Behavioral/Systems/Cognitive

Functional Magnetic Resonance Imaging Activity during the Gradual Acquisition and Expression of Paired-Associate Memory

Jon R. Law, Marci A. Flanery, Sylvia Wirth, Marianna Yanike, Anne C. Smith, Loren M. Frank, Wendy A. Suzuki, Emery N. Brown, and Craig E. L. Stark

Recent neurophysiological findings from the monkey hippocampus showed dramatic changes in the firing rate of individual hippocampal cells as a function of learning new associations. To extend these findings to humans, we used blood oxygenation level-dependent (BOLD) functional magnetic resonance imaging (fMRI) to examine the patterns of brain activity during learning of an analogous associative task. We observed bilateral, monotonic increases in activity during learning not only in the hippocampus but also in the parahippocampal and right perirhinal cortices. In addition, activity related to simple novelty signals was observed throughout the medial temporal lobe (MTL) memory system and in several frontal regions. A contrasting pattern was observed in a frontoparietal network in which a high level of activity was sustained until the association was well learned, at which point the activity decreased to baseline. Thus, we found that associative learning in humans is accompanied by striking increases in BOLD fMRI activity throughout the MTL as well as in the cingulate cortex and frontal lobe, consistent with neurophysiological findings in the monkey hippocampus. The finding that both the hippocampus and surrounding MTL cortex exhibited similar associative learning and novelty signals argues strongly against the view that there is a clear division of labor in the MTL in which the hippocampus is essential for forming associations and the cortex is involved in novelty detection. A second experiment addressed a striking aspect of the data from the first experiment by demonstrating a substantial effect of baseline task difficulty on MTL activity capable of rendering mnemonic activity as either "positive" or "negative."

Key words: medial temporal lobe; hippocampus; recollection; associative; explicit; declarative

Introduction

Memory researchers have long known that structures within the medial temporal lobe (MTL), including the hippocampal region (CA fields, dentate gyrus, and subiculum) and the parahippocampal, perirhinal, and entorhinal cortices, are necessary for normal declarative memory (for review, see Squire et al., 2004). Arbitrary association learning is one category of declarative memory that is particularly sensitive to selective damage of the MTL (for review, see Eichenbaum, 2000; Brown and Aggleton, 2001). To explore the patterns of neural activity during arbitrary association learning, Wirth et al. (2003) recorded single-unit activity in the hippocampus while monkeys performed an associative learning task. In this task, one of four saccades was randomly associated with each of several visual scenes on a trial-and-error basis (location–scene association task). Twenty-eight percent of the scene-selective hippocampal neurons exhibited changes in firing rate correlated with the animal’s behavioral performance for a particular location–scene association (changing cells). One-half of the changing cells showed a selective increase in activity that was positively correlated with learning (sustained changing cells), whereas the remaining cells responded selectively to a particular scene before learning and subsequently signaled learning by returning to baseline levels of activity (baseline sustained changing cells).

A functional magnetic resonance imaging (fMRI) study by Toni et al. (2001) used a similar associative learning task in humans to suggest that these changes in neural activity translate to an overall increase in activity within the hippocampal/parahippocampal region. However, because of its blocked design, Toni et al. (2001) could only use an averaged learning curve rather than one based on individual trial-by-trial accuracy as used by Wirth et al. (2003). Moreover, the relative patterns of activation within the individual MTL structures during learning were not described.
To examine the activation patterns in individual MTL structures during new associative learning, we conducted an fMRI study in which subjects performed an associative learning task similar to the one described by Wirth et al. (2003). To compare activity to individual learning curves, we used the same state-space smoothing algorithm for binary observations used by Wirth et al. (2003) to compute a trial-by-trial probability correct curve for individual subjects (Smith and Brown, 2003; Smith et al., 2004). This allowed us to group the fMRI activity by the probability that the participant would be correct on each individual trial. In experiment 1, we examined both the learning-related changes in activity as well as simple novelty signals throughout the individual MTL structures as well as in areas beyond the MTL. Consistent with findings that discrete MTL structures often contribute to various declarative memory tasks in a coordinated way (Squire et al., 2004), we hypothesized that similar increases in activity would be observed throughout the MTL. Experiment 2 was performed to address an unexpected finding from experiment 1. In experiment 1, we observed less blood oxygenation level-dependent (BOLD) activity in the mnemonic paired-associate task compared with the easy baseline task. In experiment 2, we addressed this issue by examining the effect of a more difficult baseline task.

Materials and Methods

Participants. Written, informed consent was obtained with local Institutional Review Board approval from 21 healthy, right-handed participants (12 females and 7 males; age range, 18–33 years; mean age, 23.2 years; recruited by advertisement) in experiment 1. Data from two participants were discarded because of scanner problems and excess movement, leaving 19 participants for data analysis. In experiment 2, written, informed consent was obtained from six healthy, right-handed participants (three females and three males; age range, 19–35 years; mean age, 24.6 years). Participants were financially reimbursed for their time.

Materials and design. The stimuli consisted of computer-generated geometric patterns of kaleidoscope-like images (Miyaishi et al., 1991) generated randomly for each participant. Stimuli were displayed using a native Visual C application on the Windows 2000 operating system, with the Allegro 4.0 gaming library (Hargreaves, 2003).

A schematic diagram of each trial is shown in Figure 1. The task was modeled directly after the task used by Wirth et al. (2003), with changes to the stimuli (kaleidoscopic vs photographs of animals or locations) and response mode (pressing one of four buttons vs visual saccade to one of four target locations). We changed the stimulus type to reduce the ability to give a clear verbal cue to each stimulus as well as to preclude any previous experience with the represented items or locations presented in the stimulus. Each trial began with the presentation of a fixation cross for 300 ms. Next a kaleidoscope image with the outlines of four squares overlaid in a horizontal row across the image was presented for 500 ms. The four square outlines corresponded with the four button response keys. A 700 ms delay period followed, during which the stimulus was removed, leaving only the four squares and fixation cross. After the delay period, the word “GO” was presented at fixation, and responses were collected within a 700 ms response window. Participants endorsed one of the four squares as the associate for the kaleidoscope stimulus by pushing the button that corresponded with one of the four squares. Immediately after the response, the endorsed square was illuminated with a translucent white box indicating which response had been recorded. Finally, 800 ms of feedback were presented, indicating whether the endorsed response had been correct, incorrect, or not made within the response window. Because only correct/incorrect feedback was displayed, the participant was encouraged to remember correct responses and to try an alternate response after making an incorrect response. The total duration for a single trial was 3000 ms.

To provide a reference for the fMRI signal and to be able to create hemodynamic response profiles for all trial types of interest, baseline (also referred to as null) trials were interspersed randomly throughout the experiment (Dale, 1999). Baseline trials were identical to the memory trials, except that a fixed, random visual static pattern was presented instead of a kaleidoscope image. In addition, one of the four squares was assigned randomly as the target on each trial and was filled with white at 50% opacity, indicating the correct response that the participant should make during the response window. The stimuli remained on the screen until the response was recorded or the trial came to an end. Thus, these baseline trials were chosen to induce zero mnemonic demand as well as to provide a stable level of activity from trial to trial, and not necessarily define an assessment of “zero activity” (Stark and Squire, 2001a).

Experiment 2 examined the below-baseline activations observed in experiment 1 and, more broadly, the effect of different baseline tasks on interpreting fMRI results. It was therefore identical to experiment 1, except that the baseline trials were divided into two categories, easy trials and difficult trials, to investigate the effect of non-mnemonic task difficulty on MTL activity. One-half of the baseline trials involved a task identical to the one used in experiment 1, in which the target visual stimulus had a 50% opacity value and the three distracter stimuli were transparent. These trials were used as the baseline in the easy baseline condition. On the remaining baseline trials, the perceptual discrimination was made more difficult. Here, the opacity of the three distracter stimuli were set equal to each other and would vary between 0 and 50% opacity on any particular trial. The opacity of the target stimulus was then set to be between 11 and 17% greater than the distracter stimuli (i.e., 100% difference was present in the easy trials). The task during the difficult baseline trials was to choose which of the four squares on the screen was the brighter (i.e., which had the higher opacity value). The three distracter stimuli would always have the same opacity, and the target would be slightly lighter. The offset value used for the target stimulus was determined for each subject during a prescan diagnostic training session. The goal was to find a level for the target offset that would result in the subject being correct on ~50% of the difficult baseline trials (chance, 25%).

Prescan training. Twenty-four to 48 h before scanning, participants were pretrained on a set of four stimuli (the reference set), which was used in all sessions for a single participant. A training run consisted of 102 trials (duration: 5 ms, 6 s), of which 72 were stimulus presentation trials (18 trials for each of the four reference images) and 30 were baseline task trials (one-half of each baseline type for experiment 2), and was given twice. If a participant was still unable to accurately perform the task, the training run was repeated until the participant was able to select the correct association for all four stimuli. After training on the reference set, each participant was also given a single diagnostic run consisting of eight
new stimuli. This diagnostic run was used to set the initial number of concurrent images a participant was to learn during the subsequent scanning session.

By familiarizing and training the participant with the reference set stimuli, we were able to compare memories that were in the process of formation with those that had been well learned some time ago. By the scan session, each stimulus in the reference set would have been seen a minimum of 42 times (18 trials per training run by two training runs, plus 7 trials in a diagnostic run; see below, Scanning session), the association learned until the participant could make the correct choice eight times in a row, and at least 24 h had passed. These stimuli therefore provided a reference point for strongly learned memories. In addition, it is possible that some degree of cellular consolidation had already occurred, thus allowing us to compare newly learned memories with recently learned memories.

**Scanning session.** Immediately before scanning, participants were given two additional training sessions on the reference set of images: once outside the scanner and once inside the scanner (a total of 36 additional trials per reference image). This ensured strong memory for the reference set and acquainted the participant with performing the task in the scanning environment. By now, the participant would have seen each reference stimulus a minimum of 78 times and made the correctly associated response for the majority of them.

During each test run, participants completed 72 associative learning trials, 30 baseline trials, and 30 reference task trials (132 trials; 6 m, 36 s) while fMRI data were collected. Whereas Wirth et al. (2003) showed monkeys four stimuli concurrently, we found that when humans were given four concurrent stimuli to learn, they acquired the associations very rapidly. To maximize the number of trials during which participants were actively learning new associations (i.e., during the transition from very poor knowledge of the association to very strong knowledge of the association), we tailored the task to each participant’s performance in two ways. First, 4, 8, or 12 novel images could be presented to the participant to be learned simultaneously during each test run. As the number of concurrent images increased, the difficulty increased and participants required more trials to learn them. Pilot data indicated that these values sufficiently spanned the range of difficulty to yield an approximately equal number of trials at each memory strength (see below, Computation of the learning curve and memory strength index). Second, training on a given set of images could span over more than one run. Because a run was composed of a fixed number of 72 stimuli trials, a run with 12 concurrent stimuli would afford the participant only 6 presentations of a given stimulus, whereas a run with 4 concurrent stimuli would afford 18 presentations of each stimulus. By presenting the same set of stimuli for two or more runs in a row, participants would be given more trials to learn the association. As the difficulty level continued to increase, scoring after each run, we varied both the number of concurrent stimuli and the number of repeated runs to maximize our available data. We attempted to reach a predetermined empirical criterion of six consecutive correct responses for at least one-half of the stimuli within one or two runs. Because some images were not learned after two runs, this criterion was a guideline and not a mandatory objective.

**fMRI imaging parameters.** Imaging was conducted on a 3.0 tesla Philips scanner equipped with a SENSE (SENSitivity Encoding) head coil at the F. M. Kirby Research Center for Functional Brain Imaging at the Kennedy Krieger Institute (Baltimore, MD). By exploiting the sensitivity profiles of multiple surface coils, SENSE imaging can undersample k-space with fewer phase encoding steps while still yielding full field-of-view images that are free of aliasing. The result is significantly reduced acquisition time and distortion attributable to magnetic susceptibility (Pruessmann et al., 1999). Functional echoplanar images were collected using a high-speed echoplanar single-shot pulse sequence with a matrix size of $80 \times 80$, an echo time of $30$ ms, a flip angle of $70^\circ$, a SENSE factor of 2, and an in-plane resolution of $3 \times 3$ mm. In each run, a total of 264 whole-brain, three-dimensional volumes were acquired with a repetition time (TR) of 1.5 s (132 trials per run, 3 s each, two acquisitions per trial). Each volume consisted of 30 triple oblique axial slices (3 mm thickness with a 1 mm interslice gap), aligned to the principle axis of both the left and right hippocampi. Data acquisition began after the fourth image to allow for stabilization of the MR signal. For anatomical localization, a standard whole-brain, three-dimensional magnetization-prepared rapid gradient echo (MP-RAGE) pulse sequence was acquired (180 1 mm³ oblique axial slices).

**Computation of the learning curve and memory strength index.** To convert the binary performance data into an estimate of memory strength, we adopted the logistic regression algorithm developed by Brown and colleagues (Smith and Brown, 2003; Smith et al., 2004) and used by Wirth et al. (2003) in their analysis of the correlation between spiking rate in the hippocampus and behavioral performance. In this approach, the data begin as a sequence of binary responses across the multiple trials for each stimulus (1 for a correct response and 0 for an incorrect response). To analyze these data, we estimated the learning curve for each stimulus. This was the probability of a correct response as a function of trial number using a state-space smoothing algorithm, in which the state equation was a Gaussian random-walk model and the observation equation was a Bernoulli model (Wirth et al., 2003; Smith et al., 2004). The algorithm uses an approximate expectation maximization algorithm (Dempster et al., 1977) to compute the maximum likelihood estimate of the learning curve as well as its associated confidence intervals. The learning curve provided a quantitative measure of how well a memory was acquired at each trial. To analyze the relationship between learning performance and the dynamics of fMRI activation, we derived a memory strength index from the learning curve for each scene using the following transformation: memory strength index 1, learning curve 0–0.2; memory strength index 2, learning curve 0.2–0.4; memory strength index 3, learning curve 0.4–0.6; memory strength index 4, learning curve 0.6–0.8; memory strength index 5, learning curve 0.8–1.0.

**Analysis of fMRI data.** Image analysis was performed using Analysis of Functional Neuroimages (AFNI) software (Cox, 1996). fMRI data were first coregistered in three dimensions and through time to reduce the effects of head motion. Six motion vectors coding the rotations and translations necessary to align the functional images were created during the coregistration process. The data from all runs from a given participant were then concatenated and analyzed using a deconvolution approach based on multiple linear regression (3dDeconvolve; http://afni.nimh.nih.gov/pub/dist/doc/manuals/3dDeconvolve.pdf). The analysis included nuisance vectors coding for drift in the MR signal, the six motion vectors, and eight sets of time-shifted vectors coded to represent the experimental conditions of interest as discussed below.

In the electrophysiological study of Wirth et al. (2003), the spike rate for individual images was analyzed as a function of trial number during learning. Unfortunately, our fMRI data do not have a sufficient signal-to-noise ratio to support an equivalent analysis. Instead, we grouped trials together based on the estimate of memory strength for each trial. By lumping together trials based on their estimated memory strength index, we were able to circumvent difficulties associated with the variance between participants in their ability to learn images and the variance within a participant in the number of trials to learn each particular image. In doing so, we acknowledge that we made the fundamental assumption that the algorithm used to assess memory strength from the binary behavioral performance data are at least monotonically related to the actual strength of a participant’s memory for the image–response association.

Of the eight vectors used to code for the effects of interest, five vectors were used to code for the memory strength of the associative learning trials based on using five bins to evenly span the full range from 0.0 to 1.0 (Fig. 2). Five bins prevented an insufficient number of trials per bin, as well asaffording sufficient resolution to detect more than simple linear or second-order trends. In addition, we separately identified and coded the initial presentation of each stimulus (these trials were therefore excluded from the lowest memory strength bin to avoid duplication of coding between conditions), because numerous neuroimaging studies have reported a “novelty” or “familiarity” effect in the form of a large drop in activity between the first and subsequent presentations of a stimulus (for review, see Schacter and Wagner, 1999). We included two final vectors to code the reference stimuli: one vector coded for reference trials during the first half of each run and a second vector coded for reference trials during the second half of each run. By coding the reference trials this way, we provided a within-participant control for the possibility that any
increase or decrease in the BOLD effect observed during acquisition of the training set might be an artifact resulting from a global drift over time. If some effect of time altered the signal and induced an artifactual effect of memory strength (e.g., via fatigue, attention, or unaccounted for scanner drift), one might assume that it would have a similar effect on the reference trials, showing a difference in BOLD signal between the first and second half of the reference trials.

In the deconvolution technique, multiple regression is used to estimate the hemodynamic responses to each of the vectors of interest. For each of our eight vectors, 11 time-shifted versions were created (each one shifted one TR further), coding for the effect on the BOLD activity from 0 to 16.5 s after trial onset. The baseline trials served as an estimate of the baseline or zero activity level and therefore were not entered into the model. All resulting fit coefficients (β coefficients) therefore represent activity for a particular trial time at a particular point in time relative to the baseline trials, forming an estimate of the hemodynamic response in each voxel and to each trial type directly from the data. The sum of the regression coefficients for the time points corresponding to the expected sum over the hemodynamic response (3–12 s after stimulus onset) was taken as the estimate of the model of the response to each trial type.

Two analyses were performed on the resulting statistical maps: one version focusing only on the MTLs and a second whole-brain analysis. For the MTL analysis, we used a modified version of the regions of interest-based alignment (ROI-AL) technique (Okado and Stark, 2003) that greatly improves alignment within the MTL, increasing statistical power and precision in cross-subject tests. The technique begins with the manual segmentation of ROIs based on anatomical structure in the Talairach-aligned high-resolution structural MP-RAGE scans (Talairach alignment provides a reasonable starting place for additional fine-tuning). Here, 10 structures in the MTL (bilateral hippocampal region and temporopolar, perirhinal, entorhinal, and parahippocampal cortices) were defined. The temporopolar portion of the perirhinal cortex, the more posterior portion of the perirhinal cortex, and the entorhinal cortex were defined according to the techniques described by Insausti et al. (1998). As in our previous research, the parahippocampal cortex was also defined bilaterally as the portion of the parahippocampal gyrus caudal to the perirhinal cortex and rostral to the splenium of the corpus callosum. The hippocampal region (the CA fields of the hippocampus, dentate gyrus, and subiculum) was also defined bilaterally.

The original version of ROI-AL technique (Stark and Okado, 2003) uses these segmentations to calculate a single 12-parameter affine transformation matrix to fine-tune the cross-participant alignment of the segmentations. The goal is to maximize the overlap of each region such that the hippocampus will align with the hippocampus, the perirhinal cortex will align with the perirhinal cortex, etc. A mathematical objective function implementing this goal is used in standard least-squares optimization of the 12-parameter transformation. In the study by Stark and Okado (2003), an MTL model had been constructed from the 20 participants’ segmentations by calculating the modal value of the aligned segmentations across participants, and a bootstrapping technique had been used to refine the model. That model, which represents a modal MTL, approximately in Talairach coordinates, was used here as the template. By using such a template, the location of results in group analyses can be compared directly to the model. In so doing, activity in the group can be classified or segmented based on known anatomical definitions from the individuals.

Here, the ROI-AL technique was modified to include a separate 12-parameter affine transformation matrix for each individual ROI (but see Okado and Stark, 2005). Thus, 10 transformation matrices were created, and 10 versions of each statistical map (and structural image) were created for each participant. To better interpret the data, the results of the 10 transformations were combined into a single statistical map. This was done in such a way that the value at any given voxel was a weighted average of the values from each of the 10 maps. Weights for a given voxel were determined by estimating the distance from each ROI to that voxel. Rather than using simple Euclidian distance from the center of mass, this distance was estimated by blurring each transformed ROI by a 5 mm full-width at half-maximum (FWHM) Gaussian kernel. Thus, the binary segmentation became a blurred map, indicating the area of influence this particular ROI was allowed to have over the combined output. Where overlap across blurred ROIs occurred, the voxel in the combination was weighted by the degree of influence (value in the blur) from each contributing ROI. Because each of the 10 transformations warps all of the data, the discrepancies across maps for these border voxels are not usually large. However, by using this weighted-average technique, we eliminate the possibility of unwanted edge effects that might arise with a more strict combination method and that might hamper cross-participant random-effects analyses. Overall, we found that by extending the ROI-AL in this manner, additional gains in statistical power and precision are achieved.

We note that when the ROI-AL technique is used to align MTL structures, coordinates in a standardized reference frame such as Talairach or MNI (Montreal Neurological Institute) space become somewhat deformed within the MTL and that data well outside the aligned regions can be grossly distorted. Therefore, to assess whole-brain activity outside of the MTL, a separate analysis was conducted on whole-brain data after alignment to the Talairach atlas (Talairach and Tournoux, 1988) (we should note the two alignments result in different voxels from each individual going into each voxel in the group tests, making these analyses independent). In both analyses, individual statistical maps were resampled to a resolution of 2.5 mm3 and smoothed with a Gaussian filter to help account for variations in the functional anatomy (MTL analysis, 2 mm; whole-brain analysis, 4 mm).

Analysis of functional ROIs. All of our analyses were conducted on functionally defined ROIs either in the entire brain or restricted to the MTL. To identify regions where activity changed in any way during the acquisition of the new associations, we used a repeated-measures ANOVA on the BOLD activity for memory strength indices 1, 3, and 5 (treating the participant as a random effect). In this way, we can identify regions that show a linear or second-order pattern using only a subset of the data and still remain highly sensitive to stepwise changes (e.g., similar activity for memory strength indices 2–5 and dissimilar activity for memory strength index 1). The resulting activation map identified voxels that differed significantly across these memory strengths without being biased toward any particular pattern of activity. It also allowed for the assessment of activity for trial types that were not included in the voxel-selection procedure (i.e., trials of memory strength indices 2 and 4, first presentation trials, and both groups of reference trials).

This technique allowed us to assess changes in activity over the course of gradual learning and to assess whether the activity of a region changed evenly over the course of learning or whether it changed more rapidly in earlier or later phases of learning (here, by assessing any curvilinearity in the relationship between memory strength and activity). Again, only a portion of the data was used to determine our ROIs. Activities associated with the first presentation, memory strength indices 2 and 4, and both reference conditions were free to vary.

In the MTL analysis, the results of the ANOVA were subjected to a voxel-wise threshold of \( p < 0.02 \) and a spatial extent threshold of 108
mm³ (corresponding α; when corrected for multiple comparisons equals 0.22) to create functionally defined ROIs. Data from all voxels within each region were then collapsed so that individual ROIs could be interrogated as to the specific effects of memory strength. Subsequent analyses were conducted using an α threshold of 0.05, which represents the final probability of a false positive in the complete analysis. The same process was used in the whole-brain data, but with a voxel-wise threshold of $p < 0.001$ and a spatial extent threshold of 140 mm³ ($α$ corrected for multiple comparisons, $<0.05$) to compensate for the greater number of comparisons being made. In both cases, $α$ values were determined by Monte Carlo simulation of intensity and spatial extent thresholds using the AlphaSim component of AFNI (Cox, 1996). Here, voxel size was set at 2.5 mm³, cluster connectivity was set at a radius of 2.6 mm, and a 4 mm FWHM smoothness was assumed for both estimates with the number of voxels and shape of the region based on masks of the MTL and of the Talairach brain.

**Results**

**Experiment 1**

Across all 19 participants, the mean number of new stimuli learned per run was 8.8 (SD, 2.0; range, 4–12), and the mean number of training trials to criterion was 5.2 (SD, 3.3; range, 2–17). Because the number of new stimuli presented in a run was tailored to the performance of individual participants, there was no clear relationship between the number of new concurrent stimuli and training trials to criterion (4.8, 3.9, 5.5, and 4.7 for 4, 6, 8, and 12 concurrent stimuli, respectively).

Reaction times to make a correct response for trials in each trial condition are shown in Table 1 (times are relative to the start of the response window). A repeated-measures ANOVA revealed a significant effect of trial condition on reaction time ($F(7,126) = 9.79; p < 0.001$). An examination of the reaction times showed a clear clustering effect with memory strength indices 1–3 forming one group and memory strength indices 4 and 5 and the two reference conditions forming the other group (reaction times to the first presentation were between those two groups). Although reliable in places (e.g., a paired $t$ test between memory strength index 3 and memory strength index 4 yielded a $t_{110} = 4.7; p < 0.001$), the absolute difference in reaction time is quite small, averaging only 30 ms between the two clusters. Given the small size of the effect, the direction of the effect (reduced reaction time in conditions with predicted greater activity), and its stepwise nature (in contrast to the predicted smooth relationship between memory strength and activity), we do not feel the reaction time difference significantly affected the fMRI results.

**fMRI results**

In two separate analyses of the fMRI data, we identified regions where changes in activity were correlated with memory strength. Because our principal focus was to identify functional differences among the MTL structures, our first analysis focused exclusively on this region. Second, because regions outside the MTL are undoubtedly involved in the acquisition and expression of paired-associate learning, a second whole-brain analysis was conducted using standard Talairach alignment.

**MTL analysis**

In the analysis constrained to the MTL, the ANOVA revealed five subregions where activity varied over the course of learning (Fig. 3). This learning-related activity was observed bilaterally in the head of the hippocampal region, bilaterally in the parahippocampal cortex, and in the right perirhinal cortex. These regions were treated as functional ROIs by collapsing across all voxels within each region and examining the average activity for each trial type within each participant. Overall, the results were highly consistent across all five regions. In both hippocampal regions as well as in the right perirhinal cortex, there was a significant initial drop in activity (all $r > 3; p < 0.001$) from the first presentation (Fig. 3, gray bars) to the memory strength index 1 trials in which the paired-associate knowledge had not been acquired but the stimuli were no longer novel. In the right and left parahippocampal regions, there were trends in this direction, but they were not significant.

After the drop observed in the initial presentation, there was a monotonic and generally linear increase in activity across the five memory strengths (red bars) within each of the MTL regions. An ANOVA was used to assess the shape of the relationship between BOLD fMRI activity and memory strength in all five regions. In each, there was a significant linear component (all $F > 7.7; p < 0.001$). In only the right parahippocampal cortex did we observe a significant quadratic component ($F = 6.9; p < 0.02$). In the right perirhinal cortex, there was a nonsignificant trend in this direction ($F = 2.5; p = 0.13; all other $p > 0.25$). No regions exhibited significant cubic components (all $p > 0.5$). Finally, once the correct response to the stimulus was learned well, activity in all of the regions stabilized, with memory strength index 5 trials exhibiting similar levels of activity as the very well learned reference trials (Fig. 3, blue bars).

One potential confound in this analysis was that memory strength would almost certainly correlate with time during the scan. Any artifact related to time that was not accounted for by our analysis could therefore have affected our results (e.g., drift in the sensitivity of the scanner to the BOLD response). To address this possibility, we analyzed the reference trials from the first half of each run (Fig. 3, left blue bar) separately from the second half of each run (Fig. 3, right blue bar). If the increase in activity as a function of memory strength was the result of a scanning artifact, one would predict a similar effect on the reference trials. No such effect was observed (all $p > 0.23$), indicating that the increase in activity across memory strengths was the result of changes in neural activity rather than time-based artifacts.

Thus, there were several clear results within the MTL. Activity in both the left and right hippocampal regions, the right perirhinal cortex, and the left and right parahippocampal cortices increased monotonically as a function of memory strength (disregarding the first presentation and reference images). This relationship was strikingly linear in the bilateral hippocampal region and in the left parahippocampal cortex but showed evidence of nonlinearity in the right parahippocampal cortex and a trend toward this in the right perirhinal cortex as well. However, it is clear that in each of these MTL regions, we observed activity correlated with behavioral indications of paired-associate memory strength. Furthermore, we observed simple novelty or “reency” signals (Brown and Aggleton, 2001) (our task does not allow us to differentiate between the two) in the same left and

<table>
<thead>
<tr>
<th>Table 1. Reaction time versus trial type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory strength</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>First</td>
</tr>
<tr>
<td>Str 1</td>
</tr>
<tr>
<td>Str 2</td>
</tr>
<tr>
<td>Str 3</td>
</tr>
<tr>
<td>Str 4</td>
</tr>
<tr>
<td>Str 5</td>
</tr>
<tr>
<td>Ref 1</td>
</tr>
<tr>
<td>Ref 2</td>
</tr>
</tbody>
</table>

Str, Strength; Ref, reference.
right hippocampal regions and the right perirhinal cortex that showed this paired-associate learning.

**Rate change in activation during gradual memory acquisition**

The fMRI memory strength curves in the right perirhinal and parahippocampal cortices were curvilinear, showing a steep slope for early learning that subsequently decreased as learning occurred. There are several possible explanations for this effect. First, Wirth et al. (2003) observed that some neurons in the hippocampus exhibited changing activity before the monkeys showed any behavioral indications of learning, suggesting that these neurons might be important for earlier stages of learning. Similarly, our use of novel and complex geometric shapes could induce greater involvement in right-hemisphere regions for the early, single-item stages of learning.

Second, the curvilinear trend could have been the result of rapid learning during the early memory strengths, followed by a saturation or ceiling effect. Our data cannot determine whether such saturation has occurred. Third, the linear versus curvilinear relationship is heavily reliant on the output of the behavioral algorithm, which may or may not produce a linear division between memory strengths. Finally, the BOLD effect itself might not have a linear response. Thus, we are more certain about the direction and monotonic nature of the observed activity change in relation to the memory strength than we are of the rate of change. However, any such collection of potential nonlinearities must offset each other perfectly to arrive at the strikingly linear relationship.

### Whole-brain analysis

A separate analysis was conducted on the whole-brain data, again identifying regions where activity varied across memory strength indices 1, 3, and 5. Regions identified in this analysis included the left hippocampus, superior frontal gyrus, medial frontal gyrus, middle frontal gyrus, inferior frontal gyrus, cingulate gyrus, left fusiform gyrus, left superior temporal gyrus, bilateral precuneus, bilateral insula, right superior parietal lobe, right subcollasal gyrus, left supramarginal gyrus, and right caudate. These areas are shown in Figure 4 and listed in Table 2. The activity in all of these regions could be clearly classified into one of two response patterns. One pattern was similar to that observed in the MTL analysis (the “rising” pattern, coded as a yellow overlay) and reflected activity that increased linearly with memory strength. An example of this pattern is shown in Figure 4b (top). The same rising pattern of activity in the MTL was also observed in the medial frontal pole and left superior frontal gyrus, suggesting a widespread network for association learning.

In contrast, a second pattern emerged (the “drop-off” pattern, coded as a red overlay) in which activity was relatively constant for the first condition and memory strength indices 1–4 but fell precipitously in the memory strength index 5 and reference trials (in post hoc t tests, all p < 0.001 when comparing strength index 4 with strength index 5). An example of this drop-off pattern is shown in Figure 4b (bottom), occurring primarily in the dorsal lateral prefrontal cortex and the superior parietal lobule. These areas have been implicated in a frontoparietal system for attention, effort, and working memory network (Cabeza and Nyberg, 2003).
Here, the pattern was strong and stable until the memory was learned well, consistent with a role for these areas in directed attention, retrieval effort, or possibly retrieval monitoring.

**Experiment 2**

One concern with the results from experiment 1 that we have not yet discussed is the location of zero on the y-axis of Figures 3 and 4. Here, zero represents activity associated with the baseline task, which was to simply indicate which of four white square outlines was filled with an opaque rather than a transparent background. Because this is a task that presumably does not require any long-term or working memory, we had anticipated that performance during these trials would result in a low-level of MTL activity relative to the experimental task. However, contrary to this expectation, activity in several MTL regions while on-task never rose appreciably above the baseline level. This result might suggest that the process of making well learned associations does not require these MTL regions. An alternative view is that this baseline task was associated with unintentional and incidental MTL activity. In fact, several studies have reported similar results in which there is greater activity during rest compared with during the active task (Gusnard and Raichle, 2001; Newman et al., 2001; Stark and Squire, 2001b; Christoff et al., 2004), with one of these reporting greater levels of MTL activity during rest and several other low-level baselines than during active, but non-mnemonic, baselines (Stark and Squire, 2001b). Furthermore, other studies have shown that activity in widespread regions of the brain including the frontal, cingulate, and parietal regions varies with the difficulty of the low-level baseline task (McKiernan et al., 2003).

Experiment 2 was designed to assess whether the below-baseline results of experiment 1 might be attributed to elevated activity during the simple baseline task. If the participants’ minds wandered during the easy baseline trials of experiment 1, incidental encoding and retrieval during the baseline task would likely ensue, resulting in increased activity in the MTL. To test this possibility, experiment 2 was identical to experiment 1, except that a difficult baseline condition was substituted for one-half of the baseline trials. The trials in the difficult baseline condition were made more perceptually difficult by adjusting the opacity of all four boxes with only one box (the target) being slightly brighter than the other three (the distractors) (Fig. 5).

**Behavioral results**

The primary factor of interest in experiment 2 was the manipulation of the baseline task. The baseline task used in experiment 1 was relatively easy, identifying a relatively opaque box versus three entirely transparent boxes on a computer screen. Performance in this easy baseline condition by the six participants in experiment 2 was at ceiling with an average of 98% correct. Performance in the difficult baseline task was significantly worse ($t_{(5)} = 11; p < 0.001$), averaging only 54% (chance is 25% correct).

**fMRI results**

In experiment 2, the only variable that was changed from experiment 1 was the introduction of a second baseline task with an increased level of perceptual difficulty that could serve as an alternate baseline. The fMRI data were analyzed twice: once using the easy baseline trials as the estimate of zero activity and a second time using the difficult baseline trials as the estimate of zero activity. Unfortunately, the smaller sample size in experiment 2 reduced our statistical power, making it impossible to define new functional ROIs and extract meaningful data from areas outside the MTL. Because the aim of experiment 2 was to clarify the results of experiment 1 within the context of the MTL, we interrogated the same five functionally determined ROIs from experiment 1.

Results from the left hippocampal ROI are shown in Figure 5. Activity for the five memory strength trials, the reference trials, and the difficult baseline trials is shown relative to the easy baseline trials in Figure 5a. In contrast, activity for the five memory strength trials, the reference trials, and the easy baseline trials are shown relative to the difficult baseline trials in Figure 5b. We emphasize here that Figure 5, a and b, plots the same memory strength data. The only change between the two panels is the choice of the baseline and hence the position of zero. Mathematically, the relationships between memory strengths are nearly identical in the two analyses: minor variances in the actual values may have resulted from the analysis methodology. As in experiment 1, BOLD activity for the paired-associate learning trials rose with memory strength regardless of the baseline condition, reaching a level similar to the well learned reference trials by memory strength index 5.

Most strikingly, switching between the two baselines caused a very large overall shift in the level of activity for the trial types of...
interest. This shift is readily seen in the change in direction of activity for each memory strength but even more clearly in the direct comparison of the two baseline tasks. When the easy baseline was used as an estimate of zero (Fig. 5a), the difficult baseline (black bar) was associated with significantly “less than zero” activity ($t_{(5)} = 6.344; p < 0.001$). Conversely, when the difficult baseline was used as an estimate of zero (Fig. 5b), the easy baseline (black bar) was associated with significantly “more than zero” activity ($t_{(5)} = 6.347; p < 0.001$).

Thus, in the left hippocampal ROI from experiment 1, there was significantly more activity during the easy baseline trials than for the difficult baseline trials. This pattern was also reliably observed in the right hippocampal region ($t > 4.314; p < 0.007$) and approached significance in the left parahippocampal cortex ($t > 2.286; p < 0.07$). In contrast, activity during the memory task in experiment 1 was consistently at or below zero (in the right perirhinal and right parahippocampal cortices, activity was mostly above baseline in experiment 1). The most parsimonious account is that the easy baseline trials did not sufficiently engage participants in the task, allowing their minds to wander and engage in non-task-related processing that required the encoding and retrieving of information. The pattern of results was not distorted between conditions of interest (in which baseline levels of activity have no real effect and are mathematically removed in the direct contrasts), but rather in the relationship to our estimate of zero activity in the region. Simply switching from the easy baseline task to a difficult baseline task resulted in the activity for memory strengths to switch from “negative” (below baseline) to “positive” (above baseline) and from a “downward-going” estimate of the hemodynamic response (Fig. 5c, dashed and gray lines) to an “upward-going” estimate of the hemodynamic response (Fig. 5c, solid line).

Finally, although only six participants were scanned in experiment 2, we were able to replicate several results from experiment 1 regarding activity change as a function of memory strength. For example, both the left and right hippocampal ROIs showed clear evidence of a linear relationship across the five memory strengths (both $p < 0.06$), and the left parahippocampal cortex activity demonstrated a purely monotonic relationship. Two factors precluded any additional detailed analysis of the relationship between the fMRI signal and memory. First, the overall sample size here was small with approximately one-third as many participants as in experiment 1. Second, two of the participants anomalously recorded very few trials in the two lowest memory strength bins (under 15 trials per bin), further adding noise to an already small sample. For comparison, participants in experiment 1 averaged ~60 trials in both of these strength bins, and none had fewer than 30 trials. Were the goal of experiment 2 the same as the goal of experiment 1, both would have been removed from the analysis, because the small number of trials prohibits an accurate assessment of activity in these bins. Neither of these two factors impeded the central aim of experiment 2, assessing the global effect of baseline difficulty on activity.

**Discussion**

Wirth et al. (2003) reported that monkey hippocampal cells changed their firing rate in conjunction with the learning of an arbitrary paired association. Using a similar behavioral paradigm in humans, we observed clear monotonic increases in the BOLD fMRI signal throughout individual MTL areas during the gradual acquisition of arbitrary associations. As participants learned the correct response associated with a stimulus through trial and error, activity in the left and right hippocampal, right perirhinal cortex, and left and right parahippocampal cortices increased monotonically in conjunction with behavioral accuracy, creating an “fMRI memory strength” signal. These findings confirm and extend the findings of Toni et al. (2001), who observed an overall increase in activity within the hippocampal/parahippocampal re-
region but did not differentiate between the different areas within the MTL. Because the BOLD fMRI signal has been shown to be more closely related to synaptic activity than to spike rate (Logothetis et al., 2001), and because components of the BOLD signal have been shown to increase rather than decrease in a local region after stimulating inhibitory input pathways (Caesar et al., 2003a,b), it is not surprising that the selective increases and decreases in neural activity observed at the single-cell level by Wirth et al. (2003) translated to an overall increase in the BOLD fMRI signal as learning progressed. Outside the MTL, regions within the temporal lobes, frontal lobes, and cingulate cortex displayed similar increases in activity. In contrast, in certain frontal and parietal regions, activity remained constant during learning and sharply decreased after learning had become relatively complete. These findings together will provide important predictions for future single-unit neurophysiological studies that explore brain areas beyond the hippocampus.

Over the course of learning the association, we can assume that processes involved in encoding and retrieval for the associative aspects increase relative to those for single-item aspects. Thus, the monotonic increases in activity observed in the MTL, frontal, and cingulate cortices over these trials are best interpreted as being related to the encoding and/or retrieval of the association itself. Here, retrieval and feedback were combined on each trial, making isolation of the two components even more problematic than single-shot encoding or retrieval studies [see Stark and Okado (2003) for difficulties even with single-shot studies]. Indeed, the asymptotic relationship observed after the drop could be the result of combining a linear increase in activity for retrieval with a step function reflecting activity for encoding. However, as activity increased with memory strength and leveled off once the association was well learned, the most parsimonious account is that activity is related to successful retrieval of the association. It is worth noting that there was no interaction of activity and area was observed in which the activity in one region increases while the other decreases, which would have suggested separate neural bases for encoding and retrieval processes. Instead, it may be that the lack of such an interaction is consistent with data showing that encoding and retrieval are complementary processes occurring simultaneously in neighboring and/or monolithic structures (Stark and Okado, 2003).

In addition to these learning-related increases in activity, we also found that the first exposure to a novel stimulus evoked a strong BOLD response that was subsequently followed by a drop in activity when the stimulus was familiar but the correct associate was still unknown. The increased activity on the first presentation trials relative to the memory strength index 1 trials cannot be attributable to associative memory formation, because there were no behavioral indications that the association had yet been learned. Therefore, it is likely that some of the details of the stimulus itself had been learned after a single exposure, allowing it to be recognized on subsequent trials. This drop in activity after the initial presentation is termed a novelty or familiarity signal in electrophysiological studies (Brown and Aggleton, 2001). Likewise, in the neuroimaging literature, a drop in activity after the initial presentation has been linked to novelty and memory encoding processes (for review, see Schacter and Wagner, 1999). In this way, the enhanced activity for the first exposure to a novel stimulus is likely to be the result of encoding details of the individual stimulus. The activity for memory strength index 1 trials is consequently much less, because the stimuli are familiar at this point. This phenomenon was observed in the perirhinal cortex as well as in the bilateral hippocampus.

Here, we show that both the initial single-item encoding effect as well as the gradual associative learning pattern was observed in the hippocampus as well as in the surrounding MTL cortices. These findings are inconsistent with views suggesting a functional separation of long-term declarative functions within the MTL. For example, it has been suggested that the hippocampus is primarily responsible for complex associations and/or relationships, whereas adjacent cortical structures are more important for simple familiarity, recognition memory, or memory for individual items (Eichenbaum et al., 1994; Aggleton and Brown, 1999). If certain structures in the MTL were, in fact, responsible for single-item learning and others for associative learning, we would have observed activity resulting from the novelty effect solely in areas specifically responsible for single-item processing, and activity during the association learning trials (i.e., memory strength indices 1–4) solely in those structures responsible for associative processing. Our data did not provide evidence for a discrete separation and thus argues strongly against such a division of labor. Instead, they support the idea that both the hippocampus and parahippocampal region are broadly involved in all types of declarative memory and that the hippocampus combines and extends the processing conducted by the adjacent cortices (Squire et al., 2004).

These findings arise from the novel techniques presented here that were designed to analyze fMRI data over the course of gradual memory acquisition. The algorithm used to transform binary correct/incorrect performance into a smooth, trial-by-trial estimate of memory strength has been successfully used to model performance in a range of memory tasks (Smith et al., 2004). Here, it provides the opportunity to treat memory formation as a dynamic rather than a one-shot process and to induce a parametric fMRI design. Memory has long been known to continue to

---

**Figure 5.** Activity in the left hippocampal region (voxels selected from experiment 1) in experiment 2 as a function of baseline difficulty. **a,** Activity for all trial types relative to the easy baseline (white, first presentation; light gray, strength indices 1–5 trials; dark gray, reference trials; black, difficult baseline trials). **b,** Activity for all trial types relative to the difficult baseline. In both, activity was quantified for each stimulus type using the raw values from the general linear model estimate ($\beta$ coefficients) of the activity summed over the $-3$–$12$ s after stimulus onset. **c, d,** Sample difficult (c) and easy (d) baseline trials. **e,** Hemodynamic response for the middle memory strength index (strength 3) trials from experiment 2 relative to the easy baseline (gray line), experiment 2 relative to the difficult baseline (black line), and experiment 1 relative to its easy baseline (dashed line). Individual $\beta$ coefficients are plotted at each time point. Str, Strength; Ref, reference.
strenthen gradually with repeated presentation (Ebbinghaus, 1895, 1913). The technique presented here allows us to image changes in activity associated with this gradual process. The dynamic behavior of structures over the course of gradual learning provides a more complete picture of the neural basis of memory encoding and retrieval than previously possible with fMRI.

**Below-baseline activation**

In experiment 1, activity in the MTL during the learning trials was less than activity during what is undeniably a non-mnemonic baseline task. Potentially, this could have resulted from a task-induced deactivation, or it might be an artifact of activity during the baseline task (Stark and Squire, 2001b). Participants reported that the high difficulty of learning trials caused them to rehearse and recollect previous learning associations during the easier baseline trials, possibly causing an artificially elevated baseline. In experiment 2, we interrogated several of the MTL regions that demonstrated learning-related changes and found that activity associated with performing the original baseline task (easy) was significantly higher than activity associated with a perceptually more difficult version of the baseline task (difficult). Because both baseline tasks were specifically designed to be independent of mnemonic processing, the enhanced activity during the easy baseline task indicates that it (and hence the baseline task in experiment 1) was indeed contaminated by some amount of uncontrolled activity.

When the activity for memory acquisition was plotted using the difficult baseline task, the previously observed negative activations for memory strength indices 1–5 and reference scenes moved above the arbitrary zero axis and became positive (Fig. 5). This effect was so strong that the difference between the baselines was often as large as the difference between memory strengths. Because BOLD fMRI is an inherently contrastive or subtractive technique, any observed activity is the difference in activity between two or more conditions. From a finding of negative activity, one can only conclude that the task of interest was associated with less activity than the comparison task. Because a change in baseline task difficulty caused a mathematical reversal in the sign of the BOLD response, our data encourage the use of comparisons of activity across conditions, rather than against an arbitrary baseline.

**Conclusions**

In conclusion, these data demonstrate learning-related activity in a paired-associates, cued-recall paradigm throughout the MTL. In the left and right hippocampi and in the left parahippocampal cortex, the fMRI memory strength curve demonstrates the involvement of these regions in associative learning and retrieval. In the right perirhinal and parahippocampal cortices, the learning curve showed aspects of both single-item and associative learning. These data mirror and extend data from recent electrophysiological studies in the monkey (Wirth et al., 2003) and demonstrate the viability and power of integrating experimental designs and results across these two methodologies. Finally, the techniques described here for assessing activity related to multtrial learning open up new avenues for furthering our understanding of the perirhinal cortex and hippocampus.

**References**

Stark CEL, Squire LR (2001b) When zero is not zero: the problem of ambiguous baseline conditions in fMRI. Proc Natl Acad Sci USA 98:12760–12766.