Published in final edited form as:

Hippocampus. 2012 June ; 22(6): 1429–1437. doi:10.1002/hipo.20982.

Hippocampal activity during recognition memory co-varies with the accuracy and confidence of source memory judgments

Sarah S. Yu^{1,2}, Jeffrey D. Johnson³, and Michael D. Rugg^{1,2}

¹Center for Vital Longevity, The University of Texas at Dallas, 1600 Viceroy Drive, Suite 800, Dallas, TX 75235

²School of Behavioral and Brain Sciences, The University of Texas at Dallas, 1600 Viceroy Drive, Suite 800, Dallas, TX 75235

³Department of Psychological Sciences, University of Missouri, 210 McAlester Hall, Columbia, MO 65211

Abstract

It has been proposed that the hippocampus selectively supports retrieval of contextual associations, but an alternative view holds that the hippocampus supports strong memories regardless of whether they contain contextual information. We employed a memory test that combined the ‘Remember/Know’ and source memory procedures, and that allowed test items to be segregated both by memory strength (recognition accuracy) and, separately, by the quality of the contextual information that could be retrieved (indexed by the accuracy/confidence of a source memory judgment). As measured by fMRI, retrieval-related hippocampal activity tracked the quality of retrieved contextual information and not memory strength. These findings are consistent with the proposal that the hippocampus supports contextual recollection rather than recognition memory more generally.

Keywords

recollection; familiarity; retrieval; fMRI

Introduction

Recognition memory is widely held to be supported by two underlying processes: *recollection* and *familiarity* (Mandler, 1980; Yonelinas, 2002; Wixted and Mickes, 2010). Recollection involves retrieval of qualitative information that supports judgments about both the prior occurrence of a test item and its encoding context. Familiarity refers to a sense of prior occurrence that can support item recognition but not contextual judgments.

Although there is broad consensus that recognition memory is supported by two memory processes, the neuroanatomical bases of recollection and familiarity are debated. According to one view (e.g. Brown and Aggleton, 2001; Eichenbaum et al., 2007) only recollection is supported by the hippocampus, while familiarity depends on extra-hippocampal regions of the medial temporal lobe (MTL), notably, the perirhinal cortex. An alternative proposal (Squire et al., 2007; Wixted et al., 2010; Wixted and Squire, 2011) is that recollection and

Corresponding author: Sarah S. Yu, Center for Vital Longevity, 1600 Viceroy Drive, Suite 800, Dallas, TX 75235, ssy100020@utdallas.edu.

Conflict of interest: The authors do not have any conflicts of interest.

familiarity both depend on the hippocampus, which imparts ‘strength’ – operationalized as the accuracy and confidence with which individual test items are recognized – to both classes of memory signal. In arguing for this view, Wixted et al. (2010) noted that in fMRI studies of retrieval-related hippocampal activity (e.g. Eldridge et al., 2000; Yonelinas et al., 2005; Montaldi et al., 2006) the recollection/familiarity distinction operationalized by ‘Remember/Know’ (Tulving, 1985) or source memory tests was invariably confounded with differences in memory strength. They argued it was this disparity in strength, and not whether an item was ‘recollected’ or ‘familiar’, that led to the differential hippocampal activity reported in these studies. Consistent with this proposal, Wais et al. (2010) reported that, relative to items wrongly endorsed as unstudied, studied recognition memory test items equated for memory strength (as operationalized by recognition confidence) elicited enhanced hippocampal responses regardless of whether an associated source memory judgment was accurate or inaccurate.

Here, we employed a memory test that permitted items to be segregated both according to whether or not they were designated as recollected on the basis of subjective report, and according to the confidence and accuracy of a source memory judgment about the items’ encoding contexts. On the basis of prior findings (reviewed by Wixted et al., 2010) we expected that memory strength would be substantially greater for items endorsed as recollected (‘Remember’ judgment) than for items endorsed as familiar (‘Know’ judgment). According to Wixted et al. (2010), this disparity in strength should be accompanied by differential hippocampal activity. By contrast, if the hippocampus selectively supports recollection of contextual details, its activity should track the amount of source-specifying information retrieved, and not memory strength *per se* (see also Diana et al., 2010; Suzuki et al., 2011). We adjudicated between these predictions by contrasting the fMRI BOLD responses elicited by test items endorsed as Know with the responses elicited by items endorsed as Remembered, segregating these latter items by the accuracy and confidence of the associated source judgment.

Materials and Methods

Subjects

Twenty-three subjects (17 female) participated in the study in return for payment. Participants ranged in age between 18 and 23 years, were right-handed and had learned English before five years of age. All subjects were recruited from the University of California, Irvine community and reported negative neurological and psychiatric histories. Prior to participating, subjects gave their informed consent in accordance with the UCI Institutional Review Board guidelines. One subject was excluded because of experimenter error during fMRI acquisition. An additional six subjects were excluded because of insufficient trials (<9) in at least one critical condition of the main analysis.

Stimulus materials

The critical experimental stimuli were drawn from 225 colored pictures of everyday objects. For each subject, 150 items were randomly selected to form the study list. The remaining 75 pictures served as new items. A test list of all 225 items was pseudo-randomized such that new pictures were intermixed with old pictures, with no more than three items of the same study status occurring consecutively. A different picture pool of 50 pictures was employed for the practice phase (30 for study, with an additional 20 added for test).

Experimental Procedure

The experiment consisted of one study-test cycle. Practice was given outside of the scanner for both the study and test tasks. Thus, subjects were aware that their memory for the

location of each study item would be tested. Following the practice session, subjects undertook the study task proper while still outside of the scanner. Study pictures were presented sequentially either on the left or right side of a display monitor, and required an indication by button-press whether the depicted object would more likely be found indoors or outdoors. Study trials consisted of the presentation of a red fixation cross (+) at the center of a gray frame for 500 ms, followed by the picture for 500 ms, followed by a black fixation cross for 2000 ms. Two buffer trials were added to the beginning and end of the study list.

Approximately thirty minutes after study, the scanned test phase began. The test list was split into three runs, each comprising 75 critical trials randomly intermixed with 25 null trials. Two buffer trials were added to the beginning of each of the three sub-lists. Test pictures were back-projected onto a screen and viewed via a mirror mounted on the scanner head coil. Each test trial consisted of the presentation of a red fixation cross at the center of a gray frame for 500 ms, which was replaced by a picture for 500 ms, followed by the cue “R/K/N?” that signaled the need to make a ‘Remember’, ‘Know’, or ‘New’ judgment via a buttonbox. The prompt was displayed for 2500 ms, and instructions emphasized the need to make the judgment as accurately as possible in a timely manner. Instructions were to endorse an item as ‘Remember’ (‘R’) if recognition was accompanied by retrieval of specific details of its study presentation, to use the ‘Know’ (‘K’) response option for items judged to be studied in the absence of the retrieval of study details, and to respond ‘New’ (‘N’) for items judged to be unstudied or for which the study status was uncertain. When an item was endorsed ‘R’ or ‘K’, a cue (“L/R?”) appeared that prompted a judgment about the item’s study location. The cue remained on the screen for 2500 ms. One of six responses could be made depending on the confidence of the location judgment: ‘Definite left’ (referred to as ‘high source confidence’ hereafter); ‘Probable left’ (i.e. ‘moderate source confidence’); ‘Possible left’ (i.e. ‘low source confidence’); ‘Definite right’; ‘Probable right’; ‘Possible right’. Subjects were instructed to respond on every trial, guessing if necessary. For each item endorsed ‘N’, a blank screen appeared for 2500 ms and the subject waited for the beginning of a new trial.

MRI Methods

A 3T Philips Achieva scanner (Philips Medical Systems, Andover, MA) equipped with an 8-channel SENSE head coil was used to acquire sagittal T1-weighted anatomical images (matrix size: 320×180 ; TR: 11 ms; TE: 4.5ms; flip angle: 18 deg; FOV: 240×157 mm; slice thickness: 0.75 (no gap); in-plane resolution: 0.75×0.75 mm) and T2*-weighted echoplanar images (EPI) (matrix size: 80×79 ; TR: 2000 ms; TE: 30ms; flip angle: 70 deg; FOV: 240×240 mm; slice thickness: 3 mm (1 mm inter-slice gap); in-plane resolution: 3×3 mm) with blood oxygenation level dependent (BOLD) contrast. Three functional runs of 240 volumes each were acquired. Each EPI volume consisted of 30 axial slices, oriented parallel to the AC-PC line, acquired sequentially in ascending order. The first five volumes of each session were discarded to allow equilibration of tissue magnetization.

fMRI Data Analysis

Functional data preprocessing and statistical analyses were performed with Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, London; <http://www.fil.ion.ucl.ac.uk/spm>) implemented in MATLAB 7.0.4 (TheMathWorks Inc, Natick, MA). The functional data were spatially realigned to the first volume of the first test run and then realigned to the mean image across the three test runs. Following spatial realignment the slices in each volume were temporally realigned to the onset of the middle slice. The data were then spatially normalized to a standard EPI template (based on the Montreal Neurological Institute (MNI) reference brain; Cocosco et al., 1997) and resampled to 3 mm isotropic voxels. Normalized images were smoothed with an 8 mm FWHM

Gaussian kernel. For each voxel, the functional timeseries was highpass-filtered to 1/128 Hz and scaled within-session to yield a grand mean of 100 across voxels and scans.

Statistical analyses were performed in two stages of a mixed effects model. Neural activity elicited by the test pictures was modeled by a delta function (impulse event) at each stimulus onset. The BOLD response to each event type was modeled by convolving these neural functions with a canonical hemodynamic response function (HRF) and its temporal and dispersion derivatives (Friston et al., 1998), yielding regressors in a General Linear Model (GLM). Parameter estimates derived from the canonical HRF were taken forward to the second level of analysis.

Limitations on trial numbers necessitated the formation of two response categories, Remember-Weak (R-Weak) and Know-All (K-All), which comprised of a mixture of trials associated with accurate and inaccurate source judgments. The source accuracy scores associated with these amalgamated categories were estimated for each subject by weighting the original score by the total number of trials in the category. To form the R-Weak category, we combined the accurate R responses that were associated with low confidence source judgments (R-Low) with all R responses that received an inaccurate source judgment. For example, if a subject made a R-Low source judgment on 10 trials with a 70% accuracy rate, and made an incorrect source judgment on 20 Remember trials, the source accuracy for the R-Weak response category would be $(0.7 \times 10)/(10+20) = 0.23$. Thus this value represents the proportion of all trials in the response category associated with an accurate source judgment. The analogous procedure was employed to estimate K-All source accuracy, wherein all accurate and inaccurate source memory trials were collapsed for items accorded K judgments.

Four events of interest were defined for the fMRI analyses: items endorsed as R that were associated, respectively, with accurate high and moderately confident source judgments (R-High and R-Mod), Remember items associated with 'Weak' source judgments (R-Weak), and all items receiving a Know judgment (K-All) (see above for details of the latter two categories). Unstudied items correctly endorsed as New (N; correct rejections) and studied items wrongly endorsed as N (misses) were also modeled. Contrasts between these items and those that went on to receive a subsequent source judgment were not included in the primary analyses because the contrasts are confounded both with the number of response cues that were presented and the number of required responses (one cue/response for items judged N, two cues/responses for items endorsed as R or K). False Alarms (R or K judgments to unstudied items) were infrequent, and modeled as events of no interest along with buffer items. Six regressors modeling movement-related variance (three rigid-body translations and three rotations determined from the realignment stage) and session-specific constant terms modeling the mean over scans were also entered into the design matrix.

To identify voxels that differentiated the four events of interest in an unbiased manner, the respective parameter estimates were subjected to one-way ANOVA using the statistical methods implemented in SPM8. The ANOVA was thresholded at $P < .001$ (2-tailed). Control of Type I error was effected by imposing a cluster-wise threshold of $P < .05$, corrected for multiple comparisons within a mask encompassing the bilateral MTL. The threshold was set at 7 contiguous voxels on the basis of a Monte Carlo simulation implemented in the Alphasim routine of the AFNI analysis package (National Institute of Mental Health, Bethesda; <http://afni.nimh.nih.gov/afni>). The resulting SPM was interrogated by inclusive masking with pre-experimentally determined, 1-tailed pair-wise contrasts (thresholded at $P < .05$), maintaining the 7 voxel extent threshold. Additionally, pair-wise t-tests were conducted on parameter estimates extracted from the hippocampal voxel demonstrating the maximum F value as identified by the aforementioned ANOVA (that is,

from a voxel selected independently of any specific contrast between the different response categories). Contrasts were also performed on parameter estimates taken from the left anterior hippocampal voxel where memory strength effects were maximal, and source accuracy effects non-significant, in the study of Wais et al. (2010). The voxel was located at $X = -30$, $y = -8$, $z = -23$ after transformation from Talairach to MNI co-ordinates (Brett et al., 2002).

Results

Behavioral results

Item memory—Table 1 shows the mean proportion of studied and unstudied test items accorded R, K and N judgments. ANOVA (factors of study status and response category; degrees of freedom corrected for non-sphericity) revealed a main effect of response category ($F_{1.6, 24.6} = 47.40$, $P < .001$) and a significant interaction between study status and response category ($F_{1.3, 18.9} = 253.25$, $P < .001$). Studied items were more likely to be given an R response than either of the other two response categories ($t_{15} = 5.96$ and $t_{15} = 12.24$ for the contrast with the K and N categories respectively, both P s $< .001$, 2-tailed). Studied items were also more likely to be endorsed as K than N ($t_{15} = 3.37$, $P < .01$, 2-tailed). Unstudied items were more likely to be correctly rejected than endorsed as R or K ($t_{15} = 24.24$ and $t_{15} = 17.39$ respectively, both P s $< .001$, 2-tailed). Subjects were more likely to misclassify new items as K than as R ($t_{15} = 4.20$, $P < .001$, 2-tailed).

Source memory—Proportions of studied items accorded R and K judgments as a function of source accuracy and confidence are shown in Table 2. A two-way ANOVA (factors of R vs. K, and response confidence) gave rise to main effects of R vs. K ($F_{1, 15} = 17.22$, $P < .001$) and confidence ($F_{3, 45} = 11.19$, $P < .001$), and to a significant interaction ($F_{2.0, 30.1} = 45.26$, $P < .001$). High confidence judgments were more frequent for items endorsed R than K ($t_{15} = 9.27$, $P < .001$), whereas the opposite was the case for low confidence judgments ($t_{15} = 4.91$, $P < .001$). As is evident from Table 2, subjects rarely made high confidence source judgments on items endorsed K; indeed, six subjects never did so.

Item and Source Strength—Because of low trial numbers, for the purposes of the fMRI analyses R items attracting low confidence and incorrect source judgments were collapsed to form an ‘R-Weak’ category. For the same reason, we created a ‘K-All’ category that was comprised of K trials associated with both accurate and inaccurate source judgments (see Materials and Methods).

Figure 1a depicts the accuracy of recognition judgments [$pHit/(pHit + pFA)$] for the different response categories. Figure 1a indicates that recognition accuracy (‘memory strength’) was markedly higher for R-Weak items than it was for K-All items (accuracy scores of .92 and .77 respectively), and then grew more modestly as a function of source confidence, approaching ceiling for the R-High category (accuracy score of .99). ANOVA contrasting the recognition accuracy scores for the R-High, R-Mod, R-Weak, and K-All categories revealed a significant main effect ($F_{1.5, 21.8} = 19.46$, $P < .001$). Follow-up t-tests revealed higher accuracy scores for R-High than R-Mod ($t_{15} = 3.10$, $P < .01$, 1-tailed), for R-Mod than R-Weak ($t_{15} = 2.30$, $P < .025$, 1-tailed), and for R-Weak than K-All ($t_{15} = 3.91$, $P < .001$, 1-tailed).

Figure 1b depicts the proportions of trials in each response category that were associated with an accurate source memory judgment. Consistent with previous findings (Hicks et al., 2002; Wais et al., 2008), source accuracy was reliably above chance for K items ($t_{15} = 6.22$, $P < .001$, 2-tailed). ANOVA contrasting the scores for the R-High, R-Mod and R-Low categories revealed a significant main effect ($F_{1.6, 24.5} = 30.67$, $P < .001$). Follow-up t-tests

revealed greater source accuracy for R-High than R-Mod ($t_{15} = 9.02$, $P < .001$, 1-tailed) and for R-Mod than R-Low ($t_{15} = 2.11$, $P < .05$, 1-tailed). Thus, as previously reported (e.g. Slotnick and Dodson, 2005), source accuracy for R items tracked response confidence. Also shown in Figure 1b are the proportions of R-Weak and K-All trials on which the source judgment was accurate. Despite the markedly greater item memory accuracy associated with the former response category, source accuracy was higher for K-All trials ($t_{15} = 4.86$, $P < .001$, 2-tailed).

fMRI results

We first used a one-way ANOVA to identify voxels where activity differed between response categories. The outcome of the ANOVA – which is unbiased with respect to any particular pair-wise contrast between the different response categories – defined the region of interest within which follow-up pair-wise contrasts were conducted using inclusive masking. The contrasts were directed towards identifying 1) voxels where responses to R-Weak items differed from those to K-All items and 2) voxels where activity elicited by R items varied as a function of source accuracy/confidence, operationalized by the contrast between R-High and R-Weak items.

The initial ANOVA identified a large cluster within the left MTL (Figure 2). The cluster contained peak voxels located in parahippocampal cortex ($-33, -40, -17$), entorhinal cortex ($-21, -25, -26$), and anterior hippocampus ($-30, -16, -20$). The pair-wise contrast between activity elicited by R-Weak and K-All items revealed a cluster of 18 voxels in the vicinity of the entorhinal cortex, and a smaller cluster of 8 voxels on the edge of parahippocampal cortex (Figure 3a). [A reviewer requested that we also report the contrast for K-All > R-Weak, on the grounds that the K-All category contains the larger proportion of source correct trials. The contrast failed to identify any significant effects within the MTL.] In striking contrast, responses elicited by R-High items were enhanced relative to responses elicited by R-Weak items in virtually the entire MTL cluster identified by the ANOVA (Figure 3b). As is illustrated in Figure 3c, differential hippocampal activity was also evident for the contrasts involving the R-Mod category (i.e. R-High > R-Mod and R-Mod > R-Weak). Thus, these analyses indicate that hippocampal activity elicited by R items tracked the confidence/accuracy of the associated source memory judgment, but that the activity elicited by R items for which source memory was weak or absent did not differ from that elicited by items endorsed K.

The analyses of the parameter estimates of the BOLD responses elicited at the peak hippocampal voxel identified by the ANOVA yielded a similar pattern of findings (Figure 4a). R-High items elicited greater activity than R-Weak items ($t_{15} = 4.55$, $P < .001$, 1-tailed), whereas the difference between R-Weak and K-All items was non-significant ($P > .4$). Additionally, R-Mod items elicited a response that was intermediate in magnitude between that elicited by R-High ($t_{15} = 1.74$, $P = .051$, 1-tailed) and R-Weak items ($t_{15} = 2.25$, $P < .025$, 1-tailed). We conducted the same analysis on the data from the left anterior hippocampal region identified as demonstrating sensitivity to memory strength but not source accuracy in the study of Wais et al. (2010; see Materials and Methods). The contrast between R-Weak and K-All items was non-significant ($P > .19$, 1-tailed), whereas responses elicited by R-High and R-Weak items differed significantly ($t_{15} = 3.22$, $P < .005$, 1-tailed).

We conducted analogous contrasts for the parameter estimates derived from the peak voxels in left entorhinal and parahippocampal cortex. In both regions the contrasts between the R-High and R-Weak response categories revealed a significant effect ($t_{15} = 3.14$, $P < .005$, and $t_{15} = 2.82$, $P < .01$, both 1-tailed, for the entorhinal and parahippocampal voxels, respectively). Significant effects were also evident in both regions for the contrast between R-High vs. R-Mod categories ($t_{15} = 2.28$, $P < .025$, and $t_{15} = 2.29$, $P < .025$, both 1-tailed).

The R-Mod vs. R-Weak contrast was not reliable in either region, whereas the R-Weak vs. K-All contrast revealed a significant entorhinal effect ($t_{15} = 2.00$, $P < .05$, 1-tailed), consistent with the finding illustrated in Figure 3a.

As reported above, recognition accuracy ('memory strength') varied across the different categories of R responses (.99, .97 and .92 for R-High, R-Mod and R-Weak categories, respectively). These differences raise the possibility that our finding of a relationship between hippocampal activity and source accuracy is due to the confounding effects of memory strength, perhaps reflecting a non-linear relationship between strength and hippocampal BOLD responses that leads to detectable modulation of hippocampal activity only when memory is especially strong (Squire et al., 2007). We addressed this possibility by contrasting the activity elicited at the left hippocampal voxel identified by the initial ANOVA in sub-groups of subjects in whom recognition accuracy (strength) was equated. Although we were unable to equate recognition accuracy between the R-High and the R-Weak response categories in a sufficient number of subjects, it was possible to form a sub-group of 13 subjects in whom accuracy was equated between the R-Weak and R-Mod categories (.96 and .97 respectively, $P > .1$), and another sub-group of 11 subjects in whom accuracy was equated between the R-Mod and R-High categories (.99 in both cases). In both sub-groups, pair-wise contrasts of the parameter estimates indicated that hippocampal activity remained sensitive to source accuracy (R-Mod vs. R-Weak: $t_{12} = 2.32$, $P < .025$; R-High vs. R-Mod: $t_{10} = 2.44$, $P < .025$; both 1-tailed, Figure 4b). Thus, when differences in recognition accuracy were eliminated, hippocampal activity continued to track the accuracy and confidence of source memory judgments. The same was true for the effects reported above with respect to the entorhinal and parahippocampal cortex. In both regions, the R-High vs. R-Mod contrast remained significant when tested in the sub-set of 11 subjects in whom recognition accuracy was matched between these response categories ($t_{10} = 2.51$, $P < .025$, $t_{10} = 2.85$, $P < .01$, both 1-tailed, for entorhinal and parahippocampal regions respectively).

For the reasons discussed in Materials and Methods, we did not include correctly rejected or missed items in the primary analyses described above. Misses were however employed by Wais et al. (2010) as a baseline against which to assess the impact of memory strength on hippocampal activity. For the sake of comparison, therefore, we contrasted the activity elicited by misses with that elicited by the different response categories analyzed above. We were unable to identify any voxels within the MTL region of interest defined by the initial ANOVA where activity elicited by these response categories exceeded the activity associated with misses. Indeed, pair-wise contrasts performed on the parameter estimates extracted from the peak hippocampal voxel revealed that the activity elicited by misses significantly exceeded that for both K-All and R-Weak items ($t_{15} = 2.77$, $P < .025$ and $t_{15} = 2.14$, $P < .05$, respectively, both 2-tailed). Similar findings emerged when these contrasts were performed on parameter estimates derived from the hippocampal region where Wais et al. (2010) reported a maximal effect of memory strength (see above). In this case, activity elicited by misses exceeded that for K-All items ($t_{15} = 2.15$, $P < .05$, 2-tailed), whereas all other contrasts were non-significant.

In a final analysis we assessed the possibility that the failure to identify differential hippocampal activity for the contrast between the R-Weak and K-All categories was a consequence of the inclusion within the K-All category of items that attracted high or moderate confidence source judgments. We formed a K-Weak response category that, like its R-Weak counterpart, contained only trials associated with low confidence or inaccurate source judgments (due to low trial numbers in one subject, this was possible for only 15 of the 16 original subjects). The R-Weak > K-Weak contrast gave rise to results essentially identical to those reported above for the R-Weak > K-All contrast.

Discussion

Behavioral data

Memory strength – operationalized by item memory accuracy – was markedly lower for items endorsed K than it was for any class of item given an R response, including items belonging to the R-Weak category. As is discussed in more detail below, recognition accuracies for the R-Weak and K-All categories are closely comparable with those argued by Wixted et al. (2010) to be sufficient to account for prior reports of recollection-selective hippocampal activity in studies using the Remember/Know procedure (e.g. Yonelinas et al., 2005; Montaldi et al., 2006; Cohn et al., 2009).

Source accuracy tracked source confidence, as has been reported previously (Slotnick and Dobson, 2005). In addition, although they were rarely of high confidence, the accuracy of source judgments was well above chance for items receiving a K judgment, also consistent with prior reports (e.g. Hicks et al., 2002; Wais et al., 2008). This latter finding has been interpreted as evidence that the recollection signal is continuous rather than discontinuous (i.e. thresholded) in character (Wais et al., 2008; Wixted and Mickes, 2010). Alternatively, the finding might indicate that source memory can be supported by familiarity as well as by recollection (e.g. Parks and Yonelinas, 2007). The present findings do not help adjudicate between these alternatives.

fMRI data

Differences in the activity elicited by correctly recognized test items were identified in a large left MTL cluster that extended from parahippocampal cortex through the entorhinal region into the anterior hippocampus (Figure 2). Follow-up analyses with either a masking or region-of-interest approach yielded the same findings: no difference in hippocampal activity could be detected for the contrast between R-Weak and K-All items, whereas robust differences in the magnitudes of the hippocampal responses were detected for the contrast between R-High and R-Weak items, the responses to R-Mod items occupying an intermediate position. Although the recognition and source accuracies of R items were positively associated in the full subject sample, hippocampal activity remained sensitive to source accuracy in sub-groups of subjects for whom recognition accuracy was matched between adjacent R response categories.

The present findings speak to the question of whether retrieval-related hippocampal activity reflects the strength of an undifferentiated memory signal (as indexed by recognition accuracy), or whether the activity reflects retrieval of qualitative information that selectively supports recollection, as has frequently been proposed (e.g. Eldridge et al., 2000; Yonelinas et al., 2005; Montaldi et al., 2006; Diana et al., 2007, 2010; Cohn et al., 2009). In arguing for the former view, Wixted et al. (2010) noted that in prior studies reporting recollection-related hippocampal activity (e.g. Eldridge, et al., 2000; Yonelinas et al., 2005; Montaldi et al., 2006; Cohn et al., 2009) the distinction between recollection and familiarity was invariably confounded with differences in memory strength (see Introduction). They further noted that in studies where subjects categorized unrecollected test items according to their familiarity strength (as indexed by confidence ratings; Yonelinas et al., 2005; Montaldi et al., 2006; Cohn et al., 2009), recognition accuracy for the most familiar items typically approximated .80, in contrast to an accuracy of .90 or more for items endorsed as recollected. Recognition accuracies in the present experiment for K-All and R-Weak items were .77 and .92 respectively, and yet these two classes of test item elicited statistically indistinguishable hippocampal responses. Thus, although caution is always required when accepting a null effect, it appears that a difference in memory strength of this magnitude is not necessarily associated with differential hippocampal activity. By contrast, among items

attracting R responses, hippocampal activity tracked the accuracy/confidence of the subsequent source memory judgment, and did so independently of recognition accuracy. Together, these findings are consistent with the proposal that retrieval-related hippocampal activity co-varies not with generic memory strength, but with the ‘strength’ or fidelity of a memory signal that carries qualitative information about the study episode (see also Montaldi and Mayes, 2010; Diana et al., 2010; Suzuki et al., 2011). It is worth noting that this finding does not necessarily invalidate ‘threshold’ models of recollection (e.g. Yonelinas, 2001), in which it is assumed that both unstudied and a variable proportion of studied items may fail to elicit a recollection signal. The findings do indicate however that once above threshold, the signal is not ‘all or none’ but varies in ‘strength’ as indexed by the amount of information retrieved (c.f. Vilberg and Rugg, 2007).

Like Wais et al. (2010), we also contrasted the activity elicited by correctly recognized items with the activity elicited by studied items incorrectly judged new (Misses), that is, items with very low memory strength. Unlike those authors, we were unable to identify any hippocampal clusters, or clusters in the MTL more generally, where recognized items elicited a response greater than the response to Misses. Rather, those differences that were detected were in the opposite direction. These findings are reminiscent of several prior reports of enhanced hippocampal responses for unfamiliar relative to familiar test items (e.g. Dudukovic and Wagner, 2007; Rombouts et al., 2001; Duzel et al., 2003; Rugg et al., 2003; Johnson et al., 2008; Suzuki et al., 2011). Such ‘new>old’ or ‘novelty’ effects are often taken as evidence of the proclivity of hippocampally-mediated encoding processes to be engaged by relatively novel or unfamiliar stimulus events (Okado and Stark, 2005). It is unclear why the present findings diverge so markedly from those of Wais et al. (2010). However, as we have already noted, the results of the contrasts involving items endorsed as New should be interpreted cautiously because of the potentially confounding effects of the different response requirements associated with these items as opposed to items endorsed R or K. This caveat is equally relevant to the findings of Wais et al. (2010), who also employed a memory test in which the number of responses differed for items judged old and new.

The present findings diverge from those of Wais et al. (2010) not only with respect to the relative magnitudes of the hippocampal activity elicited by missed items, but also with regard to the sensitivity of the hippocampus to source accuracy. Comparison of the levels of source accuracy between the two studies suggests a possible explanation for these divergent findings. In Wais et al. (2010) mean source accuracy for test items recognized with high confidence was .67, and accuracy for items recognized at the adjacent confidence level was .63. These two item classes were combined prior to fMRI analysis, meaning that source accuracy would have fallen somewhere between these two values. Source accuracy levels in the present study were considerably higher (.88 for R-High items). It is possible therefore that the failure of Wais et al. (2010) to detect effects of source accuracy in the hippocampus reflected the relatively low levels of source recollection in that study. For example, at least with the whole brain imaging methods employed in both that study and the present one, it may be that hippocampal responses associated with low levels of recollection are not detectable, perhaps due to a non-linear relationship between the strength of the recollection signal and hippocampal BOLD activity (cf. Squire et al., 2007). This proposal can also account for the absence of differential hippocampal activity between R-Weak and K-All items in the present study. By this account, the low levels of recollection elicited by R-Weak items (presumably including recollection solely of episodic information undiagnostic of source on trials where the source judgment was inaccurate, i.e. ‘non-criterial recollection’; Yonelinas and Jacoby, 1996) was sufficient to boost recognition accuracy relative to K items (when recollection was even lower or absent), but was insufficient to give rise to a larger hippocampal response than the one elicited by K items.

The present findings indicate that retrieval-related activity in the hippocampus is modulated by the amount and quality of the contextual information retrieved in association with a studied test item. As discussed above, these findings are consistent with the long-standing idea that the hippocampus plays a selective role in recollection, but it is also worth considering them in light of the recent proposal that the hippocampus supports memories composed of multiple ‘attributes’ rather than just a single attribute, regardless of whether the memories support recollection or familiarity (Wixted and Squire, 2011). On the assumption that accurate source judgments depend upon the retrieval of multi-attribute memories (the item along with, in the present case, its spatial context) the present findings are consistent with this proposal. Its full evaluation, however, will require the development of experimental procedures that permit the neural correlates of the retrieval of equally ‘strong’ multi- and single-attribute memories to be directly contrasted.

As is evident from Figure 3b, the contrast between R-High and R-Weak items identified voxels not only within the hippocampus, but across much of the left MTL. Consistent with prior reports (e.g. Kirwan and Stark, 2004; Diana et al., 2010), these results indicate that the hippocampus is not the only MTL region where activity is enhanced when recollection is successful. Among the MTL regions identified by our initial one-way ANOVA, activity in the entorhinal cortex was enhanced for R-Weak compared with K-All items. It is unclear whether this effect should be interpreted as evidence for the sensitivity of this region to differences in memory strength (given the marked difference in strength between the items in the two response categories) or recollection [on the assumption that R-Weak items were associated with a somewhat stronger recollection signal than K-All items (see above)]. The finding that entorhinal activity was enhanced for R-High relative to R-Mod items even when overall memory strength for these two classes of items was equated (see Results) suggests that the latter interpretation may be the more likely.

In conclusion, the present findings suggest that retrieval-related hippocampal activity is sensitive to the amount of episodic (contextual) information that is retrieved. The findings therefore support proposals that the hippocampus selectively supports recollection (e.g. Aggleton and Brown, 1999; Eichenbaum, et al., 2007), or, perhaps, multi-attribute memories more generally (Wixted and Squire, 2011), rather than an undifferentiated memory signal that varies solely in its strength. Additionally, the findings suggest that, despite the confound that is typically present between R and K judgments and recognition accuracy (Wixted et al., 2010), the enhanced hippocampal activity often reported to be associated with R judgments is attributable to the greater levels of recollection with which they are associated. This is a consequence of the fact that a markedly higher proportion of R than K judgments are associated with retrieval of qualitative information about the study episode (see Figure 1 and Wixted and Mickes, 2010).

Acknowledgments

Grant sponsor: NIH Grant number: R01-MH072966

References

- Aggleton JP, Brown MW. Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behav Brain Sci.* 1999; 22:425–444. [PubMed: 11301518]
- Brett M, Johnsrude IS, Owen AM. The problem of functional localization in the human brain. *Nat Rev Neurosci.* 2002; 3:243–249. [PubMed: 11994756]
- Brown MB, Mickes L. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? *Nat Rev Neuro.* 2010; 2:51–61.

- Cocosco CA, Kollokian V, Kwan RS, Evans A. Brainweb: online interface to a 3D MRI simulated brain database. *NeuroImage*. 1997; 5:S425.
- Cohn M, Moscovitch M, Lahat A, McAndrews MP. Recollection versus strength as a primary determinant of hippocampal engagement at retrieval. *Proc Natl Acad Sci USA*. 2009; 106:22451–22455. [PubMed: 20007783]
- Diana RA, Yonelinas AP, Ranganath C. Medial temporal lobe activity during source retrieval reflects information type, not memory strength. *J Cogn Neurosci*. 2010; 22:1808–1818. [PubMed: 19702458]
- Dudukovic NM, Wagner AD. Goal-dependent modulation of declarative memory: neural correlates of temporal recency decisions and novelty detection. *Neuropsychologia*. 2007; 45:2608–2620. [PubMed: 17499318]
- Duzel E, Habib R, Rotte M, Guderian S, Tulving E, Heinze H-J. Hippocampal and parahippocampal activity during associative recognition memory for spatial and nonspatial stimulus configurations. *J Neurosci*. 2003; 23:9439–9444. [PubMed: 14561873]
- Eichenbaum H, Yonelinas AP, Ranganath C. The medial temporal lobe and recognition memory. *Ann Rev Neurosci*. 2007; 30:123–152. [PubMed: 17417939]
- Eldridge LL, Knowlton BJ, Furmanski CS, Bookheimer SY, Engel SA. Remembering episodes: a selective role for the hippocampus during retrieval. *Nat Neurosci*. 2000; 3:1149–1152. [PubMed: 11036273]
- Friston KJ, Fletcher PC, Josephs O, Holmes A, Rugg MD, Turner R. Event-related fMRI: Characterizing differential responses. *NeuroImage*. 1998; 7:30–40. [PubMed: 9500830]
- Hicks JL, Marsh RL, Ritschel L. The role of recollection and partial information in source monitoring. *J Exp Psychol Learn Mem Cogn*. 2002; 28:503–508. [PubMed: 12018502]
- Johnson JD, Muftuler LT, Rugg MD. Multiple repetitions reveal functionally and anatomically distinct patterns of hippocampal activity during continuous recognition memory. *Hippocampus*. 2008; 18:975–980. [PubMed: 18548578]
- Kirwan CB, Stark CE. Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. *Hippocampus*. 2004; 14:919–930. [PubMed: 15382260]
- Mandler G. Recognizing: the judgment of previous occurrence. *Psych Rev*. 1980; 87:252–271.
- Montaldi D, Mayes AR. The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus*. 2010; 20:1291–1314. [PubMed: 20928828]
- Montaldi D, Spencer TJ, Roberts N, Mayes AR. The neural system that mediates familiarity memory. *Hippocampus*. 2006; 16:504–520. [PubMed: 16634088]
- Okado Y, Stark CE. Neural activity during encoding predicts false memories created by misinformation. *Learn Mem*. 2005; 12:3–11. [PubMed: 15687227]
- Parks CM, Yonelinas AP. Moving beyond pure signal-detection models: comment on Wixted (2007). *Psych Rev*. 2007; 114:199–202.
- Rombouts SA, Barkhof F, Witter MP, Machielsen WC, Scheltens P. Anterior medial temporal lobe activation during attempted retrieval of encoded visuospatial scenes: an event-related fMRI study. *NeuroImage*. 2001; 14:67–76. [PubMed: 11525338]
- Rugg MD, Henson RNA, Robb WGK. Neural correlates of retrieval processing in the prefrontal cortex during recognition and exclusion tasks. *Neuropsychologia*. 2003; 41:40–52. [PubMed: 12427564]
- Squire LR, Wixted JT, Clark RE. Recognition memory and the medial temporal lobe: a new perspective. *Nat Rev Neurosci*. 2007; 8:872–883. [PubMed: 17948032]
- Slotnick SD, Dodson CS. Support for a continuous single-process model of recognition memory and source memory. *Mem Cognit*. 2005; 33:151–170.
- Suzuki M, Johnson JD, Rugg MD. Recollection-related hippocampal activity during continuous recognition: a high-resolution fMRI study. *Hippocampus*. 2011; 21:575–583. [PubMed: 20232398]
- Tulving E. Memory and consciousness. *Canadian Psychology*. 1985; 26:1–12.
- Vilberg KL, Rugg MD. Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia*. 2007; 45:2216–2225. [PubMed: 17449068]

- Wais PE, Mickes L, Wixted JT. Remember/know judgments probe degrees of recollection. *J Cogn Neurosci*. 2008; 20:400–405. [PubMed: 18004949]
- Wais PE, Squire LR, Wixted JT. In search of recollection and familiarity signals in the hippocampus. *J Cogn Neurosci*. 2010; 22:109–123. [PubMed: 19199424]
- Wixted JT, Squire LR. The medial temporal lobe and the attributes of memory. *Trends Cogn Sci*. 2011; 15:210–217. [PubMed: 21481629]
- Wixted JT, Mickes L. A continuous dual-process model of remember/know judgments. *Psych Rev*. 2010; 117:1025–1054.
- Wixted JT, Mickes L, Squire LR. Measuring recollection and familiarity in the medial temporal lobe. *Hippocampus*. 2010; 20:1195–1205. [PubMed: 20848603]
- Yonelinas AP. Components of episodic memory: the contribution of recollection and familiarity. *Philos Trans R Soc Lond B Biol Sci*. 2001; 356:1363–1374. [PubMed: 11571028]
- Yonelinas AP. The nature of recollection and familiarity: a review of 30 years of research. *J Mem Lang*. 2002; 46:411–517.
- Yonelinas AP, Jacoby LJ. Noncriterial recollection: familiarity as automatic, irrelevant recollection. *Conscious Cogn*. 1996; 5:131–141.
- Yonelinas AP, Otten LJ, Shaw KN, Rugg MD. Separating the brain regions involved in recollection and familiarity in recognition memory. *J Neurosci*. 2005; 25:3002–3008. [PubMed: 15772360]

Figure 1.

a: Recognition accuracy as a function of the subsequent source memory judgment. b: Proportion of trials associated with accurate source memory judgment in each response category. Dotted line indicates chance performance.

Figure 2. Outcome of the omnibus ANOVA employed to identify response category effects ($P < .001$, 2-tailed). The sections are at the levels of parahippocampal cortex ($y = -40$), entorhinal cortex ($y = -25$), and the anterior hippocampus ($y = -16$).

Figure 3.

a: Outcome of the pair-wise contrast between the activity elicited by R-Weak and K-All items. The three sections are at the levels of parahippocampal cortex ($y = -40$), entorhinal cortex ($y = -25$), and anterior hippocampus ($y = -16$). b: Coronal sections as in 3a for the pair-wise contrast between the activity elicited by R-High and R-Weak items. c: Left anterior hippocampal voxels demonstrating greater activity ($P < .05$, 1-tailed) for the R-High than the R-Mod categories (left: 13 voxels), and for the R-Mod than the R-Weak categories (right: 22 voxels). Sections at $y = -16$.

Figure 4.

Parameter estimates from the peak voxel in the left anterior hippocampus ($-30 -16 -20$) identified by the one-way ANOVA. a: Average parameter estimates across all subjects for each response category. b: Estimates for pairs of response categories from subject subgroups in whom recognition accuracy was equated. Left: R-High and R-Mod; Right: R-Mod and R-Weak.

Table 1

Proportions of studied and unstudied test items accorded ‘Remember’ (‘R’), ‘Know’ (‘K’), or ‘New’ (‘N’).

Studied items		Unstudied Items			
R	K	N	R	K	N
0.65	0.24	0.11	0.04	0.12	0.85

Table 2

Proportions of correctly recognized studied items endorsed R and K and accorded correct and incorrect confidence source judgments. Correct judgments are segregated by response confidence.

	High	Mod	Low	Incorrect
Remember ('R')	0.40	0.25	0.12	0.23
Know ('K')	0.07	0.29	0.33	0.32