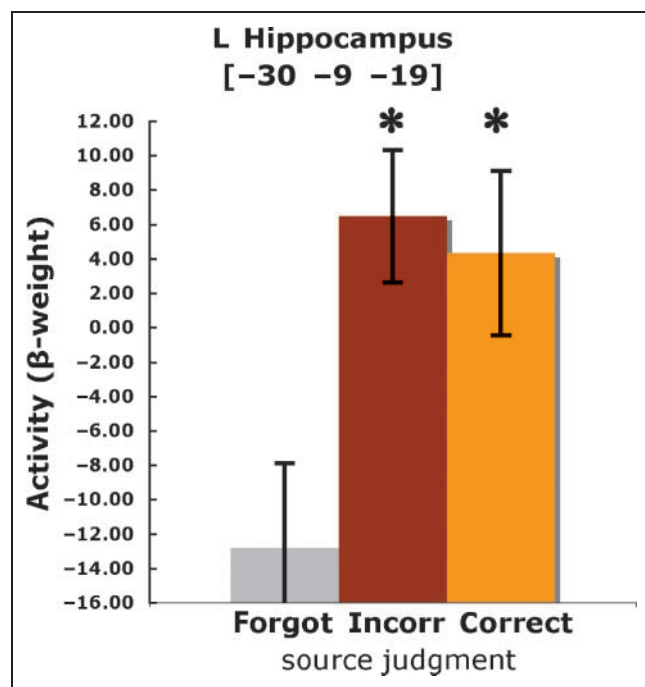

Figure 3. Activation identified for separate contrasts of correct source judgments versus forgotten items and incorrect source judgments versus forgotten items. To equate for memory strength, the source-correct and source-incorrect data were based on old decisions made with high confidence. Error bars for the two source categories represent the *SEM* of the difference scores for each comparison, whereas the error bar for the forgotten items represents the root mean square of the *SEM* values associated with the two individual comparisons (* denotes a significant difference relative to forgotten items, *p*-corrected < .05).

or "definitely"). Thus, some source-incorrect decisions were made with high source confidence and may have reflected false recollection. Conceivably, the hippocampal activity associated with source-incorrect decisions in the analysis described above reflects false recollection, as has been reported in other paradigms (Schacter & Slotnick, 2004; Cabeza, Rao, Wagner, Mayer, & Schacter, 2001). To address this issue, we used voxel-based *t* tests (thresholded at $p < .001$) to contrast activity associated with true source judgments, source guesses, and false source judgments against the activity associated with forgotten items. Once again, in order to eliminate a strength confound that would otherwise exist, only old decisions made with high confidence (5 or 6) were included in the following analyses. For each identified cluster (*p*-corrected < .05), signal data were also extracted for the other source conditions.

Regions in the posterior hippocampus, bilaterally, exhibited significantly greater activity associated with true source decisions than for forgotten items (Figure 4A). In signal data extracted from this cluster, the activity levels for source guesses and false source decisions were both numerically higher than the level associated with forgotten items and did not differ significantly from the level associated with true source decisions. In a separate contrast, activity associated with source guesses was significantly greater than the activity associated with forgotten items in the right posterior hippocampus (Figure 4B). The location of this cluster was virtually identical to the location of the cluster in the right hippocampus identified by the comparison between true source decisions and forgotten items (shown in Figure 4A). The activity levels extracted from this cluster for true source decisions and false source decisions were numerically higher than the level associated with forgotten items and did not differ from the level associated with true source decisions.

The findings presented in Figure 4A and B may indicate both a recollection and a familiarity signal in the hippocampus when confidence in memory is strong. Evidence for a recollection signal comes from the fact that the activity associated with true source memories significantly exceeded the activity associated with forgotten items (Figure 4A). Such an interpretation is consistent with much prior evidence, suggesting that the hippocampus plays an important role in recollection. For example, Manns, Hopkins, Reed, Kitchener, and Squire (2003) found that selective hippocampal lesions were associated with clear deficits in recall performance, which is generally assumed to be based exclusively on recollection.

Evidence for a possible familiarity signal in the hippocampus comes from the fact that the activity associated with source guesses (using only old decisions made with high confidence) significantly exceeded the activity associated with forgotten items (Figure 4B). This finding has not been previously reported, perhaps because when old decisions made with lower confidence are included in the analysis, rather than only old decisions made with
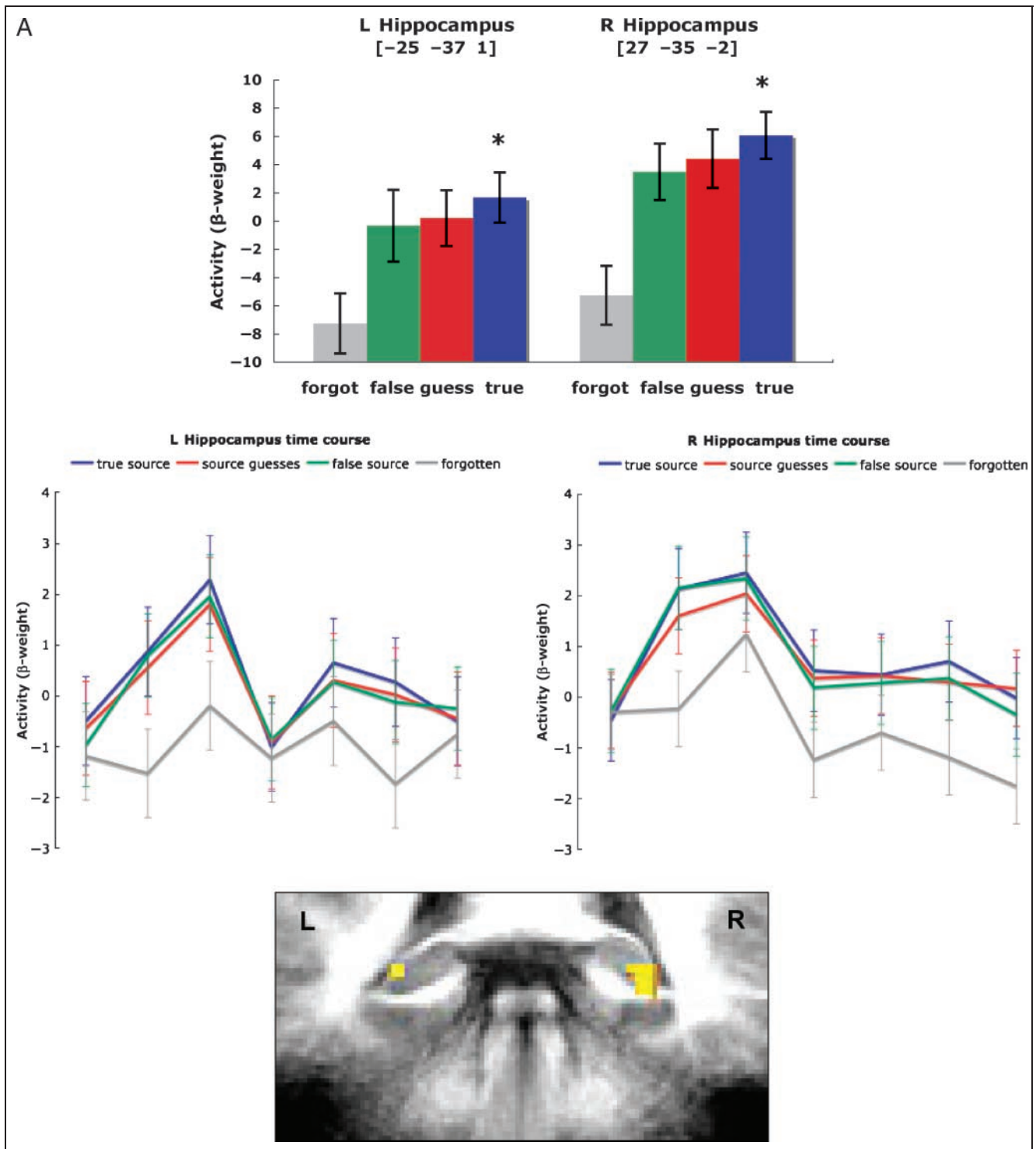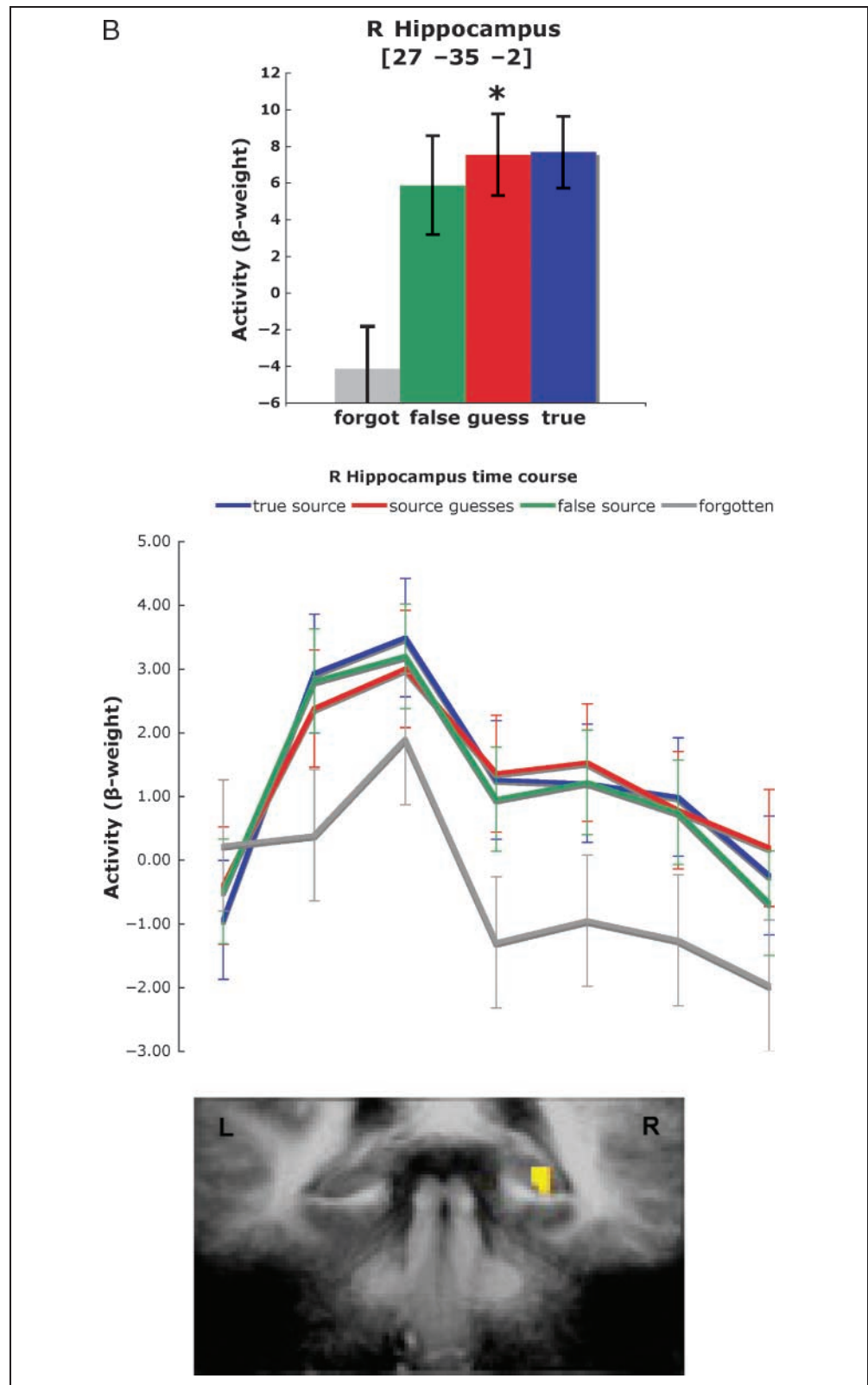
Figure 4. (A) Activation identified in the hippocampus for the contrast of true source judgments versus forgotten items. (B) Activation identified in the hippocampus for the contrast of source guesses versus forgotten items. Bar graphs show signal correlated with forgotten items and with recognition decisions that were accompanied by true, guess, or false source judgments. Also illustrated are the time courses of the hrf correlated with these results and MRI sections locating the activation on averaged anatomical images for our participants. Error bars for the three source categories represent the *SEM* of the difference scores for each comparison (relative to forgotten items), whereas the error bar for the forgotten items represents the root mean square of the *SEM* values associated with the three individual comparisons (* denotes a significant difference relative to forgotten items, *p*-corrected < .05).

Figure 4. (*continued*)

high confidence, then memories associated with incorrect source decisions (or source guesses) are too weak, on average, to elicit an fMRI signal in the hippocampus (Wais, 2008; Squire et al., 2007).

In order to test for activity in the hippocampus that might be selectively associated with decisions accompanied by source recollection, we next contrasted activity associated with true source decisions against activity asso-

ciated with source guesses (again using only old decisions made with high confidence). No statistically significant regions ($p < .001$) were identified for this contrast in the hippocampus. Taken together, these results suggest that hippocampal activity signals strong memory, whether or not source recollection is involved.

With the same analysis just described, the voxel-based t tests also revealed several areas of activity in the parahippocampal gyrus. Specifically, in left parahippocampal cortex, activity associated with true source decisions was significantly greater than activity associated with forgotten items (Figure 5A), just as was the case in the hippocampus bilaterally (Figure 4A). Also as in the hippocampus, in signal data extracted from this cluster, the activity levels for source guesses and false source memories were both numerically greater than that associated with forgotten items and did not differ significantly from the level associated with true source decisions. In left perirhinal cortex, the activity associated with source guesses was significantly greater than the activity associated with forgotten items (Figure 5B), just as was the case in the right hippocampus (Figure 4B). Also as in the right hippocampus, the activity levels associated with true source decisions and false source decisions extracted from this cluster were both numerically greater than that associated with forgotten items and did not differ significantly from the level for source guesses. Lastly, a contrast between activity associated with true source decisions versus activity associated with source guesses identified no significant regions within the parahippocampal gyrus. Thus, the pattern in the parahippocampal gyrus was similar to that seen in the hippocampus.

Next, because both the dorsolateral and the ventrolateral regions of prefrontal cortex (DLPFC and VLPFC, respectively) have been associated with source memory processes in prior fMRI studies (Badre & Wagner, 2007; Ranganath & Blumenfeld, 2007), we examined the whole-brain data to determine whether activity in prefrontal cortex (PFC) associated with true source judgments, source guesses, or false source judgments was greater than activity associated with forgotten items (again using only old decisions made with high confidence). Voxel-based t tests (thresholded at $p < .001$) identified one region in left DLPFC (approximately BA 47/BA 11) and one region in left VLPFC (approximately BA 44) where the activity associated with true source judgments was significantly greater than for forgotten items. To test whether activity was selectively associated with decisions accompanied by source recollection, we next performed the contrast of true source decisions versus source guesses in the whole-brain data (again using only old decisions made with high confidence). Unlike in the MTL, this contrast identified two areas that were significantly more active when source information was recollected. Activation in left VLPFC (Figure 6A), approximately BA 45, and right DLPFC (Figure 6B), approximately BA 46, increased sig-

nificantly during responses correlated with true source memory as compared to source guesses. No clusters of activity in PFC were identified by the contrast of true source decisions versus false source decisions. Thus, even when memories were equated for confidence, a signal associated with source recollection was identified in left VLPFC and right DLPFC.

All of the preceding analyses were designed to test for activity correlated with the presence or absence of source recollection after memory strength was equated. To determine whether our findings would be similar to those reported in previous studies that did not take steps to equate for memory strength, we also conducted an analysis that was based on the notion that old–new confidence ratings of 6 denote recollection-based decisions, whereas confidence ratings of less than 6 denote familiarity-based decisions (Yonelinas, 1994). This theory has often been used to guide fMRI analyses in the past (Daselaar, Fleck, & Cabeza, 2006; Montaldi, Spencer, Roberts, & Mayes, 2006; Yonelinas, Otten, Shaw, & Rugg, 2005; Ranganath et al., 2004) even though considerable evidence suggests that weak memories are associated with lower degrees of recollection, not the absence of recollection (Slotnick & Dodson, 2005). Voxel-based t tests of the LDDMM data for the MTL (thresholded at $p < .005$) identified a region in the right hippocampus where the activity correlated with hits rated 6 was greater than the activity associated with forgotten items (Figure 7). By contrast, no regions were identified in which the activity associated with hits rated 5 differed significantly from the activity associated with forgotten items. This result is similar to the pattern of data that has been interpreted previously to indicate that the hippocampus selectively serves recollection (Daselaar et al., 2006; Montaldi et al., 2006; Yonelinas et al., 2005; Ranganath et al., 2004). An alternative interpretation suggested by all of the other data reported above is that this result indicates instead that hippocampal activity is readily detectable when confidence in memory is strong.

DISCUSSION

This study attempted to investigate neural activity associated with memory retrieval when memory strength was equated for decisions based on recollection and decisions based on familiarity. Previous research on this issue, which has suggested a functional dissociation within the MTL, relied on methods that distinguished strong recollection-based memories from weak memories that were thought to be based on familiarity (Appendix; cf. Wais, 2008). These weak memories were associated with either low confidence per se (Daselaar et al., 2006; Ranganath et al., 2004), know judgments (Eldridge, Engel, Zeineh, Bookheimer, & Knowlton, 2005), or the failure to recollect task-relevant source details (Davachi, Mitchell, & Wagner, 2003). We removed the memory

Figure 5. (A) Activation identified in left parahippocampal cortex (sagittal image) for the contrast of true source decisions versus forgotten items. (B) Activation identified in left perirhinal cortex (coronal image) for the contrast of source guesses versus forgotten items. Bar graphs show signal correlated with forgotten items and with recognition decisions that were accompanied by true, guess, or false source judgments. Also illustrated are the time courses of the hrf correlated with these results, and coronal sections locating the activation on averaged anatomical images for our participants. Error bars for the three source categories represent the *SEM* of the difference scores for each comparison (relative to forgotten items), whereas the error bar for the forgotten items represents the root mean square of the *SEM* values associated with the three individual comparisons (* denotes a significant difference relative to forgotten items, *p*-corrected < .05).
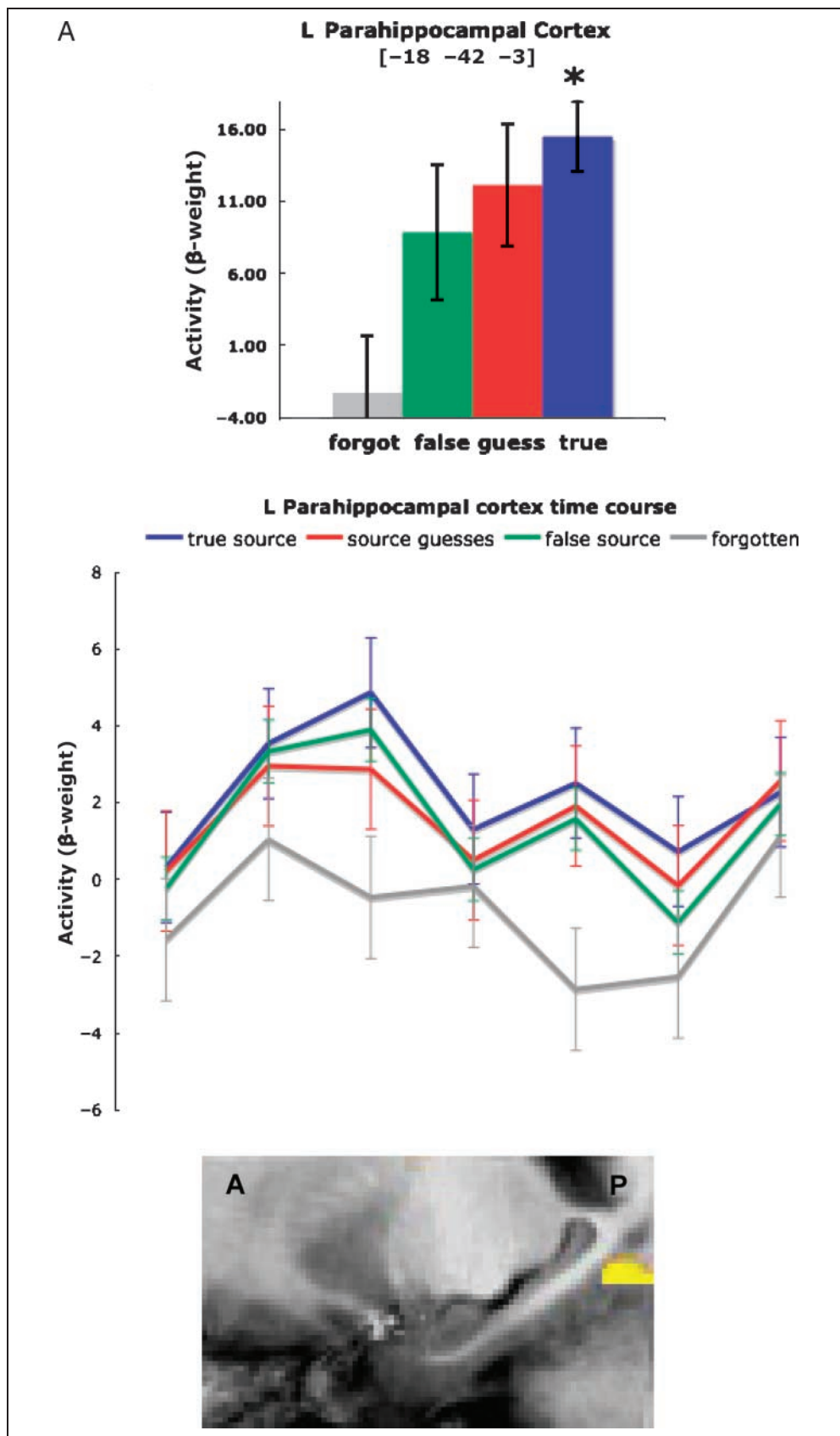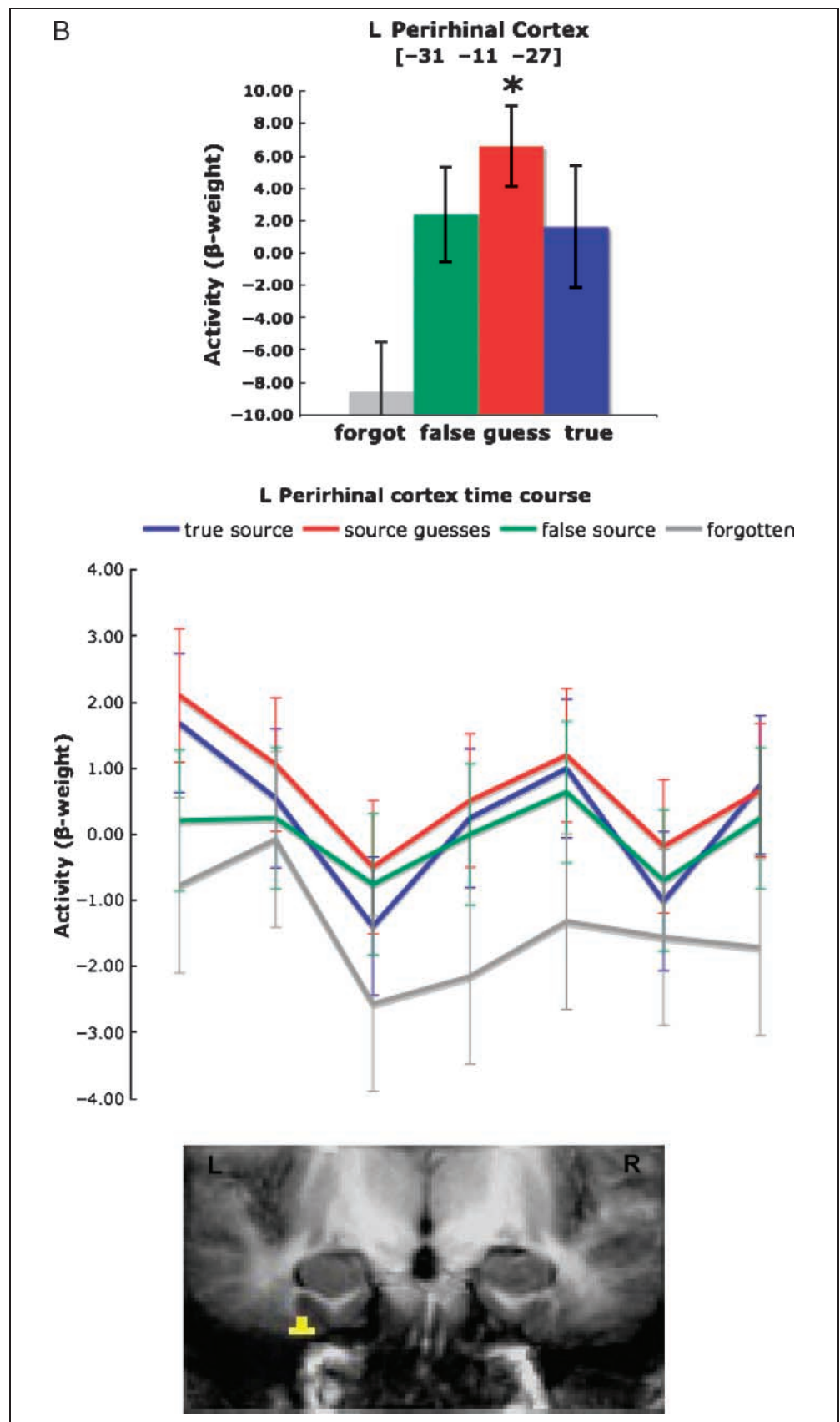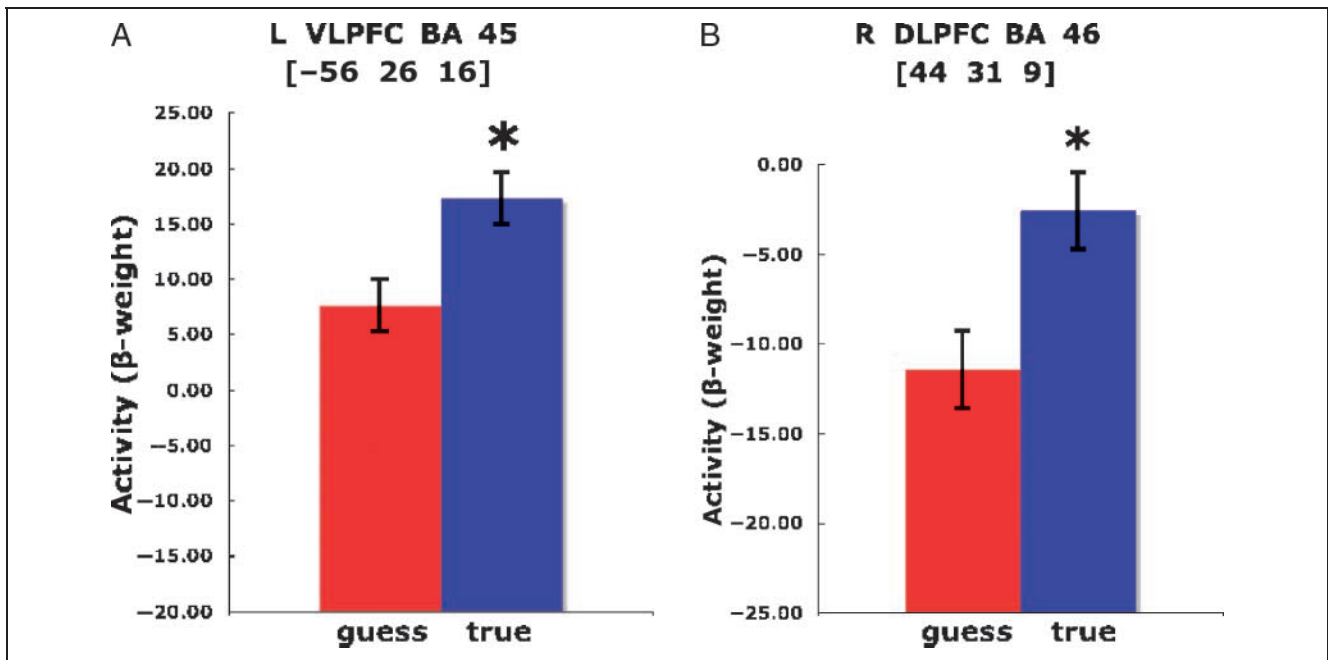
Figure 5. (*continued*)

Figure 6. Activation identified in prefrontal cortex for the contrast of high-confidence hits that were accompanied by true source judgments versus high-confidence hits that were accompanied by source guesses. (A) Left VLPFC (BA 45) and (B) right DLPFC (BA 46). Error bars represent the *SEM* of the difference scores (* denotes a significant difference between true source and source guesses, *p*-corrected < .05).

strength confound by taking old–new decisions made with relatively high confidence and then separating them into categories according to whether source memory was correct or incorrect. Under those conditions, hippocam-
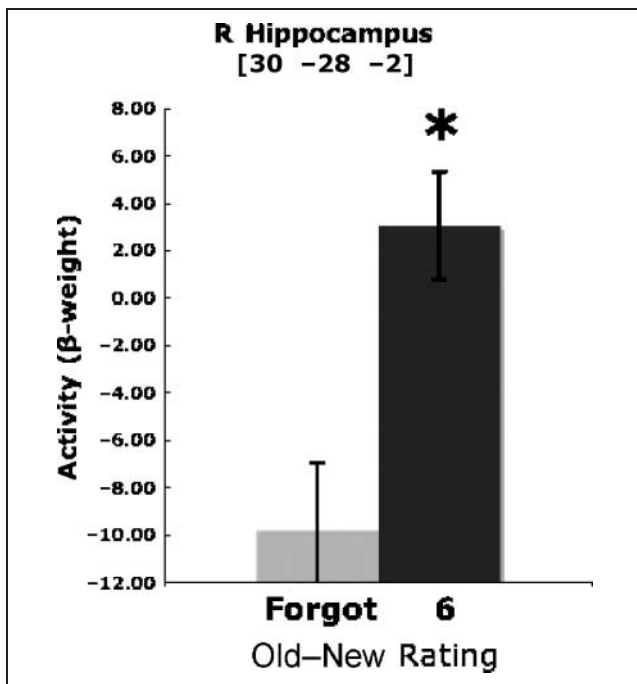

Figure 7. Activation identified for the contrast of hits rated as old–new "6" versus forgotten items. Error bars represent the *SEM* of the difference scores (* denotes a significant difference relative to forgotten items, *p*-corrected < .05).

pal activity was elevated (and equally so) for both source-correct and source-incorrect decisions relative to forgotten items. The same was true for high-confidence old–new decisions that were followed by source guesses (i.e., low-confidence source decisions associated with chance accuracy). In a conceptually similar study, Kirwan, Wixted, and Squire (2008) investigated activity at encoding and found that hippocampal activity increased with subsequent item memory strength even in the absence of source recollection. We next consider what our results mean depending on how source recollection failure is interpreted.

Interpretation 1: Source Recollection Failure Implies Familiarity-based Item Decisions

If the failure to recollect task-relevant source detail following a correct old decision is indicative of a familiarity-based decision (as has been assumed in prior studies), then our results suggest that hippocampal activity is associated with both recollection and familiarity. Interpreted in this fashion, our findings are at odds with the view that accords to the hippocampus a selective role in recollection and perirhinal cortex a selective role in familiarity (Eichenbaum et al., 2007; Brown & Aggleton, 2001). Moreover, our findings suggest that the failure of prior fMRI studies to detect increased hippocampal activity associated with familiarity-based responses may have more to do with the failure to detect a weak memory than with the absence of familiarity-related hippocampal activity (Stark & Squire, 2001).

An implication of this view is that the typical relationship between memory strength and neural activity in the hippocampus, as measured by fMRI, is nonlinear (Johnson et al., 2008; Squire et al., 2007). According to this idea, elevated hippocampal activity (relative to forgotten items) may be detectable using fMRI primarily when memory is strong (e.g., for source-correct decisions). Methods other than fMRI may be more sensitive to elevated hippocampal activity associated with weaker memories (e.g., source-incorrect decisions). In agreement with this idea, evidence from cellular recordings of single hippocampal neurons in humans indicates that some cells are more active when an item is correctly declared to be old (compared to forgotten items) even when source recollection fails, and they are more active still when source recollection succeeds (Rutishauser, Mamelak, & Schuman, 2006). This study did not equate old–new memory strength for source-correct and source-incorrect items as we did, but it nevertheless detected elevated hippocampal activity for weak, source-incorrect items.

Interpretation 2: Source Guesses Imply Familiarity-based Item Decisions

Although source-incorrect judgments are often thought to reflect familiarity-based item decisions, some incorrect source decisions are made with high confidence and might reflect false recollection of the alternative source. If so, then hippocampal activity associated with incorrect source judgments may reflect false recollection. To address this possibility, we partitioned high-confidence old–new decisions into three categories: true source memory (correct source decisions made with relatively high confidence), source guesses (correct and incorrect source decisions made with low confidence and low accuracy), and false source memory (incorrect source decisions made with relatively high confidence). The results were that hippocampal activity was higher for true source memories (strong recollection) as well as for source guesses (strong memories uncontaminated by either true or false source recollection) than for forgotten items. Furthermore, the level of increased activity in the hippocampus associated with true source memories did not differ from the level associated with source guesses. If the absence of source recollection (i.e., a source guess) is indicative of a familiarity-based decision, then, again, these results suggest that hippocampal activity is associated with both recollection and familiarity.

Whole-brain analyses showed greater activity in regions of PFC for decisions based on source recollection (true source) compared to decisions made in the absence of source recollection (source guesses) even when strength was equated. If the absence of source recollection is assumed to denote a familiarity-based decision, then these results suggest that activity in the MTL is not related to whether memory is based on recollection or familiarity, whereas a specific recollection-related signal is evident in left mid-VLPFC.

This interpretation is consistent with much prior research on the role of PFC in recollection and familiarity. For example, our finding that DLPFC (BA 46) was more active when participants made a true source decision than when they made a source guess decision is consistent with work showing that this prefrontal region is recruited when recognition is accompanied by recollection of the correct source cue, but not when recognition occurs in the absence of recollection (Ranganath, Heller, & Wilding, 2007). Additionally, our finding that the activity associated with true source memories was greater than that associated with source guess memories in left mid-VLPFC (BA 45) is consistent with the conclusion that this region is recruited during the selection of goal-relevant details (Badre & Wagner, 2007; Dobbins & Wagner, 2005). That is, according to this view, when recollection occurs, source decisions require postretrieval selection because not all of the recollected details are necessarily relevant to the source decision. VLPFC activity associated with source-correct decisions may reflect this postretrieval selection process.

Finally, if source guesses are assumed to reflect familiarity-based decisions, then our results would also be consistent with studies showing that patients with PFC lesions exhibit selective source memory deficits (Janowsky, Shimamura, & Squire, 1989), whereas item and source memory are comparably impaired in patients with hippocampal lesions (Gold et al., 2006).

Interpretation 3: Source Guesses Reflect Task-irrelevant Recollection

Another possible interpretation of our findings is that the hippocampal activity associated with the absence of task-relevant recollection reflects task-irrelevant recollection (i.e., the recollection of idiosyncratic information encoded during the learning episode). This possibility is difficult to rule out, and it attends all prior efforts to identify activity correlated with familiarity as well. For example, previous studies that interpreted decisions based on low confidence, or know judgments, or incorrect source judgments to reflect familiarity all made the assumption that no idiosyncratic recollection occurred. In our study, if task-irrelevant recollection occurred on most source guess trials, then our results would not weigh against the notion that the hippocampus selectively subserves recollection. Yet under this same assumption, our results would seem to weigh against the related notion that perirhinal cortex selectively subserves familiarity (Brown & Aggleton, 2001). This is because activity associated with source guesses, if these guesses are, in fact, contaminated by task-irrelevant recollection, was significantly elevated in perirhinal cortex compared to forgotten items, just as it was in the hippocampus.

Summary

The present research draws attention to a strength confound that exists in prior fMRI studies of recollection and familiarity, and it represents an attempt to compare the activity associated with high-strength recollection-based decisions with the activity associated with similarly high-strength familiarity-based decisions. Although more than one interpretation of our results is possible, these findings raise the possibility (unaddressed by prior research; cf., Wais, 2008) that familiarity-based activity in the hippocampus can be detected using fMRI when memory is strong. More work is needed to definitively resolve this issue, but our results should encourage the search for additional behavioral methods to isolate strong, familiarity-based decisions when seeking to identify the neuroanatomical basis of recollection and familiarity in the MTL.

APPENDIX

Model-based Interpretations

The high-threshold/dual-process model (Yonelinas, 1999) assumes that responses based on recollection and responses based on familiarity can be easily isolated from each other because they are independent processes that contribute to different recognition decisions. More specifically, the high-threshold model views the strength of familiarity as being continuous (i.e., ranging from weak confidence based on familiarity to strong confidence based on familiarity), whereas it views recollection as a categorical process that, when it occurs, only gives rise to the strongest confidence. Because recollection reliably yields responses with the highest memory confidence, this model assumes that the occurrence of recollection preempts familiarity (which is variable in strength) and that responses made with the highest confidence are typically based on recollection. If recollection does not occur, then the response is exclusively based on familiarity instead (Parks & Yonelinas, 2007). Accordingly, recognition decisions made with the highest confidence can be taken to primarily denote recollection because decisions based on highest-confidence familiarity are usually few in number, whereas recognition decisions made with confidence below that threshold can be taken to denote familiarity.

Similar considerations apply to experiments that use the Remember/Know/New procedure (Gardiner & Richardson-Klavehn, 2000; Tulving, 1985). Standard instructions for this procedure ask subjects to respond ''Remember'' based on recollection whenever it occurs and to otherwise respond ''Know'' based on a strong sense of familiarity. The Remember/Know procedure, as it has typically been applied in neuroimaging studies, assumes that decisions are based exclusively on recollection or familiarity. The most common interpretation of the Remember/Know procedure also assumes that the recollection process will preempt the familiarity process, just as the high-threshold view does.

Like other dual-process models, the aggregated-strength/dual-process model (Wixted, 2007) assumes that recollecting the contextual details associated with an item is a separate process from appreciating the familiarity of an item. However, according to this view, the strength of recollection underlying recognition responses varies from weak to strong, just as the strength of underlying familiarity varies from weak to strong (Slotnick & Dodson, 2005). In addition, it assumes that recollection and familiarity are aggregated to determine the memory strength of a particular item. That is, this model uniquely holds that both processes contribute to individual recognition decisions (Wixted, 2007; Kelley & Wixted, 2001). This aggregated strength view of recognition is compatible with standard signal-detection theory, and it implies that the strength of memory per se cannot be interpreted as a sign of underlying recollection or familiarity.

REFERENCES

Badre, D., & Wagner, A. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia, 45,* 2883–2901.

Brown, M., & Aggleton, J. (2001). Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nature Reviews Neuroscience, 2,* 51–61.

Cabeza, R., Rao, S., Wagner, A., Mayer, A., & Schacter, D. (2001). Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 4805–4810.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29,* 162–173.

Daselaar, S., Fleck, M., & Cabeza, R. (2006). Triple dissociation in the medial temporal lobes: Recollection, familiarity and novelty. *Journal of Neurophysiology, 96,* 1902–1911.

Davachi, L., Mitchell, J., & Wagner, A. (2003). Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 2157–2162.

Dobbins, I., & Wagner, A. (2005). Domain-general and domain-sensitive prefrontal mechanisms for recollecting events and detecting novelty. *Cerebral Cortex, 15,* 1768–1778.

Dunn, J. (2004). Remember–know: A matter of confidence. *Psychological Review, 111,* 524–542.

Dunn, J. C. (2008). The dimensionality of the remember–know task: A state–trace analysis. *Psychological Review, 115,* 426–446.

Eichenbaum, H., Yonelinas, A., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30,* 123–152.

Eldridge, L., Engel, S., Zeineh, M., Bookheimer, S., & Knowlton, B. (2005). A dissociation of encoding and retrieval processes in the human hippocampus. *Journal of Neuroscience, 25,* 3280–3286.

Eldridge, L., Knowlton, B., Furmanski, C., Bookheimer, S., & Engel, S. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience, 3,* 1149–1152.

Gardiner, J., & Richardson-Klavehn, A. (2000). Remembering and knowing. *The Oxford Handbook of Memory.* New York: Oxford University Press.

Gold, J., Smith, C., Bayley, P., Shrager, Y., Brewer, J., Stark, C., et al. (2006). Item memory, source memory, and the medial temporal lobe: Concordant findings from fMRI and memory-impaired patients. *Proceedings of the National Academy of Sciences, U.S.A., 103,* 9351–9356.

Janowsky, J., Shimamura, A., & Squire, L. (1989). Source memory impairment in patients with frontal lobe regions. *Neuropsychologia, 27,* 1043–1056.

Johnson, J., Muftuler, L., & Rugg, M. (2008). Multiple repetitions reveal functionally- and anatomically-distinct patterns of hippocampal activity during continuous recognition memory. *Hippocampus, 18,* 975–980.

Kelley, R., & Wixted, J. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27,* 701–722.

Kensinger, E., & Schacter, D. (2006). Amygdala activity is associated with the successful encoding of item, but not source, information for positive and negative stimuli. *Journal of Neuroscience, 26,* 2564–2570.

Kirwan, C., Jones, C., Miller, M., & Stark, C. (2007). High-resolution investigation of the medial temporal lobe. *Human Brain Mapping, 10,* 959–966.

Kirwan, C., Wixted, J., & Squire, L. (2008). Activity in the medial temporal lobe predicts memory strength, whereas activity in the prefrontal cortex predicts recollection. *Journal of Neuroscience, 28,* 10541–10548.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.

Manns, J., Hopkins, R., Reed, J., Kitchener, E., & Squire, L. (2003). Recognition memory and the human hippocampus. *Neuron, 37,* 171–180.

Mickes, L., Wixted, J., & Wais, P. (2007). A direct test of the unequal-variance signal-detection model of recognition memory. *Psychonomic Bulletin & Review, 14,* 858–865.

Miller, M., Beg, M., Ceritoglu, C., & Stark, C. (2005). Increasing the power of functional maps of the medial temporal lobe by using large deformation diffeomorphic metric mapping. *Proceedings of the National Academy of Sciences, U.S.A., 102,* 9685–9690.

Montaldi, D., Spencer, T., Roberts, N., & Mayes, A. (2006). The neural system that mediates familiarity memory. *Hippocampus, 16,* 504–520.

Parks, C., & Yonelinas, A. (2007). Moving beyond pure signal-detection models: Comment on Wixted (2007). *Psychological Review, 114,* 188–202.

Ranganath, C., & Blumenfeld, R. (2007). Prefrontal cortex and human memory: An integrative review of findings from neuropsychology and neuroimaging. *Neuroscientist, 13,* 280–291.

Ranganath, C., Heller, A., & Wilding, E. (2007). Dissociable correlates of two classes of retrieval processing in prefrontal cortex. *Neuroimage, 35,* 1663–1673.

Ranganath, C., Yonelinas, A., Cohen, M., Dy, C., Tom, S., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia, 42,* 2–13.

Rutishauser, U., Mamelak, A., & Schuman, E. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus–amygdala complex. *Neuron, 49,* 805–813.

Schacter, D., & Slotnick, S. (2004). The cognitive neuroscience of memory distortion. *Neuron, 44,* 149–160.

Slotnick, S., & Dodson, C. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition, 33,* 151–170.

Squire, L., Wixted, J., & Clark, R. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience, 8,* 872–883.

Stark, C., & Squire, L. (2001). When zero is not zero: The problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 12760–12766.

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain. 3-dimensional proportional system: An approach to cerebral imaging.* New York: Thieme Medical Publishers.

Tulving, E. (1985). Memory and consciousness. *Canadian Journal of Psychology, 26,* 1–12.

Wais, P. (2008). fMRI signals associated with memory strength in the medial temporal lobes: A meta-analysis. *Neuropsychologia, 46,* 3185–3196.

Wais, P., Mickes, L., & Wixted, J. (2008). Remember/Know judgments probe degrees of recollection. *Journal of Cognitive Neuroscience, 20,* 400–405.

Weis, S., Specht, K., Klaver, P., Tendolkar, I., Willmes, K., Ruhlmann, J., et al. (2004). Process dissociation between contextual retrieval and item recognition. *NeuroReport, 1,* 2729–2733.

Wixted, J. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152–176.

Yonelinas, A. (1994). Receiver operating characteristics in recognition memory: Evidence for a dual process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341–1354.

Yonelinas, A. (1999). The contribution of recollection and familiarity to recognition and source memory: An analysis of receiver operating characteristics and a formal model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1415–1434.

Yonelinas, A., Otten, L., Shaw, K., & Rugg, M. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience, 25,* 3002–3008.