

The Hippocampus Supports Both Recollection and Familiarity When Memories Are Strong

Christine N. Smith,¹ John T. Wixted,² and Larry R. Squire^{1,2,3,4}

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

Recognition memory is thought to consist of two component processes—recollection and familiarity. It has been suggested that the hippocampus supports recollection, while adjacent cortex supports familiarity. However, the qualitative experiences of recollection and familiarity are typically confounded with a quantitative difference in memory strength (recollection > familiarity). Thus, the question remains whether the hippocampus might in fact support familiarity-based memories whenever they are as strong as recollection-based memories. We addressed this problem in a novel way by using the Remember/Know procedure, which allowed us to explicitly match the confidence and accuracy of Remember and Know decisions. As in earlier studies, recollected items had higher accuracy and confidence than familiar items, and hippocampal activity was higher for recollected items than for familiar items. Furthermore, hippocampal activity was similar for familiar items, misses, and correct rejections. When the accuracy and confidence of recollected and familiar items were matched, the findings were dramatically different. Hippocampal activity was now similar for recollected and familiar items. Importantly, hippocampal activity was also greater for familiar items than for misses or correct rejections (as well as for recollected items vs misses or correct rejections). Our findings suggest that the hippocampus supports both recollection and familiarity when memories are strong.

Introduction

The acquisition of declarative memory depends on the integrity of the medial temporal lobe (the hippocampus, the dentate gyrus and subicular complex, together with entorhinal, perirhinal, and parahippocampal cortices). One of the most widely studied examples of declarative memory is recognition memory—the ability to judge an item as having been encountered previously. There is broad agreement that recognition memory consists of two distinct components, recollection and familiarity (Atkinson and Juola, 1974; Mandler, 1980). Recollection involves remembering specific details about the learning episode. Familiarity refers to remembering that an item was encountered previously, but without the ability to identify any information about the learning episode.

There has been considerable interest in the neuroanatomy of recollection and familiarity, particularly in the possibility that structures within the medial temporal lobe might differentially and uniquely support these functions. For example, it has been proposed that recollection depends on the hippocampus and familiarity on the adjacent perirhinal cortex (for reviews, see Brown and Aggleton, 2001; Diana et al., 2007; Eichenbaum et al.,

2007; Skinner and Fernandes, 2007). In lesion studies as well as in neuroimaging studies, a number of methods have been used to separate recollection and familiarity, including high or low confidence ratings, the presence or absence of source recollection, and Remember or Know judgments (Squire et al., 2007).

In fMRI studies, a common finding has been that hippocampal activity is higher for recollection-based decisions than for familiarity-based decisions (Eldridge et al., 2000; Cansino et al., 2002; Yonelinas et al., 2005; Daselaar et al., 2006; Montaldi et al., 2006; Otten, 2007). In addition, familiarity-based decisions often do not appear to engage the hippocampus. Thus, hippocampal activity is often no different when items are recognized based on familiarity than when items are not recognized as having appeared on an earlier list (Eldridge et al., 2000; Davachi et al., 2003; Montaldi et al., 2006; Vilberg and Rugg, 2007).

The interpretation of this rather consistent picture is complicated by the fact that the methods used to differentiate recollection from familiarity also invariably differentiate strong memories from weak memories. That is, recollection-based decisions are typically associated with higher confidence and/or higher accuracy than familiarity-based decisions (even though this need not be the case). For example, old/new judgments made with high confidence, old judgments made in association with correct source judgments, and Remember judgments are all made with higher confidence and/or higher accuracy than judgments made with low confidence, with incorrect source judgments, or Know judgments (for high vs low confidence, see Reed et al., 1997; Mickes et al., 2007; for correct vs incorrect source, see Slotnick and Dodson, 2005; Gold et al., 2006; for Remember vs Know, see Dunn, 2004, 2008; Wixted and Stretch, 2004; Rotello and Zeng, 2008).

Received July 6, 2011; accepted Aug. 25, 2011.

Author contributions: C.N.S., J.T.W., and L.R.S. designed research; C.N.S. performed research; C.N.S. analyzed data; C.N.S., J.T.W., and L.R.S. wrote the paper.

This work was supported by the Medical Research Service of the Department of Veterans Affairs, NIMH (Grants 24600 and 082892), and the Metropolitan Life Foundation. We thank Anna van der Horst, Jennifer Frascino, Annette Jensenon, Zhuang Song, Ashley Knutson, and Craig Stark for assistance.

Correspondence should be addressed to Dr. Larry R. Squire, Veterans Affairs Medical Center 116A, 3550 La Jolla Village Drive, San Diego, CA 92161. E-mail: lsquire@ucsd.edu.

DOI:10.1523/JNEUROSCI.3438-11.2011

Copyright © 2011 the authors 0270-6474/11/3115693-10\$15.00/0

In a recent attempt to address this strength confound (Cohn et al., 2009), participants were instructed to make a high-confidence Remember judgment when details about a previously presented item were recollected and to make a high-confidence Familiar judgment (equivalent to a high-confidence Know judgment) when no details about the item could be recollected. However, the behavioral accuracy scores computed from hit and false alarm rates indicate that Remember judgments and high-confidence Familiar judgments differed considerably in memory strength despite the intention to avoid this problem (Wixted et al., 2010).

We have addressed this problem in a novel way by using the Remember/Know procedure. Participants were first asked to make an old/new judgment for each studied item according to a 1–20 confidence scale and were then asked to judge each item according to whether it was Remembered, Known, or a Guess (Rotello and Zeng, 2008; Wixted and Mickes, 2010). We then assessed brain activity before and after explicitly matching the confidence and accuracy of Remember and Know decisions.

Materials and Methods

Participants. Sixteen right-handed volunteers (mean age, 27 years; age range, 19–36 years; seven female) recruited from the University community gave written informed consent before participation and were compensated monetarily.

Materials. The stimuli were 360 nouns with a mean frequency of 27 (range, 1–191) and concreteness rating >500 (mean, 573) obtained from the MRC Psycholinguistics Database (Wilson, 1988). Five 60-word lists were used for study, and one 60-word list provided foils for the retrieval test. The ratio of targets to foils (300:60) maximized the number of trials available for the primary analyses (Remember hits, Know hits, and Misses). An equal number of targets and foils would have made the scan time excessively long (>2 h). The assignment of study lists to the study and retrieval test conditions was randomized across participants. All words were presented in black font on a white background.

Procedure. Before scanning, participants saw 300 words and were told that their memory would be tested. They made a pleasant/unpleasant judgment for each word (2.5 s presentation time, 500 ms intertrial interval) by pressing one of two marked buttons on a laptop computer keyboard (Fig. 1). The study session was divided into four equal blocks of 75 trials each, with short breaks between blocks.

Following the study session (~20 min), participants took a memory test in the MRI scanner for 300 target words and 60 foil words. Participants were scanned in six separate runs (~2 min delay between runs), such that each run contained 50 target words and 10 foils. For each word, participants made an old/new recognition judgment (4 s/word) using a 20-point scale (1 = definitely new, 20 = definitely old) (Fig. 1). Participants were instructed to use the entire 20-point scale. For words declared old, participants judged whether the word was recollected, was familiar, or was a guess (2 s/word), following a modified Remember–Know–Guess procedure (Rajaram, 1996; Wixted and Mickes, 2010). The modified instructions emphasized that participants should use the Remember response only if they could actually describe specific details about the experience of studying the word. They were told that they should use the Know response if they thought the word was familiar but could not recollect any details of their encounter with the word.

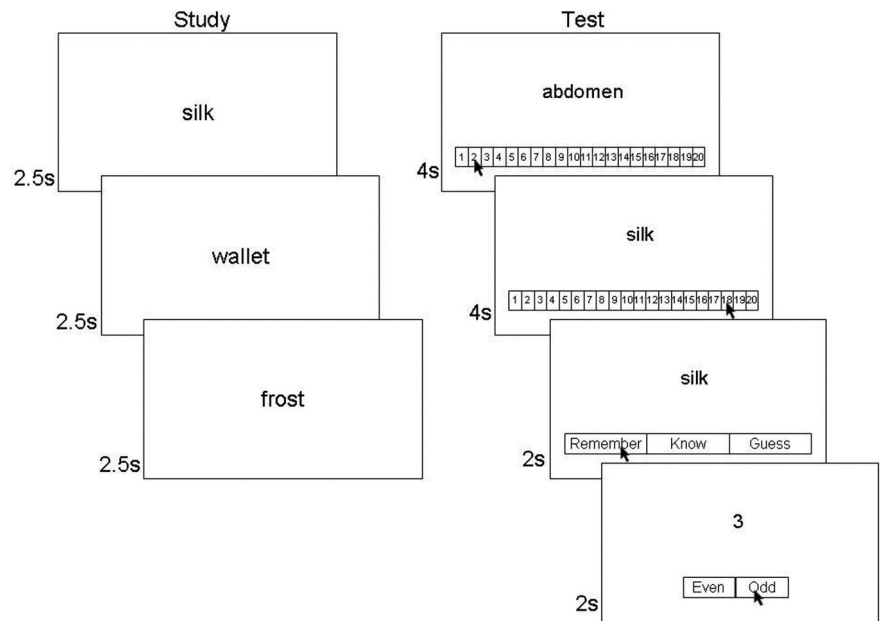


Figure 1. Task design. Participants studied 300 words ~20 min before scanning. In the scanner, participants saw 300 targets and 60 foils and performed an old/new recognition memory task using a 20-point scale (1 = definitely new, 20 = definitely old). For words designated as old (rating, 11–20), participants also made a Remember/Know/Guess judgment. The test words were separated by zero to seven baseline trials where participants made odd/even judgments. Participants moved an arrow to indicate their judgment on each trial.

Participants made their responses by moving the cursor of an MRI-compatible mouse (Current Designs) to the appropriate location on the screen (i.e., a number from 1 through 20 and the words Remember, Know, and Guess) (Fig. 1). An odd/even digit task (Stark and Squire, 2001) was intermixed with word presentation and served as a baseline against which the hemodynamic response was estimated. For the digit task, participants saw a digit from 1 to 9 (1.75 s duration followed by a 0.25 s interval) and indicated whether the digit was odd or even by moving the mouse cursor. Each scan run began with five digit trials and ended with seven digit trials. After the presentation of each word, zero to seven digit trials were given (101 total digit trials per scan run). Words were more likely to be followed by few digit trials (e.g., 0, 1, or 2 trials) than many digit trials (e.g., 5, 6, or 7 trials). The mean intertrial interval between words was 3.4 s (range, 0–14 s). Participants were given a short practice block before scanning to ensure that they understood the task and how to use the MRI-compatible mouse.

For all behavioral responses, the vertical position of the mouse cursor was fixed over the response options and the cursor could be moved only to the left or right. The starting position of the cursor was randomized across trials. In the event that participants made an erroneous response, they were instructed to indicate the error by pressing the mouse button on the subsequent trial. Three participants made erroneous responses during testing (mean for these participants, 5.3 trials; range, 1–11 trials). Trials with erroneous responses were discarded.

fMRI imaging. Imaging was performed on a 3T GE scanner at the Center for Functional MRI (University of California, San Diego). Functional images were acquired using a gradient-echo, echo-planar, T2*-weighted pulse sequence (2000 ms TR; 30 ms TE; 90° flip angle; 64 × 64 matrix size; 25 cm field of view). The duration of the experiment within each scan run varied according to the number of old and new responses participants made (i.e., trials for words designated as old lasted 6 s, whereas trials for words designated as new lasted 4 s). If participants finished the experiment before the maximum number of MR volumes had been acquired (281 MR volumes; i.e., the number of MR volumes needed if they had indicated every word was old), they viewed a screen instructing them to take a break from the task until the scanner stopped (~30 s). Any MR volumes acquired during the break period were excluded from analysis. The first five MR volumes acquired were discarded to allow for T1 equilibration. Thirty-six oblique coronal slices (slice

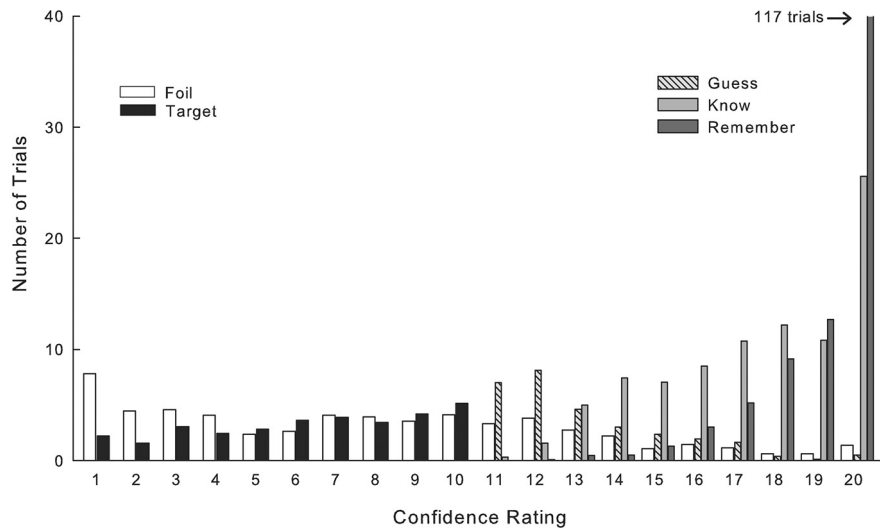


Figure 2. Distribution of targets and foils as a function of the confidence rating (1–20) assigned to each item (averaged across participants). For targets declared old (confidence rating 11–20), responses are shown separately for Remember judgments, Know judgments, and Guesses. There were no Remember judgments for targets assigned a confidence rating of 11.

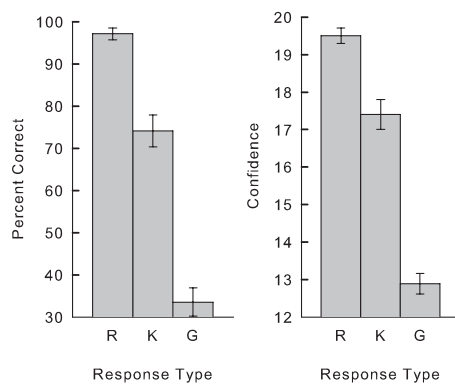


Figure 3. Accuracy (percentage correct; left) and confidence (right) for words designated as Remember (R), Know (K), or Guess (G). Note that accuracy and confidence were much higher for Remember trials than for Know trials. Error bars indicate SEM.

thickness = 4.8 mm) were acquired perpendicular to the long axis of the hippocampus and covering the whole brain. Following the six functional runs, high-resolution structural images were acquired using a T1-weighted inversion recovery spoiled gradient-recalled pulse sequence (25.6 cm field of view; 8° flip angle; 2.9 ms TE; 172 slices; 1.0 mm slice thickness; 256 × 256 matrix size).

fMRI data analysis. fMRI data were analyzed using the AFNI suite of programs (Cox, 1996). Functional data were corrected for field inhomogeneities with field mapping data collected before functional scanning, coregistered in three dimensions with the whole-brain anatomical data, slice-time corrected, and coregistered through time to reduce effects of head motion. Large motion events, defined as MR volumes in which there was >0.3° of rotation or >0.6 mm of translation in any direction, were excluded from the deconvolution analysis by censoring the excluded time points but without affecting the temporal structure of the data. We also excluded the MR volumes immediately preceding and following the motion-contaminated MR volumes.

Behavioral vectors were created that coded each retrieval trial according to the old/new status of the word and the old/new judgment to create four categories: hits [correct old responses (11–20) to a target], misses [incorrect new responses (1–10) to a target], correct rejections [correct new responses (1–10) to a foil], and false alarms [incorrect old responses (11–20) to a foil]. Two separate models were created. For the first model ($N = 16$), the vectors for the hits were divided further into Remember

hits, Know hits, and Guess hits, collapsing across memory confidence (mean number of trials for Remember hits = 148.7 ± 10.9 ; mean number of trials for Know hits = 89.2 ± 9.4 ; mean number of trials for Guess hits = 29.7 ± 6.3). A second model was created to equate accuracy and confidence for Remember hits and Know hits ($N = 16$). For each participant, trials from one or more levels of confidence were combined so that the average accuracy was the same or similar for Remember hits and for Know hits. Trials were combined from the highest confidence level(s) possible until there were sufficient trials for estimating the hemodynamic response function (mean number of trials for Remember hits = 119.9, range = 25–231; mean number of trials for Know hits = 35.7; range = 14–107). From the full set of hits, 80.5% of the Remember hits and 39.6% of the Know hits were used to create Strong Remember and Strong Know conditions. The remaining hit trials with lower confidence were modeled for Remember hits and for Know hits to create Weak Remember and Weak Know conditions. As is typically the case, there were

very few trials in the Weak Remember condition (29.1 trials on average, and eight of the 16 participants had <10 trials). In contrast, there were ample weak Know hits for fMRI analysis (mean number of trials for Weak Know hits = 53.9; range = 11–127).

The behavioral vectors, six vectors that coded for motion (three for translation and three for rotation), and three polynomial vectors that coded for linear, quadratic, and cubic drift in the MRI signal were used in deconvolution analyses of the fMRI time series data. The deconvolution method does not assume a shape of the hemodynamic response, and the fit of the data to the model was estimated for each time point independently (0–14 s after trial onset). The resultant fit coefficients (β coefficients) represent activity versus baseline in each voxel for a given time point and for each of the response categories. For comparisons that involved only words designated as old (e.g., Remember hits vs Know hits), this activity was summed over the expected hemodynamic response (2–14 s) and taken as the estimate of the response (relative to the digit task baseline).

As indicated above, the trial length was shorter for words designated as new (correct rejections and misses) than for words designated as old (hits and false alarms) (4 vs 6 s, respectively), because the Remember/Know/Guess judgment was omitted for words designated as new. Accordingly, for comparisons that involved new judgments (e.g., comparisons involving Misses or Correct Rejections), the β coefficients for the response categories were summed over the first 2–8 s of the modeled hemodynamic response (rather than 2–14 s). Note that when all analyses were limited to 2–8 s of the modeled hemodynamic response (not just analyses involving new judgments), our main findings remained the same.

Initial spatial normalization was accomplished using each participant's structural MRI scan to transform the data to the atlas of Talairach and Tournoux (1988). Statistical maps were also transformed to Talairach space, resampled to 2 mm³, and smoothed using a Gaussian filter (4 mm FWHM) that respected the anatomical boundaries of the several medial temporal lobe (MTL) regions defined for each individual participant (see below). Specifically, the smoothing was performed within each of the anatomically defined MTL regions, but smoothing was not extended beyond the edges of these regions to prevent activity from one region (e.g., parahippocampal cortex) from being blurred into another, adjacent region (e.g., hippocampus). This was accomplished by creating a separate mask for each region, smoothing the data within that mask, and then recombining the smoothed data. The Talairach-transformed data were used in the whole-brain analyses. Anatomical regions were manually segmented in 3D on the Talairach-transformed anatomical

images for the hippocampus, temporal polar, entorhinal, perirhinal, and parahippocampal cortices on each side. Temporal polar, entorhinal, and perirhinal cortices were defined according to the landmarks described by Insausti et al. (1998b). The caudal border of the perirhinal cortex was defined as 4 mm caudal to the posterior limit of the gyrus intralimbicus, as identified on coronal sections (Insausti et al., 1998b). The parahippocampal cortex was defined bilaterally as the portion of the parahippocampal gyrus caudal to the perirhinal cortex and rostral to the splenium of the corpus callosum (Insausti et al., 1998a).

We used a recent instantiation of an ROI alignment technique (ROI-ANTS) (Yassa et al., 2010; Lacy et al., 2011) to optimally align regions of the medial temporal lobe across participants (Yassa and Stark, 2009). This method uses Advanced Normalization Tools, which implements SyN (symmetric normalization), a powerful diffeomorphic registration algorithm (Klein et al., 2009). A customized anatomical space was constructed based on the Talairach-transformed structural scans from the 16 participants in the study. Each participant's grayscale scan and hand-drawn ROI segmentation of the hippocampus were used simultaneously to warp the structural scan into the customized anatomical space (Yushkevich et al., 2009).

Parameter estimate maps for each participant were entered into group-level analyses and in all cases thresholded at a voxelwise p value of <0.01 . For the MTL analyses, group statistic maps were masked to include only regions of the MTL. A cluster correction technique was used to correct for multiple comparisons in all group-level analysis, and Monte Carlo simulations (AlphaSim software) were used to determine how large a cluster of voxels was needed to be statistically meaningful ($p < 0.05$) (Forman et al., 1995; Xiong et al., 1995). Within the volume of the MTL, the minimum cluster extent was 17 contiguous voxels; for the volume of the entire brain, the minimum cluster extent was 48 voxels.

Results

Behavioral findings

Participants distributed their responses over the entire 1–20 scale (Fig. 2). High-confidence responses (ratings of 19 and 20) were primarily associated with Remember judgments, but high-confidence Know judgments were abundant as well. Guess judgments were predominantly associated with lower confidence ratings (ratings of 11 and 12).

Accuracy [percentage correct = $100 * \text{hit rate}/(\text{hit rate} + \text{false alarm rate})$] and confidence were higher for words designated as Remember ($97.1 \pm 1.4\%$ correct; 19.5 ± 0.2 confidence rating) than for words designated as Know ($74.1 \pm 3.8\%$ correct; 17.4 ± 0.4 confidence rating) ($ps < 0.001$; Fig. 3). Accuracy and confidence were lower for words designated as Guess ($33.5 \pm 3.3\%$ correct; 12.9 ± 0.3 confidence rating) than for words designated as either Remember or Know ($ps < 0.001$). The hit rates for Remember, Know, and Guess judgments were 0.50, 0.30, and 0.10, respectively, and the false alarms rates were 0.02, 0.12, and 0.17.

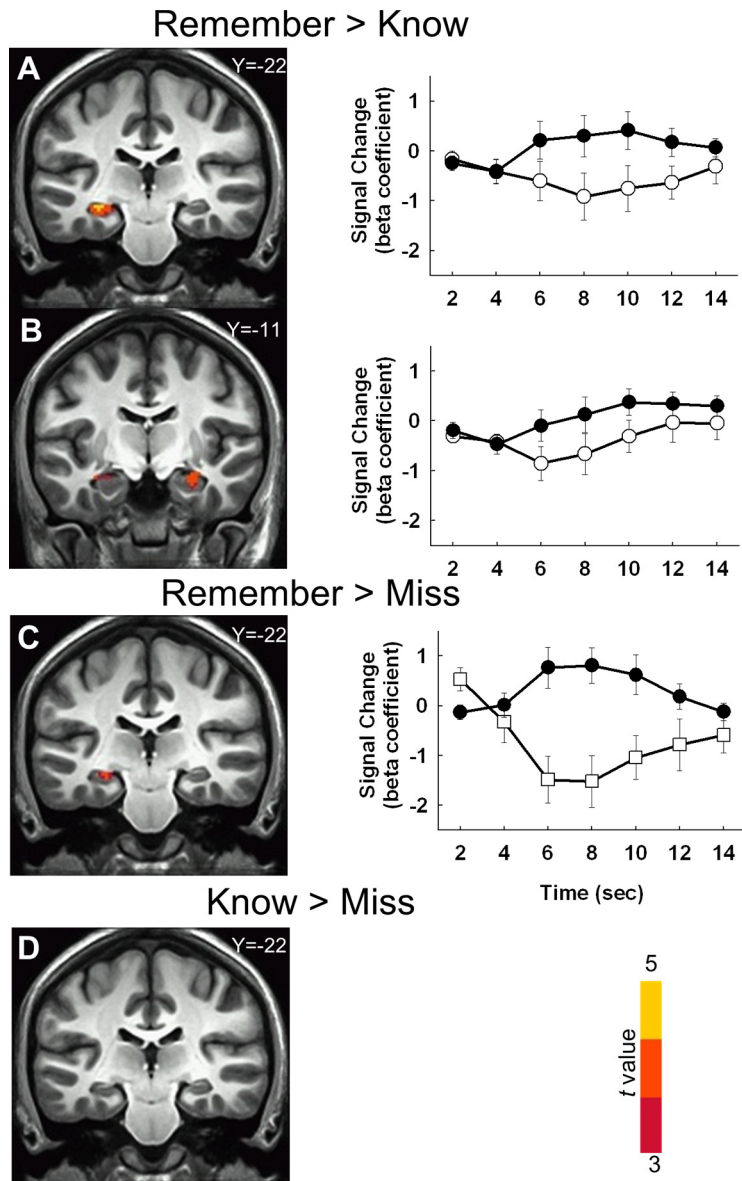


Figure 4. Coronal sections and hemodynamic response functions for clusters identified in comparisons involving Remember hits, Know hits, and Misses. **A**, A cluster in left hippocampus/parahippocampal cortex was identified where brain activity was higher for Remember hits than for Know hits (parahippocampal cortex is not visible in this section). Average hemodynamic response functions (relative to baseline) are depicted for Remember hits (black circles) and Know hits (white circles) for all voxels in the cluster. **B**, A cluster in right hippocampus was also identified from this same comparison. **C**, A cluster in left hippocampus was also identified where brain activity was higher for Remember hits than for Misses. Hemodynamic response functions are depicted for Remember hits (black circles) and Misses (white squares) for the cluster. **D**, No clusters were identified in the medial temporal lobe for the contrast of Know hits versus Misses. The anterior/posterior Talairach coordinate for each coronal section appears in the upper right of each section. Error bars indicate SEM.

fMRI findings

The first analysis of brain activity followed procedures that have been used previously with similar data (Eldridge et al., 2000; Yonelinas et al., 2005; Montaldi et al., 2006). First, we looked for clusters in the medial temporal lobe where activity for Remember hits was higher than for Know hits. Three clusters were identified, including left hippocampus/parahippocampal cortex, right hippocampus, and left temporopolar cortex (Fig. 4A,B; Table 1). Next, in separate analyses, we directly compared Remember hits and Know hits to Misses and Correct Rejections (the two latter representing responses where participants reported no experience of a memory). Activity for Remember hits was higher in left

Table 1. Medial temporal lobe analysis

	Talairach coordinates			<i>t</i> value (peak)	Cluster size (μ l)
	<i>x</i>	<i>y</i>	<i>z</i>		
Remember hits > Know hits (Fig. 4A,B)					
L. Hippocampus/parahippocampal cortex	−31	−25	−4	4.5	1248
R. Hippocampus	29	−11	−10	3.8	328
L. Temporopolar cortex	−27	13	−36	3.7	168
Remember hits > Misses (Fig. 4C)					
L. Hippocampus	−17	−39	4	4.7	648
Remember hits > Correct Rejections					
L. Hippocampus	−27	−23	−6	4.3	968
Know hits > Misses (Fig. 4D)					
No clusters found					
Know hits > Correct Rejections					
No clusters found					
Strong Remember hits > Strong Know hits (Fig. 6A)					
No clusters found					
Strong Remember hits > Misses (Fig. 6B)					
L. Hippocampus	−17	−37	4	4.3	704
Strong Remember hits > Correct Rejections					
L. Hippocampus	−27	−15	−14	4.6	928
Strong Know hits > Misses (Fig. 6C)					
L. Hippocampus/entorhinal cortex/perirhinal cortex/parahippocampal cortex	−19	−21	−16	4.5	1400
Strong Know hits > Correct Rejections					
L. Hippocampus	−21	−15	−16	4.4	232
L. Hippocampus	−31	−17	−8	4.3	232
R. Hippocampus	29	−19	−8	4.2	400
L. Parahippocampal cortex	−31	−35	−2	5.0	968
Strong Know hits > Weak Know hits					
L. Hippocampus/parahippocampal cortex	−33	−31	−4	4.9	1576
R. Hippocampus	33	−15	−8	6.8	1256
R. Hippocampus	31	−35	0	3.6	192
L. Perirhinal cortex	−31	−7	−22	4.8	168
R. Parahippocampal cortex	15	−33	−6	4.5	288

Talairach coordinates indicate the location of the voxel that had the peak *t* value. L, Left; R, right.

hippocampus than activity for Misses (Fig. 4C) or Correct Rejections (Table 1). In contrast, there were no clusters detected in the medial temporal lobe when comparing activity for Know hits to either Misses (Fig. 4D) or Correct Rejections (Table 1). The findings illustrated in Figure 4 replicate what has been reported previously in fMRI studies of Remembering and Knowing (Eldridge et al., 2000), as well as in other similar studies (Yonelinas et al., 2005; Montaldi et al., 2006).

One might conclude from our findings that activity in hippocampus is related to recollection but not to familiarity. Note, however, that words designated as Remember and words designated as Know differed not only with respect to the reported presence or absence of recollection but also with respect to the strength of the memory (Fig. 3). Specifically, words designated as Remember were recognized with both high accuracy and high confidence (accuracy >95% correct; confidence rating >19), whereas words designated as Know were recognized with lower accuracy and lower confidence (accuracy <75% correct; confidence rating <18). To compare Remember and Know responses without this difference in memory strength, we matched the accuracy and confidence ratings associated with Remember hits and Know hits (Fig. 5). These we termed Strong Remember responses ($97.7 \pm 1.4\%$ correct; 19.5 ± 0.2 confidence rating) and Strong Know responses ($94.4 \pm 2.6\%$ correct; 19.2 ± 0.3 confidence rating).

After memory strength was matched for Remember hits and Know hits, we repeated the original data analyses. These analyses

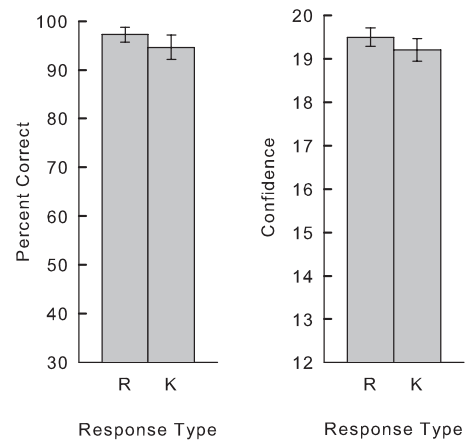


Figure 5. Accuracy (percentage correct, left) and confidence (right) for those words judged specifically to be old with high confidence that were later designated as Remember (R) or Know (K). Approximately 65% of the data from Figure 2 were used to match the strength of Remember and Know responses. Error bars indicate SEM.

yielded strikingly different results than was found before memory strength was matched. First, we found no clusters in the medial temporal lobe where activity was higher for Remember hits than for Know hits (i.e., Strong Remember vs Strong Know; Fig. 6A, Table 1). Second, clusters were now identified where activity was higher for Know hits than for Misses or Correct Rejections. Specifically, a comparison of Strong Know hits and Misses identified a cluster in left hippocampus/entorhinal/perirhinal/parahippocampal cortex (Fig. 6C). In addition, a comparison of Strong Know hits and Correct rejections identified clusters in bilateral hippocampus and left parahippocampal cortex (Table 1). Lastly, we compared Strong Remember hits to Misses and Correct Rejections. For these comparisons (Fig. 6B, Table 1), we identified the same clusters as in the original analyses (Fig. 4C, Table 1).

It is worth pointing out that perirhinal cortex activity was unique to the Strong Know versus Miss contrast. Specifically, a cluster in left perirhinal cortex exhibited more activity in association with Strong Know versus Miss contrast than in association with the Strong Remember versus Miss contrast (440μ l; $-31, -21, -24$). This finding is interesting in light of the putative role of perirhinal cortex in familiarity-based responses (Brown and Aggleton, 2001; Diana et al., 2007; Eichenbaum et al., 2007; Skinner and Fernandes, 2007).

Our findings suggest that activity in the medial temporal lobe identified earlier in the contrast between Remember hits and Know hits (Fig. 4A,B) is related to differences in memory strength and not to the presence of recollection. If so, one might expect to obtain similar findings in the medial temporal lobe in other comparisons involving conditions that differ in memory strength. Accordingly, we compared Strong Know hits to Weak Know hits. These two response categories differed considerably in memory strength (Strong Know: $95.9 \pm 2.1\%$ correct, 19.2 ± 0.3 confidence rating; Weak Know: $69.7 \pm 4.7\%$ correct, 16.2 ± 0.3 confidence rating). As expected, there was considerable overlap between the regions identified in this new contrast (Strong Know hits vs Weak Know hits; Table 1) and the regions identified in the original Remember hits versus Know hits contrast (Fig. 4A,B). Specifically, brain activity was higher for Strong Know judgments than for Weak Know judgments in the hippocampus bilaterally, the parahippocampal cortex bilaterally, and left perirhinal cortex. A similar analysis involving Remember judgments could not be performed because there were an insufficient number of Remem-

ber judgments made with low confidence (see Materials and Methods, fMRI data analysis, above).

Note that when we substantially relaxed the voxelwise threshold for the comparison of Strong Remember hits versus Strong Know hits, we were able to detect a cluster in left hippocampus that exhibited higher activity for Strong Remember judgments than for Strong Know judgments (voxelwise p value = 0.15 uncorrected). However, at the same threshold, we also detected a cluster in right hippocampus and bilateral clusters in parahippocampal gyrus that exhibited the opposite pattern: higher activity for Strong Know judgments than for Strong Remember judgments. These opposite patterns of results are what might be expected if the threshold is too low (i.e., the statistical test would be expected to yield false positives).

In summary, when memory strength was high for items designated as Remember and lower for items designated as Know, several regions in the medial temporal lobe exhibited higher activity for Remember responses than for Know responses (Fig. 4*A,B*). However, when memory strength was high for items designated as Know (and equivalent in memory strength to the items designated as Remember), no regions in the medial temporal lobe were identified in this same contrast (Fig. 6*A*). Furthermore, when memory strength was high for items designated as Remember and also high for items designated as Know, brain activity in the medial temporal lobe was higher for both Remember and Know responses than for Misses or Correct Rejections (Fig. 6*B,C*).

Although no regions in the medial temporal lobe distinguished items designated as Remember from items designated as Know after memory strength was matched, a number of neocortical regions were identified, i.e., higher activity for Strong Remember hits than for Strong Know hits (Fig. 7, Table 2). These regions were left anterior cingulate and superior frontal gyrus, right medial frontal and rectal gyri, left posterior cingulate/precuneus, left angular gyrus, and bilateral cuneus. Other neocortical regions, as well as right thalamus and bilateral cerebellum, exhibited the opposite pattern, i.e., higher activity for Strong Know hits than for Strong Remember hits (Fig. 7, Table 2).

Discussion

We assessed brain activity when recognition memory judgments were associated with recollection and when these judgments were associated with familiarity. Participants first judged each studied item according to the confidence of their old/new decision (1 = definitely new, 20 = definitely old). For words judged old (ratings of 11–20), participants then decided whether the item was Re-

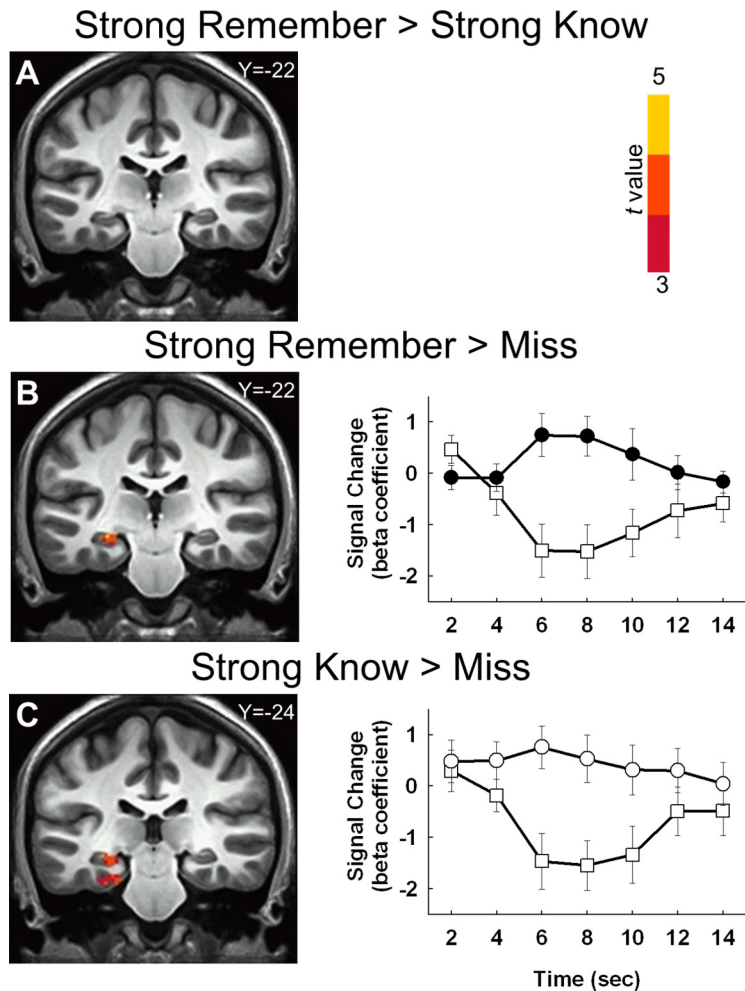


Figure 6. Coronal sections and hemodynamic response functions for clusters identified in comparisons involving Strong Remember hits, Strong Know hits, and Misses (and after matching for the strength of Remember and Know responses). *A*, No clusters were identified in the medial temporal lobe for the contrast of Strong Remember hits versus Strong Know hits. *B*, In contrast, a cluster in left hippocampus was detected where activity was higher for Strong Remember hits than for Misses. Average hemodynamic response functions (relative to baseline) are depicted for Strong Remember hits (black circles) and Misses (white squares) for all voxels in the cluster. *C*, A cluster in left hippocampus/parahippocampal cortex was also detected where activity was higher for Strong Know hits than for Misses. Hemodynamic response functions are depicted for Strong Know hits (white circles) and Misses (white squares) for the cluster. The anterior/posterior Talairach coordinate for each coronal section appears in the upper right of the panel. Error bars indicate SEM.

membered, Known, or a Guess. The behavioral results yielded the expected memory-strength difference between Remember judgments (average accuracy = 97% correct, average confidence = 19.5), thought to denote recollection, and Know judgments (average accuracy = 74% correct, average confidence = 17.5), thought to denote familiarity. When correct Remember judgments were compared with correct Know judgments, brain activity was detected in several regions of the medial temporal lobe, including hippocampus (Fig. 4*A,B*; Table 1). Moreover, hippocampal activity was higher for correct Remember judgments than for Misses (Fig. 4*C*) or Correct Rejections (Table 1), but no activity in the hippocampus or elsewhere in the medial temporal lobe was detected in similar comparisons between correct Know judgments and Misses (Fig. 4*D*) or Correct Rejections (Table 1). These findings replicate what has been reported in other, similar studies, and they appear to support the view that activity in the hippocampus is related to recollection

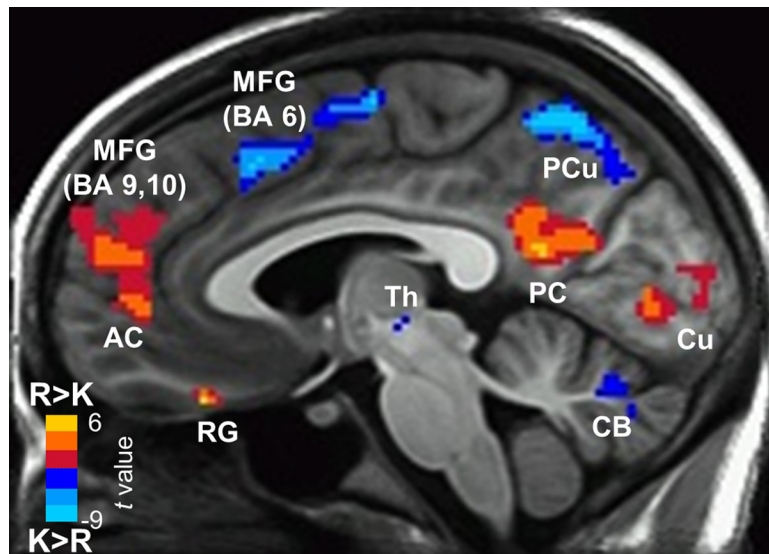


Figure 7. Sagittal section displaying clusters that exhibited higher brain activity for Strong Remember hits than for Strong Know hits ($R > K$). Clusters were not detected in the medial temporal lobe for this comparison (Fig. 6), but clusters were detected in a number of regions including bilateral medial frontal gyrus (MFG; BA 9, 10), rectal gyrus (RG), anterior cingulate gyrus (AC), posterior cingulate gyrus (PC), and cuneus (Cu). Other regions exhibited higher activity for Strong Know hits than for Strong Remember hits ($K > R$), including bilateral medial frontal gyrus (BA 6), precuneus (PCu), thalamus (Th), and cerebellum (CB). Left/right Talairach coordinate of sagittal section is -4 mm.

but not to familiarity (Eldridge et al., 2000; Yonelinas et al., 2005; Montaldi et al., 2006).

Yet, words designated as Remember and words designated as Know differed not only with respect to the reported presence or absence of recollection but also with respect to the strength of the memory itself. When the strength difference was removed by equating both the accuracy scores and the confidence ratings for items given Remember and Know judgments, the results were dramatically different. First, regions in the medial temporal lobe no longer distinguished Remember judgments from Know judgments (though regions of neocortex, especially in the frontal lobe, did make this distinction). Second, hippocampal activity was higher for both Remember judgments and Know judgments than for Misses or Correct Rejections. These findings suggest that the hippocampus is associated with elevated activity for both recollection-based and familiarity-based decisions, contrary to what has often been concluded in the past.

It is important to emphasize that our findings do not imply (and we do not claim) that hippocampal activity associated with Strong Know responses must be equivalent to hippocampal activity associated with Strong Remember responses (or that the activity associated with any particular strength of Know responses should be equivalent to the activity associated with a similar strength of Remember responses). What our findings do imply (and what we do claim) is that a familiarity signal (Strong Know $>$ Weak Know; Strong Know $>$ Miss; Strong Know $>$ Correct Rejections) is evident in the hippocampus when Know judgments are as strong as Remember judgments.

Our findings further suggest that familiarity, like recollection, must be sufficiently strong before elevated activity will be observed in the hippocampus (Song et al., 2011). Strong familiarity is associated with those Know judgments that are made with both high confidence and high accuracy. One possible concern is that high-confidence Know judgments actually reflect recollection and are not based strictly on familiarity. In fact, considerable

evidence shows that, although Know judgments are associated with much less recollection than Remember judgments, they are almost always associated with some amount of recollection (Eldridge et al., 2005; Wais et al., 2008). Yet, even high-confidence Know judgments are associated with the same small amount of recollection that is typically associated with weak Know judgments. For example, Wixted and Mickes (2010) used the same method as ours but also tested for the presence of recollection objectively using source memory questions (e.g., questions about the location and/or color of the study word when it was presented on a computer screen). For Know judgments, item recognition accuracy increased from 71% to 75% to 88% correct as item confidence increased from 15–16 to 17–18 to 19–20, respectively (on a 1–20 scale). Importantly, source accuracy for Know judgments was low and did not change as a function of item confidence (source accuracy was 58%, 57%, and 58% correct, respectively, as item confidence increased from 15–16 to 17–18 to 19–20). See In-

gram et al. (2011) for a similar result using a different Remember/ Know procedure.

Thus, high-confidence Know responses are like lower-confidence Know responses in that they contain little recollection (though some small amount of recollection can always be detected in conjunction with Know responses). Accordingly, strong Know judgments achieve their high strength because they involve a greater degree of familiarity—not a greater degree of recollection—than weak Know judgments. The point is that high-confidence Know responses are associated with no more recollection than that which occurs in the Know condition of typical Remember/ Know studies where one finds that hippocampal activity is no different for Know judgments, Misses, and Correct Rejections.

A number of fMRI studies have used a modified Remember/ Know procedure to compare recollection-based decisions with high-confidence, familiarity-based decisions (Yonelinas et al., 2005; Cohn et al., 2009). However, in these studies, as in other Remember/ Know studies, the recollection condition involved considerably stronger memory than the familiarity condition. In a similar study, Montaldi et al. (2006) used a novel (familiarity-only) procedure to try to minimize the contribution of recollection. Participants studied complex visual scenes and were instructed (and trained) to scan them superficially. After a 2 d delay, a recognition memory test was administered in the scanner. Participants were asked to rate the items for familiarity and to avoid effortful recollection (but to nevertheless report recollection if it did occur). “Old” decisions were made using a rating scale of F1, F2, F3 (low, medium, and high degrees of familiarity, respectively) and R (recollection). Hippocampal activity was selectively elevated for the few R judgments that occurred and not for F3 judgments. Montaldi and Mayes (2010) recently undertook a new analysis of accuracy scores for F3 and R in that study and reported them to be similar (88% and 89% correct, respectively). Yet, it is important to note that the two hippocampal clusters identified in that study were found only when a more

Table 2. Whole-brain analysis

	Brodmann area	Talairach coordinates			<i>t</i> value (peak)	Cluster size (μ l)
		<i>x</i>	<i>y</i>	<i>z</i>		
Strong Remember hits > Strong Know hits (Fig. 7)						
R. Medial frontal/rectal gyri	11	1	39	−14	5.4	2424*
L. Medial frontal/anterior cingulate gyri	9/10/32	−5	55	18	4.0	2424*
L. Superior frontal/medial frontal gyri	8	−17	27	42	4.4	2368
L. Posterior cingulate gyrus/precuneus	31	−9	−51	20	4.6	2416*
L. Angular gyrus	39	−55	−65	34	4.1	904
Cuneus	18/30	5	−81	6	4.0	1248*
	18	21	−95	24	4.1	736
	18	−11	−87	16	3.6	624*
Strong Know hits > Strong Remember hits (Fig. 7)						
Bil. Medial frontal/middle frontal/anterior cingulate gyri/R. superior frontal gyrus/basal ganglia	6/8/9/10/32	7	25	38	−8.6	61176*
L. Superior/middle frontal gyri	8/10	−47	31	32	−5.7	7744
Inferior frontal gyrus	44/45	51	21	12	−3.8	592
	6	−41	−7	32	−4.7	1488
L. Insular cortex/inferior frontal gyrus	13/44/45	−57	9	2	−5.6	3512
L. Superior temporal gyrus	22	−65	−9	10	−4.8	536
R. Middle temporal gyrus	19	35	−57	16	−5.3	536
L. Fusiform gyrus/cerebellum		−29	−27	−22	−5.1	424
R. Posterior cingulate gyrus	23	5	−35	26	−4.6	704
R. Inferior/superior parietal lobule	7/40	39	−63	48	−6.6	8728
Precuneus	7	5	−69	48	−6.9	9928*
	7	−27	−65	48	−4.4	1120
L. Cuneus	19	−25	−73	34	−4.3	2264
R. Middle temporal/middle occipital gyri	19	29	−73	18	−5.0	1416
R. Inferior/middle occipital gyri	19	31	−83	−8	−5.0	1040
L. Inferior temporal/middle occipital gyri	19	−49	−73	0	−4.7	984
R. Thalamus		7	−19	4	−4.9	912*
Cerebellum		−41	−67	−20	−4.6	4352
		−17	−71	−20	−5.2	3272*
		21	−35	−32	−4.8	2864
		9	−49	−20	−4.6	752

Talairach coordinates indicate the location of the voxel that had the peak *t* value. Bil, Bilateral; L, left; R, right.

*Clusters pictured in Figure 7.

lenient threshold was used than was used for all the other data analyses. Furthermore, the clusters were quite small (one contained two voxels and the other contained six voxels) and no correction for multiple comparisons was used. Additional work is merited with this particular method before concluding that strong, familiarity-based memory does not yield hippocampal activity. In any case, our findings suggest that under typical experimental conditions—the conditions used in most studies using the Remember/Know procedure—strong familiarity-based memory is associated with elevated activity in the hippocampus.

If the hippocampus did not support familiarity but only supported recollection, then patients with hippocampal lesions would be expected to commonly experience a strong sense of familiarity, i.e., a sense of having encountered an item previously even though nothing specific about a prior encounter can be recalled. This is a familiar experience for most people (for example, recognizing someone's face with certainty without being able to recollect anything about the person). Still, the experience is rather rare because, ordinarily, a strong sense of familiarity is accompanied by recollection. However, if hippocampal lesions selectively impair recollective ability, then this experience should not be rare for patients with such lesions. In a recent study designed to document the frequency of this phenomenon, i.e., how often hippocampal patients experience strong, familiarity-based recognition in the absence of recollection, the finding was that, if anything, the phenomenon occurs less often in patients than in controls (Kirwan et al., 2010). This finding from patients is consistent with the idea that the hippocampus supports familiarity (as well as recollection), and it accords with the neuroimaging

data presented here that strong, familiarity-based recognition memory is associated with elevated activity in the hippocampus.

No medial temporal lobe structures distinguished items designated as Remember from items designated as Know after memory strength was matched. These findings are consistent with the idea that the processes of recollection and familiarity are both supported by the hippocampus. The findings are also consistent with a single-process view that draws no distinction between recollection and familiarity. However, consistent with a dual-process view, a number of neocortical regions did discriminate between these processes. In particular, dorsomedial and dorsolateral prefrontal cortex exhibited more activity for Strong Remember judgments than for Strong Know judgments (Fig. 7, Table 2). These findings are consistent with two neuroimaging studies that also associated recollection with frontal cortical activity after memory strength was accounted for (Kirwan et al., 2008; Wais et al., 2010). In addition, our findings are concordant with a substantial neuropsychological literature linking frontal lobe function to source memory, free recall, and other measures of recollection (Janowsky et al., 1989; Moscovitch and Winocur, 1995; Wheeler et al., 1995).

Our results weigh against the suggestion that the hippocampus does not support familiarity and help to explain why so many findings seem to suggest otherwise. Specifically, prior studies using the Remember/Know procedure (as well as other procedures) have almost invariably compared strong memories (Remember) to weak memories (Know). It may be the case with fMRI that elevated activity in the hippocampus is unlikely to be detected when memory is weak (Song et al., 2011). The fact that a memory

strength confound can explain why earlier Remember/Know studies failed to detect familiarity-based activity in the hippocampus should not be taken to mean that memory strength is a concept that usefully informs the functional organization of medial temporal lobe structures. Nor should our findings suggest that any observed functional differences between medial temporal lobe structures will disappear once memory strength is equated. The point instead is that the role of the hippocampus in familiarity-based recognition has been obscured by a methodological confound. The fact that hippocampal activity is associated with both recollection and familiarity once memory strength is equated at a high level suggests that the functional organization of the medial temporal lobe will be best understood in terms unrelated to the distinction between recollection and familiarity (Wixted and Squire, 2011).

References

- Atkinson RC, Juola JF (1974) Search and decision processes in recognition memory. In: Contemporary developments in mathematical psychology (Krantz DH, Atkinson RC, Suppes P, eds.), pp 243–290. San Francisco: Freeman.
- Brown MW, Aggleton JP (2001) Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nat Rev Neurosci* 2:51–61.
- Cansino S, Maquet P, Dolan RJ, Rugg MD (2002) Brain activity underlying encoding and retrieval of source memory. *Cereb Cortex* 12:1048–1056.
- Cohn M, Moscovitch M, Lahat A, McAndrews MP (2009) Recollection versus strength as the primary determinant of hippocampal engagement at retrieval. *Proc Natl Acad Sci U S A* 106:22451–22455.
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Daselaar SM, Fleck MS, Cabeza R (2006) Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. *J Neurophysiol* 96:1902–1911.
- 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 U S A* 100:2157–2162.
- Diana RA, Yonelinas AP, Ranganath C (2007) Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends Cogn Sci* 11:379–386.
- Dunn JC (2004) Remember-know: a matter of confidence. *Psychol Rev* 111:524–542.
- Dunn JC (2008) The dimensionality of the remember-know task: a state-trace analysis. *Psychol Rev* 115:426–446.
- Eichenbaum H, Yonelinas AP, Ranganath C (2007) The medial temporal lobe and recognition memory. *Annu Rev Neurosci* 30:123–152.
- Eldridge LL, Knowlton BJ, Furmanski CS, Bookheimer SY, Engel SA (2000) Remembering episodes: a selective role for the hippocampus during retrieval. *Nat Neurosci* 3:1149–1152.
- Eldridge LL, Engel SA, Zeineh MM, Bookheimer SY, Knowlton BJ (2005) A dissociation of encoding and retrieval processes in the human hippocampus. *J Neurosci* 25:3280–3286.
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* 33:636–647.
- Gold JJ, Smith CN, Bayley PJ, Shrager Y, Brewer JB, Stark CE, Hopkins RO, Squire LR (2006) Item memory, source memory, and the medial temporal lobe: concordant findings from fMRI and memory-impaired patients. *Proc Natl Acad Sci U S A* 103:9351–9356.
- Ingram K, Mickes L, Wixted JT (2011) Recollection can be weak and familiarity can be strong. *J Exp Psychol Learn Mem Cogn*, In press.
- Insausti R, Insausti AM, Sobreviela M, Salinas A, Martinez-Penuela J (1998a) Human medial temporal lobe in aging: Anatomical basis of memory preservation. *Microsc Res Tech* 43:8–15.
- Insausti R, Juottonen K, Soininen H, Insausti AM, Partanen K, Vainio P, Laakso MP, Pitkänen A (1998b) MR volumetric analysis of the human entorhinal, perirhinal, and temporopolar cortices. *Am J Neuroradiol* 19:659–671.
- Janowsky JS, Shimamura AP, Squire LR (1989) Source memory, impairment in patients with frontal lobe lesions. *Neuropsychologia* 27:1043–1056.
- Kirwan CB, Wixted JT, Squire LR (2008) Activity in the medial temporal lobe predicts memory strength, whereas activity in the prefrontal cortex predicts recollection. *J Neurosci* 28:10541–10548.
- Kirwan CB, Wixted JT, Squire LR (2010) A demonstration that the hippocampus supports both recollection and familiarity. *Proc Natl Acad Sci U S A* 107:344–348.
- Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang MC, Christensen GE, Collins DL, Gee J, Hellier P, Song JH, Jenkinson M, Lepage C, Rueckert D, Thompson P, Vercauteren T, Woods RP, Mann JJ, Parsey RV (2009) Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* 46:786–802.
- Lacy JW, Yassa MA, Stark SM, Muftuler LT, Stark CE (2011) Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learn Mem* 18:15–18.
- Mandler G (1980) Recognizing: the judgment of previous occurrence. *Psychol Rev* 87:252–271.
- Mickes L, Wixted JT, Wais PE (2007) A direct test of the unequal-variance signal detection model of recognition memory. *Psychon Bull Rev* 14:858–865.
- Montaldi D, Mayes AR (2010) The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus* 20:1291–1314.
- Montaldi D, Spencer TJ, Roberts N, Mayes AR (2006) The neural system that mediates familiarity memory. *Hippocampus* 16:504–520.
- Moscovitch M, Winocur G (1995) Frontal lobes, memory, and aging. *Ann N Y Acad Sci* 769:119–150.
- Otten LJ (2007) Fragments of a larger whole: retrieval cues constrain observed neural correlates of memory encoding. *Cereb Cortex* 17:2030–2038.
- Rajaram S (1996) Perceptual effects on remembering: recollective processes in picture recognition memory. *J Exp Psychol Learn Mem Cogn* 22:365–377.
- Reed JM, Hamann SB, Stefanacci L, Squire LR (1997) When amnesic patients perform well on recognition memory tests. *Behav Neurosci* 111:1163–1170.
- Rotello CM, Zeng M (2008) Analysis of RT distributions in the remember-know paradigm. *Psychon Bull Rev* 15:825–832.
- Skinner EL, Fernandes MA (2007) Neural correlates of recollection and familiarity: a review of neuroimaging and patient data. *Neuropsychologia* 45:2163–2179.
- Slotnick SD, Dodson CS (2005) Support for a continuous (single-process) model of recognition memory and source memory. *Mem Cognit* 33:151–170.
- Song Z, Wixted JT, Smith CN, Squire LR (2011) Different nonlinear functions in hippocampus and perirhinal cortex relating functional MRI activity to memory strength. *Proc Natl Acad Sci U S A* 108:5783–5788.
- Squire LR, Wixted JT, Clark RE (2007) Recognition memory and the medial temporal lobe: a new perspective. *Nat Rev Neurosci* 8:872–883.
- Stark CE, Squire LR (2001) When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proc Natl Acad Sci U S A* 98:12760–12766.
- Talairach J, Tournoux P (1988) A co-planar stereotaxic atlas of the human brain. Stuttgart, Germany: Thieme.
- Vilberg KL, Rugg MD (2007) Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia* 45:2216–2225.
- Wais PE, Mickes L, Wixted JT (2008) Remember/know judgments probe degree of recollection. *J Cogn Neurosci* 20:400–405.
- Wais PE, Squire LR, Wixted JT (2010) In search of recollection and familiarity signals in the hippocampus. *J Cogn Neurosci* 22:109–123.
- Wheeler MA, Stuss DT, Tulving E (1995) Frontal lobe damage produces episodic memory impairment. *J Int Neuropsychol Soc* 1:525–536.
- Wilson MD (1988) The MRC psycholinguistic database: machine readable dictionary, version 2. *Behav Res Methods Instrum Comput* 20:6–11.
- Wixted JT, Mickes L (2010) A continuous dual-process model of remember/know judgments. *Psychol Rev* 117:1025–1054.

- Wixted JT, Squire LR (2011) The medial temporal lobe and the attributes of memory. *Trends Cogn Sci* 15:210–217.
- Wixted JT, Stretch V (2004) In defense of the signal detection interpretation of remember/know judgments. *Psychon Bull Rev* 11:616–641.
- Wixted JT, Mickes L, Squire LR (2010) Measuring recollection and familiarity in the medial temporal lobe. *Hippocampus* 20:1195–1205.
- Xiong J, Gao JH, Lancaster JL, Fox PT (1995) Clustered pixels analysis for functional MRI activation studies of the human brain. *Hum Brain Mapp* 3:287–301.
- Yassa MA, Stark CE (2009) A quantitative evaluation of cross-participant registration techniques for MRI studies of the medial temporal lobe. *Neuroimage* 44:319–327.
- Yassa MA, Stark SM, Bakker A, Albert MS, Gallagher M, Stark CE (2010) High-resolution functional MRI of hippocampal CA3 and dentate gyrus in patients with amnesic mild cognitive impairment. *Neuroimage* 51:1242–1252.
- Yonelinas AP, Otten LJ, Shaw KN, Rugg MD (2005) Separating the brain regions involved in recollection and familiarity in recognition memory. *J Neurosci* 25:3002–3008.
- Yushkevich PA, Avants BB, Pluta J, Das S, Minkoff D, Mechanic-Hamilton D, Glynn S, Pickup S, Liu W, Gee JC, Grossman M, Detre JA (2009) A high-resolution computational atlas of the human hippocampus from postmortem magnetic resonance imaging at 9.4 T. *Neuroimage* 44:385–398.