Most functional magnetic resonance imaging (fMRI) studies examining working memory (WM) load have focused on the prefrontal cortex (PFC) and have demonstrated increased prefrontal activity with increased load. Here we examined WM load effects in the medial temporal lobe (MTL) using an fMRI Sternberg task with novel complex visual scenes. Trials consisted of 3 sequential events: 1) sample presentation (encoding), 2) delay period (maintenance), and 3) probe period (retrieval). During sample encoding, subjects saw either 2 or 4 pictures consecutively. During retrieval, subjects indicated whether the probe picture matched one of the sample pictures. Results revealed that activity in the left anterior hippocampal formation, bilateral retrosplenial area, and left amygdala was greater at retrieval for trials with larger memory load, whereas activity in the PFC was greater at encoding for trials with larger memory load. There was no load effect during the delay. When encoding, maintenance, and retrieval periods were compared with fixation, activity was present in the hippocampal body/tail and fusiform gyrus bilaterally during encoding and retrieval, but not maintenance. Bilateral dorsolateral prefrontal activity was present during maintenance, but not during encoding or retrieval. The results support models of WM predicting that activity in the MTL should be modulated by WM load.

Keywords: delayed match to sample, episodic memory, hippocampus, neuroimaging, subiculum

Introduction

Models of working memory (WM) have proposed a mechanism whereby persistent spiking at a cellular level could be used to maintain multiple items in a WM buffer (Lisman and Idiart 1995; Jensen and Lisman 1998, 2005; Koene and Hasselmo 2007). These mechanisms could also be critical for holding novel information across a delay and could allow multiple items to be linked together across time (Hasselmo and Stern 2006). These models use persistent spiking mechanisms that have been demonstrated in intracellular recordings from slices of entorhinal cortex (Klink and Alonso 1997; Egorov et al. 2002; Fransén et al. 2006) and amygdala (Egorov et al. 2006). Therefore, some of these models focus on WM in medial temporal lobe (MTL) regions instead of prefrontal cortex (PFC) (Jensen and Lisman 2005; Koene and Hasselmo 2007).

Previous functional neuroimaging work established that the PFC is recruited to maintain items in WM in humans (Jansma et al. 2000; Jha and McCarthy 2000; Druzzgal and D’Esposito 2001; Rypma et al. 2002; Woodward et al. 2006; Altamura et al. 2007) and that activity in the PFC increases when more items are being held online during the memory delay (Rypma et al. 1999, 2002; Druzzgal and D’Esposito 2003; Veltman et al. 2003; Cairo et al. 2004). The stimuli that these studies used, such as letters (Rypma et al. 1999, 2002; Veltman et al. 2003; Cairo et al. 2004; Altamura et al. 2007) or words and faces (Druzzgal and D’Esposito 2003), were highly familiar to the subjects, and these studies did not report activity in the MTL.

In contrast, recent functional magnetic resonance imaging (fMRI) studies suggest that when stimuli are trial unique, the MTL regions are recruited to maintain this novel information (Ranganath and D’Esposito 2001; Stern et al. 2001; Schon et al. 2004; Nichols et al. 2006). Furthermore, this delay activity is related to long-term memory encoding (Schon et al. 2004). These results are consistent with neurophysiological recording and functional mapping studies of WM in animals (Eacott et al. 1994; Suzuki et al. 1997; Davachi and Goldman-Rakic 2001) and with neuropsychological studies in humans that have demonstrated that the MTL is crucial for WM with novel information even for memory delays as short as 4 s (Nichols et al. 2006; Olson et al. 2006). Most critically, this research has raised doubts that WM and long-term memory are 2 separate entities as both PFC and MTL have been shown to be critical for both types of memory (Ranganath and Blumenfeld 2005; Jonides et al. 2008).

Computational modeling also predicts that MTL regions are critical for maintaining trial-unique stimuli and would be modulated by WM load (Lisman and Idiart 1995; Jensen et al. 1996; Jensen and Lisman 2005; Koene and Hasselmo 2007). Our study was designed to test the model-based prediction that fMRI activity in the anterior MTL would increase as a function of WM load. We tested this prediction by combining fMRI with a Sternberg task paradigm during which subjects had to remember either 2 or 4 sequentially presented stimuli over a brief WM delay. This paradigm allowed for the investigation of load effects separately during encoding, maintenance, and retrieval.

Materials and Methods

Subjects

Eighteen healthy young individuals (age: 22 ± 5 years; 8 males) from the Boston University community participated in this fMRI study after providing informed consent in a manner approved by both the Partners Human Research Committee of the Massachusetts General Hospital and the Boston University Charles River Campus Institutional Review Board. Subjects were included if they did not have any history of or current neurological or psychiatric symptoms and did not have any conditions that are counterindicators for magnetic resonance imaging. Vision was normal or corrected to normal. All subjects were right handed.

Task Procedures

Stimuli

As in our previous work (Sherman et al. 2003; Schon et al. 2004, 2005), we used a set of 600 digital color photographs of unfamiliar trial-unique complex visual outdoor scenes as stimuli.
Sternberg Task

Subjects performed a Sternberg task (Sternberg 1966) while in the fMRI scanner (Fig. 1). There were 2 conditions, LOAD2 and LOAD4. Each trial consisted of an initial encoding phase during which the subjects saw either 2 (LOAD2, 4 s) or 4 (LOAD4, 8 s) sequentially presented scenes. The encoding period was then followed by a variable-length delay period (4, 6, or 8 s; maintenance phase), which, in turn, was followed by a retrieval phase (2 s), and lastly, the retrieval phase was followed by a variable-length intertrial interval (8, 10, 12 s; fixation/ITI).

During the encoding phase, each scene was presented for an average of 1600 ms (ranging from 1400 to 1800 ms) and was followed by a variable temporal jitter during which the screen was black (mean, 400 ms; range, 200–600 ms; uniform distribution in steps of 100 ms). This temporal jitter during the encoding phase was included in order to allow use of the identical task in a future intracranial electroencephalography (iEEG) and magnetoencephalography study. As in previous studies (Sakai and Passingham 2003; Cairo et al. 2004; Ranganath et al. 2005; Piekema et al. 2006), the variable length of both the delay period and the ITI reduces multicollinearity between covariates (i.e., overlap between hemodynamic responses) that model temporally adjacent events (encoding, maintenance, and then retrieval) by introducing differential overlap. Adding variable-length delay and ITI periods thus allows assessment of the hemodynamic response separately for encoding, maintenance, and retrieval periods. In addition, we performed a separate behavioral pilot study with 15 healthy young subjects (age: 20 ± 2 years) that demonstrated no effect of delay length (4, 6, or 8 s) on accuracy or reaction time (RT) and no delay length by memory load interaction on these behavioral measures.

During the retrieval phase, subjects indicated with a button press response whether the probe picture was identical to one of the sample pictures seen during that trial (Match). Subjects were asked to respond as quickly and as accurately as possible. Matching stimuli at retrieval were equally likely to have been encountered before in any of the temporal positions during the encoding phase (positions 1 and 2 for LOAD2 trials and positions 1, 2, 3, and 4 for LOAD4 trials). The probability of Match trials for each condition was 0.5.

Task Procedure

One day before scanning, subjects viewed task instructions and practiced the task on a computer screen using a different set of unfamiliar, trial-unique scenes. The following day, subjects performed 8 runs of the task with 18 trials each in the scanner. There were 9 LOAD2 trials and 9 LOAD4 trials per run. Thus, each subject performed a total of 144 trials (72 trials per memory load). Scanning took approximately 60 min. We used PsyScope X Build 46 (http://psy.ck.sissa.it/) for task presentation and recording of RTs and accuracy on an Intel Macintosh MacBookPro laptop.

fMRI Data Acquisition

All imaging data were acquired on a 3-T MAGNETOM Trio scanner (Siemens AG, Medical Solutions, Erlangen, Germany) using a 12-channel Tim® Matrix coil at the Athinoula A. Martinos Center for Biomedical Imaging at the Massachusetts General Hospital in Charlestown, MA. We acquired 2 high-resolution T1-weighted magnetization-prepared rapid gradient echo structural scans (time repetition [TR], 2530 ms; time echo [TE], 3.39 ms; flip angle, 7°; 128 slices; matrix size, 256 × 192; field of view, 256; in-plane resolution, 1 mm²; slice thickness, 1.33 mm). Following the structural scans, we acquired 8 functional T2*-weighted blood oxygen level-dependent (BOLD) scans with 216 images each during which the subjects performed the task. Thirty slices were...
aligned in a coronal oblique fashion perpendicular to the long axis of the hippocampus. This slice prescription (TR, 2000 ms; TE, 30 ms; flip angle, 90°; matrix size, 64 × 64; field of view, 200 mm; in-plane resolution, 1.125 mm²; slice thickness, 5 mm; 1 mm gap between slices; interleaved slice acquisition) optimized signal in the MTL while at the same time covering the whole brain.

fMRI Data Preprocessing

fMRI data were preprocessed with SPM5 software (Friston et al. 1994, Friston, Holmes, Poline, et al. 1995; Friston, Holmes, Worsley, et al. 1999). Preprocessing included 1) Reorienting of all BOLD images such that the origin (i.e., coordinate \( \mathbf{x} = [0 0 0] \)) was at the anterior commissure. 2) Slice-timing correction of all BOLD images temporally smoothed all images to the first slice acquired in time. 3) Motion correction included realigning and unwarping the BOLD images in order to correct for variance due to susceptibility-by-movement interactions (Andersson et al. 2001). This step created a mean BOLD image and 6 movement parameters (3 translations and 3 rotations), the latter of which were later entered into the general linear model as covariates of no interest. In step 4), the structural images were coregistered to the mean BOLD image for later visualization of functional activity. In the following step 5), the high-resolution structural images were segmented into white and gray matter images. A bias-corrected structural image was also created at this step using the default tissue probability maps as priors. SPM5 uses a modified version of the ICBM (International Consortium for Brain Mapping) Tissue Probability Atlas, and images are registered to the Montreal Neurological Institute (MNI) space. Bias correction of smooth image intensity variations may allow more accurate spatial registration (Ashburner and Friston 2005). 6) For later signal averaging across subjects, the bias-corrected structural images and the coregistered BOLD images were spatially normalized into standard ICBM/MNI space using parameters derived during segmentation of the structural images and using nonlinear image registration (including resampling to 1 mm³ isotropic voxels for structural images and 3 mm³ isotropic voxels for BOLD images). The normalized structural images of all subjects were averaged, and the average was used for statistical overlay of the statistical parametric maps (SPMs) in Figures 2 and 3. 7) Finally, the BOLD images were spatially smoothed using a 6-mm full-width at half-maximum Gaussian kernel.

Data Analysis

Behavioral Data Analysis

We calculated median RTs for each subject and each condition and then averaged them across subjects to obtain the mean. Averaged RTs and accuracy (proportion of correct responses) were analyzed with a 2-factor, repeated-measures analysis of variance (ANOVA) with the factors load (LOAD2 vs. LOAD4) and trial (Match vs. Nonmatch) in order to demonstrate that a load effect was modulated by the 2 behavioral measures and that this load effect was modulated by trial type (Match vs. Nonmatch). All behavioral analyses use an alpha level of 0.05.

In addition, we performed a separate analysis in order to demonstrate absence of an effect of delay length (4, 6, or 8 s) on accuracy or RT and absence of a delay length by memory load interaction on these behavioral measures.

fMRI Data Analysis

Sixteen separate regressors were created for each subject as a function of load (LOAD2 vs. LOAD4), event (encoding vs. maintenance vs. retrieval vs. fixation/ITI), and accuracy (correct vs. incorrect) and convolved with the canonical hemodynamic response function in SPM5. Thus, all conditions were explicitly modeled as a regressor, including fixation/ITI. The 6 movement parameters from the realignment procedure were additionally added as covariates of no interest to account for residual movement-related spurious activation. The design matrix was constructed separately for each subject and analyzed using the modified general linear model approach in SPM5. \( T \) contrasts assessing WM load effects compared LOAD4 versus LOAD2 separately for each event of a trial (encoding, maintenance, and retrieval vs. fixation/ITI) optimized signal in the MTL while at the same time covering the whole brain.

Figure 2. Load-dependent activations during encoding and retrieval (LOAD4 > LOAD2). SPMs are displayed on anatomical images derived from an average obtained from the normalized structural images of all subjects using a statistical threshold of \( p_{FDR} <0.05 \) with SVC for ROIs outside PFC. Time courses display percent signal change with respect to the overall ROI mean and are separated into 3 plots: 1) Peri-sample onset time courses (left), 2) Postsample time courses (center), and 3) Peri-probe onset time courses (right) for Nonmatch trials. Peri-sample onset time courses illustrate activity during encoding with zero corresponding to the onset of the first sample scene during the encoding phase, postsample time courses illustrate activity during the maintenance phase with zero corresponding to the onset of the last sample scene during the encoding phase (i.e., scene #2 for LOAD2 trials and scene #4 for LOAD4 trials), and peri-probe onset time courses illustrate activity during retrieval with zero corresponding to the onset of the probe during the retrieval period. Blue lines indicate activity during LOAD4 trials, and red lines indicate activity during LOAD2 trials. (A) WM load effect during encoding in the right and left anterior VLPFC and the right DLPFC. (B) WM load effect during retrieval in the left anterior hippocampal formation, in the left and right retrosplenial area, and in the left amygdala. R, right; L, left.
(retrieval), and $t$ contrasts assessing load-unrelated effects compared encoding, maintenance, and retrieval versus fixation/ITI. Only correct trials were included for these comparisons. In a second step, we created group-averaged SPMs by entering the resulting contrast images into 1-sample $t$-tests using subject as a random factor. The group SPMs were thresholded at $p < 0.00005$ with whole-brain correction for delay > fixation/ITI and $p_{FDR} < 0.05$ with whole-brain correction for encoding > fixation/ITI and for retrieval > fixation/ITI and $p_{FDR} < 0.05$ with whole-brain correction for delay > fixation/ITI. Time courses display percent signal change with respect to the overall ROI mean and are separated into 3 plots: 1) Peri-sample onset time courses (left), 2) Postsample time courses (center), and 3) Peri-probe onset time courses (right). Peri-sample onset time courses illustrate activity during encoding with zero corresponding to the onset of the first sample scene during the encoding phase, postsample time courses illustrate activity during the maintenance phase with zero corresponding to the onset of the last sample scene during the encoding phase (i.e., scene #2 for LOAD2 trials and scene #4 for LOAD4 trials), and peri-probe onset time courses illustrate activity during retrieval with zero corresponding to the onset of the probe during the retrieval period. Blue lines indicate activity during LOAD4 trials, and red lines indicate activity during LOAD2 trials.

**Figure 3.** Load-independent activations during encoding, maintenance (delay), and retrieval. SPMs are displayed on anatomical images derived from an average obtained from the normalized structural images of all subjects using a statistical threshold of $p_{FDR} < 0.00005$ with whole-brain correction for encoding > fixation/ITI and for retrieval > fixation/ITI and $p_{FDR} < 0.05$ with whole-brain correction for delay > fixation/ITI. Time courses display percent signal change with respect to the overall ROI mean and are separated into 3 plots: 1) Peri-sample onset time courses (left), 2) Postsample time courses (center), and 3) Peri-probe onset time courses (right). Peri-sample onset time courses illustrate activity during encoding with zero corresponding to the onset of the first sample scene during the encoding phase, postsample time courses illustrate activity during the maintenance phase with zero corresponding to the onset of the last sample scene during the encoding phase (i.e., scene #2 for LOAD2 trials and scene #4 for LOAD4 trials), and peri-probe onset time courses illustrate activity during retrieval with zero corresponding to the onset of the probe during the retrieval period. Blue lines indicate activity during LOAD4 trials, and red lines indicate activity during LOAD2 trials.

(A) Activity during encoding versus fixation/ITI in the right hippocampal body/tail, the right fusiform gyrus, and the right VLPFC. (B) Activity during maintenance (delay) versus fixation/ITI in the right anterior DLPFC. (C) Activity during retrieval versus fixation/ITI in the right hippocampal body/tail, the right fusiform gyrus, and the right anterior DLPFC. (D) Activity during retrieval that is greater than activity during encoding (retrieval > encoding) in the right hippocampal body/tail and in the right anterior DLPFC. R, right; L, left.
cases, the threshold extent was 5 voxels. A statistical threshold of $p_{\text{vox}} < 0.05$ was employed whenever activity was expected to be less robust (i.e., when assessing delay-period activity and when assessing differential activity as a function of WM load), whereas a statistical threshold of $p_{\text{vox}} < 0.00005$ was used whenever activity was expected to be very robust, such as encoding or retrieval versus ITI/fixation. The FDR measure (Genovese et al. 2002) controls the expected "proportion" of false-positive voxels among all suprathreshold voxels; thus, a more stringent threshold was used whenever a very large number of activated voxels was expected (Nichols and Hayasaka 2003). SVC was restricted to an a priori-defined ROI as described below.

**ROI Selection**

SVC included a large a priori-defined mask generated using the WFU_PickAtlas tool (Maldjian et al. 2003) and included the hippocampus, parahippocampal gyrus, amygdala, fusiform gyrus, lingual gyrus, and posterior cingulate cortex, bilaterally. We performed SVC on regions within this mask because of the known susceptibility-induced signal reduction in the anterior MTL (Veitman et al. 2000; Greicius et al. 2003). ROIs were selected based on previous fMRI studies on WM, recognition memory, and long-term encoding with complex visual scenes (Stem et al. 1996, 2001; Epstein et al. 1999; Menon et al. 2000; Rombouts et al. 2001; Schon et al. 2004; Castelo et al. 2006). These studies have demonstrated a role for the MTL (including perirhinal/entorhinal cortex [Brodmann Area (BA) 28/34/35], posterior perihippocampal cortex, and hippocampus), the fusiform gyrus (BA 36/37), the lingual gyrus (BA 19/18), and the retrosperical cortex (BAs 30 and 29) in memory for complex visual scenes. We used this large mask for SVC for comparisons assessing WM load effects. Because previous studies have indicated a WM load effect in the lateral PFC (Rypma et al. 1999, 2002; Druzel and D’Esposito 2003; Veitman et al. 2003; Cairo et al. 2004), additional ROIs included the dorsolateral PFC (DLPFC) (BAs 9 and 46) and the ventrolateral PFC (VLPFC) (BAs 44, 45, and 47). MTL areas were anatomically localized using anatomical maps, including the probabilistic cytoarchitectonic maps by Amunts et al. (2005), provided in the SPM5 Anatomy toolbox (Eickhoff et al. 2005, 2006, 2007).

**Extraction of Signal Intensities and Percent Signal Change Calculation**

Raw signal intensities were extracted using the Volumes toolbox (http://sourceforge.net/projects/spmtools) extension for SPM5. Using the activation peaks from the group analyses within our ROIs, we extracted raw time series separately for each ROI and each subject from a sphere with a 5-mm radius and the ROI peak as the center coordinate. Percent signal change over time was calculated to explore temporal dynamics as follows: ([signal intensity during task − average signal intensity in ROI]/average signal intensity in ROI) × 100. Thus, the baseline used to calculate percent signal change for each ROI was the overall mean of all voxels that are included in the 5-mm sphere of that ROI (~15 voxels). Similar to previous papers reporting tasks with variable-length delay periods (Sakai and Passingham 2003; Ranganath et al. 2005), we constructed peri-sample onset, postsample, and peri-probe onset time courses (see Figs 2 and 3). In these line graphs, percent signal change is displayed on the y axis and time (in seconds, with 0 as onset) is displayed on the x axis. Peri-sample onset time courses illustrate activity during encoding with zero corresponding to the onset of the first sample scene during the encoding phase, postsample time courses illustrate activity during the maintenance phase with zero corresponding to the onset of the last sample scene during the encoding phase (i.e., scene 1/2 for LOAD2 trials and scene 3/4 for LOAD4 trials), and peri-probe onset time courses illustrate activity during retrieval with zero corresponding to the onset of the probe during the retrieval period. For each time point, averages and standard errors of the mean were calculated.

**Results**

Our results revealed a WM load effect in the left anterior MTL, including the hippocampal formation and the amygdala. Additional load effects were localized in a region that included both the retrosperical cortex and the anterior lingual gyrus, bilaterally; we have labeled this region the retrosperical area. This WM load effect was present during retrieval specifically during Nonmatch trials, that is, during the correct rejection of new scenes (lures). In addition, activity in the hippocampal body/tail was greater during retrieval than during encoding but was not modulated by WM load.

**Behavioral Results**

A 2-factor, repeated-measures ANOVA with the factors load (LOAD2 vs. LOAD4) and trial (Match vs. Nonmatch) on the dependent measures RTs (correct trials) and accuracy revealed a main effect of load on both RT ($F_{2,34} = 28.35$, $P < 0.05$) and accuracy ($F_{1,17} = 5.72$, $P < 0.05$), demonstrating that subjects were faster and made fewer errors on LOAD2 trials than on LOAD4 trials. Whereas the main effect of trial for RT was not significant ($F_{1,17} = 2.38$, NS), subjects made significantly fewer errors on Nonmatch trials (when lures needed to be rejected) than on Match trials (when the probe picture was identical to one of the sample pictures) as indicated by a significant main effect of trial for accuracy ($F_{1,17} = 1.07$, NS). Additional 2-factor repeated-measures ANOVAs investigated whether load effects (LOAD2 vs. LOAD4) were modulated by delay length (4 vs. 6 vs. 8 s) for the dependent measures RT (correct trials) and accuracy. These analyses revealed only a main effect of load for RT (LOAD4 > LOAD2; $F_{1,17} = 25.25$, $P < 0.05$) and for accuracy (LOAD2 > LOAD4; $F_{1,17} = 9.49$, $P < 0.05$). As expected, there was no significant main effect of delay length ($F_{2,34} = 0.30$, NS, and $F_{2,34} = 0.93$, NS, for RT and accuracy, respectively) and no significant load × delay length interaction for RT ($F_{2,34} = 1.40$, NS) or for accuracy ($F_{2,34} = 0.45$, NS).

**fMRI Results**

**Effects of WM Load during Retrieval**

WM load modulated activity in the left hippocampal formation, the left amygdala, and the retrosperical area, bilaterally. When activity during the retrieval phase of LOAD4 trials was compared with activity during the retrieval phase of LOAD2 trials (LOAD4 > LOAD2) across all trials (Match and Nonmatch), we observed greater MTL activity in the left anterior hippocampal formation (possibly subiculum) ($[x y z] = [−24 – 21 – 18]$; $Z = 3.92$) (Fig. 2B) and the left amygdala (possibly basolateral nucleus) ($[x y z] = [−24 – 3 – 12]$; $Z = 4.30$). Additional load effects during retrieval within our ROIs were present in a region that spanned both the retrosperical cortex and the anterior lingual gyrus, bilaterally. This activity was centered on the junction between the parietooccipital fissure and the anterior calcaneus sulcus (retrosperical area/anterior lingual gyrus) ($[x y z] = [−15 – 45 0]$; $Z = 4.41$; $[x y z] = [15 – 45 – 6]$; $Z = 3.40$; BA 30/19) (Fig. 2B) and is labeled retrosperical area in the figures and discussion. Peri-probe time courses illustrate that this load effect was present for Nonmatch trials when new scenes (lures) were correctly rejected (Fig. 2B). Consistent with this, when we repeated the SPM analysis separately for Match and Nonmatch trials, the load effect was present in the same areas for Nonmatch trials (i.e., correct rejections), but not for Match trials (i.e., hits) (not depicted). When activity during the probe period of Nonmatch trials was directly compared with activity during the probe period of Match trials regardless
of WM load (Nonmatch > Match), none of the above areas showed differential activation, indicating that activity in these regions does not reflect incidental long-term memory encoding of the novel, nonmatching stimuli. In addition, there was no statistically significant difference between Match and Nonmatch trials during retrieval as assessed by comparing ([LOAD4 Nonmatch > LOAD2 Nonmatch] > [LOAD4 Match > LOAD2 Match]), nor any activation differences for Match trials greater than Nonmatch trials regardless of WM load during the retrieval period. We did observe a load effect neither in the PFC nor in brain areas outside our ROIs during retrieval.

Effects of WM Load during Encoding
When activity during the encoding phase of LOAD4 trials was compared with activity during the encoding phase of LOAD2 trials (LOAD4 > LOAD2), there were no suprathreshold voxels in any of our ROIs outside the PFC using the ROI mask for SVC at a threshold of $p_{FWE} < 0.05$. Within the lateral PFC, using $p_{FWE}$ correction across the whole brain, we observed greater activity with higher load for this comparison in the right and left anterior VLPFC (inferior frontal gyrus [IFG], $[x, y, z] = [42 45 3]$; $Z = 4.54$; $[x, y, z] = [-39 48 -3]$; $Z = 3.98$; BA 10) and in the right DLPFC (middle frontal gyrus, $[x, y, z] = [148 27 36]$; $Z = 3.64$; BA 9) (Fig. 2A). The peri-sample onset and postsample time courses also demonstrate greater activity during the encoding phase of LOAD4 trials than during the encoding phase of LOAD2 trials in these VLPFC and DLPFC regions. Activity in other regions demonstrating a load effect during encoding included the right and left supramarginal gyrus, posterior occipital areas, the anterior cingulate, left and right insula, the midcingulate gyrus, and precuneus.

Effects of WM Load during Maintenance
When activity during the maintenance phase of LOAD4 trials was compared with activity during the maintenance phase of LOAD2 trials (LOAD4 > LOAD2), there was no differential activity in any of our ROIs, including the lateral PFC. In addition, there was no differential activity in any area outside of the ROIs for this comparison, demonstrating the absence of a WM load effect during the maintenance phase.

WM Load–Independent Effects in the MTL and PFC during Encoding
Activation during encoding recruited the hippocampal body/tail, independent of WM load. When encoding was compared with fixation/TTI, activity was present in the fusiform gyrus, bilaterally ($[x, y, z] = [27 -54 -9]$; $Z = 6.51$; $[x, y, z] = [-27 -63 -15]$; $Z = 6.20$; BA 37) and in the hippocampal body/tail, bilaterally ($[x, y, z] = [-21 -36 3]$; $Z = 6.15$; $[x, y, z] = [18 -33 -3]$; $Z = 5.90$) (Fig. 3A). Activity in these regions was not modulated by WM load as demonstrated by the peri-sample onset and postsample time courses (Fig. 3A). Within the lateral PFC, the right VLPFC (inferior frontal sulcus, $[x, y, z] = [45 15 27]$; $Z = 5.00$; BA 45) was recruited during encoding. During encoding, activity in other regions included posterior occipital areas, the left pre-SMA, the left and right intraparietal sulci, the left putamen, and the mediodorsal thalamic nucleus.

WM Load–Independent Effects in the MTL and PFC during Retrieval
When retrieval was compared with fixation/TTI, the observed activity pattern was similar to that seen during encoding versus fixation/TTI and included the right and left hippocampal body/tail ($[x, y, z] = [21 -30 -3]$; $Z = 6.13$; $[x, y, z] = [-21 -30 -9]$; $Z = 6.23$) and the right and left fusiform gyrus ($[x, y, z] = [24 -45 -15]$; $Z = 6.29$; $[x, y, z] = [-24 -45 -12]$; $Z = 6.00$; BA 37) (Fig. 3C). Activity in these regions was not modulated by WM load, as demonstrated by the postsample and peri-probe onset time courses (Fig. 3C). Within the PFC, the right and left anterior DLPFC ($[x, y, z] = [27 51 15]$; $Z = 5.22$; $[x, y, z] = [-30 48 21]$; $Z = 5.36$; BA 10) and the right and left VLPFC (inferior frontal sulcus, $[x, y, z] = [45 9 27]$; BA 45; $Z = 4.82$; $[x, y, z] = [-51 9 30]$; BA 45; $Z = 5.50$) were recruited during retrieval. During retrieval, activity in other regions included posterior occipital areas, the right posterior superior temporal gyrus, the midcingulate gyrus, the dorsal anterior cingulate cortex/pre-SMA, the insula, the right posterior superior temporal sulcus, the left frontal eye fields (BA 6), the left and the right mediodorsal thalamic nuclei, the left and the right intraparietal sulci, as well as the left and the right putamen.

A direct comparison between encoding and retrieval periods across WM loads revealed greater activity during retrieval compared with encoding in the right and left hippocampal body/tail ($[x, y, z] = [24 -27 -6]$; $Z = 5.34$; $[x, y, z] = [-21 -30 -9]$; $Z = 5.80$) and in the right and left anterior DLPFC ($[x, y, z] = [27 51 15]$; $Z = 5.09$; $[x, y, z] = [-33 48 18]$; $Z = 5.37$; BA 10) (Fig. 3D). There was no differential activity for the reverse contrast (encoding > retrieval).

WM Load–Independent Effects in the PFC, but not the MTL, during Maintenance
When maintenance was compared with fixation/TTI, there was no activity in any of the ROIs outside the PFC. Within the lateral PFC, the right and left anterior DLPFC ($[x, y, z] = [36 42 30]$; $Z = 4.52$; $[x, y, z] = [-33 45 30]$; $Z = 4.39$; BA 10) were active during the delay period when novel scenes needed to be maintained during the WM delay (Fig. 3B). Activity in these regions was not modulated by WM load as demonstrated by the postsample and peri-probe onset time courses (Fig. 3B). During maintenance, activity in other regions included the basal ganglia, anterior insula, posterior occipital regions, the left intraparietal sulcus, and the dorsal anterior cingulate/pre-SMA.

Discussion
WM load effects occurred in the left anterior MTL, including the hippocampal formation and the amygdala, and bilaterally in the retrosplenial area. This WM load effect was present only during retrieval, specifically for Nonmatch trials when lures were correctly rejected, but not when old stimuli were correctly recognized as a Match (i.e., hit).

Our results reveal additional evidence for rejecting the view that WM and long-term memory are 2 separate entities (for review, see Ranganath and Blumenfeld 2005; Jonides et al. 2008). Previous work has provided evidence that the same PFC regions are recruited during WM and during long-term memory. This has been demonstrated for both encoding (Ranganath et al. 2003) and retrieval (Cabeza et al. 2002; Ranganath et al. 2003). Most importantly, recent work has attributed a critical role for the MTL in encoding and maintenance during WM (Ranganath and D’Esposito 2001; Stern et al. 2001; Schon et al. 2004; Karlsgodt et al. 2005; Hasselmo and Stern 2006; Nichols et al. 2006) in addition to episodic encoding and retrieval (e.g., Stern et al. 1996; Gabrieli
et al. 1997). Our findings demonstrate not only that MTL regions are also recruited during retrieval in the context of a WM task (Cabeza et al. 2002; Karlsdottir et al. 2005) but also that this region can be modulated by cognitive load during WM retrieval when items are correctly rejected in the case of Nonmatch trials.

These results match predictions from computational models (Lisman and Idiart 1995; Jensen et al. 1996; Jensen and Lisman 2005; Koene and Hasselmo 2007) based on cellular mechanisms of persistent spiking in parahippocampal regions (Klink and Alonso 1997; Egorov et al. 2002). According to these models, multiple items can be held in a WM buffer in a sequential order during the WM delay. These models predict load-related MTL activity during the delay period of the Sternberg task, but it is equally likely that the load effect would be observed during retrieval because subjects would replay the items held in the WM buffer sequentially until a matching item is found (Sternberg 1966; Lisman and Idiart 1995; Jensen and Lisman 1998). Replay during retrieval could also explain why the load effect was present only for Nonmatch trials when lures were correctly rejected, but not for Match trials (i.e., hits). In the case of Nonmatch trials, subjects have to sequentially replay all 4 items during LOAD4 trials, but only 2 items during LOAD2 trials. On Match trials, replay would end when the correct match is encountered (Jensen and Lisman 1998). Although this replay idea is appealing as a possible interpretation for our findings, parallel search process models have sometimes been favored as they fit behavioral RT data better than serial scanning process models (Jonides et al. 2008), and our behavioral RT data also do not support serial scanning, possibly because of the small number of items viewed during the encoding phase. The load effect is also consistent with other models of WM that buffer multiple items in parallel but use different magnitudes of activity rather than different spiking phases (Grossberg 1978; Bullock 2004). It is unlikely that this load effect can be explained by WM rehearsal processes because the stimuli used were novel complex visual outdoor scenes, which are difficult to label verbally.

Our results are consistent with the idea that the hippocampus may act as a match/mismatch detector, as put forth in a series of recent papers by Kumaran and Maguire (2006, 2007a, 2007b). In their fMRI studies, the left hippocampus was recruited whenever a mismatch was detected between a stored representation, such as a sequence of items, and perceptual input. In our Sternberg task, it is possible that the observed hippocampal activation during retrieval might have been due to mismatch detection during Nonmatch trials when the scene encountered during the probe period did not match any of the scenes encountered during the sample period. This mismatch (i.e., Nonmatch) condition was also the condition that demonstrated a load effect (LOAD4 > LOAD2) in a similar left hippocampal region in our study. Although our paradigm was not designed to assess the role of the hippocampus as a match/mismatch detector, this comparator role of the hippocampus may be more general than originally proposed (Eichenbaum and Buckingham 1990; Hasselmo and Wybly 1997; Vinogradova 2001; Lisman and Grace 2005; Kumaran and Maguire 2006, 2007a, 2007b). According to these models, the hippocampus generates a mismatch signal during novelty detection when prior expectations are violated. Our behavioral task should not have produced prior expectations because 50% of all trials were Nonmatch trials (i.e., they occurred by chance). Therefore, we argue that the hippocampus may act as a mismatch detector whenever the perceptual reality does not match the stored representation, regardless of prior expectations. Whereas this contrasts with the conclusion of Kumaran and Maguire that the hippocampus is not critical for associative novelty per se, the idea that the hippocampus may be a general comparator is consistent with our observed load effect in this region because it should take longer to compare the stored representation of larger stimulus sets (i.e., LOAD4 trials) with the perceptual input than to compare the stored representation of smaller stimulus sets (i.e., LOAD2 trials) with the perceptual reality.

The WM load effect observed in the left anterior hippocampal formation (including subicum) during retrieval for Nonmatch trials when lures were correctly rejected is consistent with a high-resolution fMRI study demonstrating greater left subicum activity for subsequent correctly rejected lures (i.e., Nonmatch) than for hits (i.e., Match) (Kirwan and Stark 2007). Thus, our data demonstrate that areas that have previously been implicated in episodic retrieval are also recruited during WM retrieval. Anatomical tracer studies have demonstrated strong reciprocal connections between the subicum and both the retrosplenial cortex (Morris et al. 1999; Kobayashi and Amaral 2003, 2007) and the lateral PFC (Barbas and Blatt 1995).

The WM load effect was clearly evident on the behavioral level as well. Subjects made significantly more errors and were slower on LOAD4 trials than on LOAD2 trials. Whereas this behavioral load effect and its associated MTL activity could alternatively be interpreted as being related to cognitive effort, the fact that the MTL activity was greater for Nonmatch trials on which subjects also made fewer errors than on Match trials discredits this alternative explanation.

An iEEG study by Axmacher et al. (2007) recently demonstrated a WM load effect in the rhinal cortex using a Sternberg task paradigm. In their study, gamma power increased with increasing WM load. In supplemental fMRI data, they found increased activity with greater WM load in the same region within the left anterior hippocampal formation that we report here. However, in contrast to our findings, they observed their fMRI load effect during maintenance, specifically during the late delay, but not during retrieval. The results of the 2 fMRI studies may actually be consistent with each other, as the load effect reported by Axmacher et al. (2007) during the late delay period might have been associated with activity at retrieval due to collinearity between regressors assessing activity during the late delay and activity during retrieval in their supplemental fMRI study. Our results move beyond this study by demonstrating that the WM load effect in the anterior hippocampal formation is present only during Nonmatch trials when lures were correctly rejected, but not during Match trials when old scenes were correctly recognized as old.

A recent functional connectivity study, using a Sternberg task with unfamiliar faces (Rissman et al. 2008), has demonstrated that the correlation between the IFG and the hippocampus increases with increasing WM load during the maintenance delay. However, inconsistent with our results and with the data presented by Axmacher et al. (2007), only the IFG, but not the hippocampus, increased linearly with increasing WM load using standard univariate analysis methods. This discrepancy is likely attributable to that study’s reduced power as indicated by smaller number of trials and subjects.
The fact that we did not observe a WM load effect during the delay period in the PFC may be due to the stimulus material used. Our data suggest that this area is not recruited if stimuli are unfamiliar and trial unique (Stern et al. 2001).

Consistent with previous WM studies, we observed a WM load effect in the PFC. Specifically, we observed a load effect in the anterior VLPFC, bilaterally, and in the right DLPFC. Unlike previous reports (Rypma et al. 1999, 2002; Druzel and D’Esposito 2003; Veltman et al. 2003; Cairo et al. 2004), we observed this effect during encoding, but not during maintenance. Whereas we observed load-independent activity in the anterior DLPFC during the delay period, in contrast to these other studies, this area was not modulated by WM load in our study. It is possible that PFC activity is increased with increasing WM load during maintenance when the memoranda are highly familiar, as in most previous studies (Rypma et al. 1999, 2002; Veltman et al. 2003; Cairo et al. 2004), but not when they are unfamiliar and trial unique as in our study. This idea is consistent with the suggestion that the PFC is needed when task monitoring requirements are high (Owen et al. 1996; Owen 2000; Petrides 2000; Petrides et al. 2002) and is consistent with the putative role of the DLPFC in executive control (Postle et al. 1999; Smith and Jonides 1999; Menon et al. 2001; Wagner et al. 2001; Lie et al. 2006). Previous work in our laboratory has demonstrated greater PFC activity for 2-back task performance with familiar stimuli than with novel stimuli and MTL activity for 2-back task performance if stimuli were novel, but not if stimuli were familiar (Stern et al. 2001). Similarly, using a delayed matching-to-sample task, we have recently demonstrated that delay-period activity in the PFC is also modulated by whether the stimuli were unfamiliar (trial unique) or preexposed (Schon et al. 2008). Thus, the MTL and the PFC may be differentially recruited for WM based on whether the stimuli are unfamiliar and trial unique or highly familiar (Hasselmo and Stern 2006).

The retrosplenial cortex has not been implicated previously in WM studies. Most fMRI studies on WM that used novel stimuli have focused on the hippocampus and related regions within the MTL (Ranganath and D’Esposito 2001; Stern et al. 2001; Schon et al. 2004; Nichols et al. 2006) and not on functionally related or anatomically connected brain areas such as the retrosplenial area. In addition, studies that specifically set out to investigate brain areas involved in WM retrieval are sparse (Oztekin et al. 2008) and have not included the retrosplenial area. This study demonstrates that brain regions that have previously been implicated in episodic retrieval are also recruited for WM retrieval, including the MTL and the retrosplenial cortex. Studies on scene recognition (Suzuki et al. 2005; Epstein et al. 2007) and autobiographical memory retrieval (Gillova et al. 2004; Steinworth et al. 2006) point to the role of the retrosplenial cortex in episodic retrieval (see also Wiggs et al. 1999), a mechanism that, as we demonstrate here, may be modulated by the number of items held in a WM buffer. Consistent with our findings, anatomical investigations using tracer methods have found strong bidirectional anatomical connections between the retrosplenial cortex and the hippocampal formation, including the subiculum, and the entorhinal cortex (Morris et al. 1999; Kobayashi and Amaral 2003, 2007).

Episodic retrieval of autobiographical memories has also been shown to recruit the amygdala when the retrieved content is emotional (Dolcos et al. 2005; Cabeza and St Jacques 2007; Daselaar et al. 2008). Similarly, the left amygdala has been shown to be recruited during retrieval of emotional information (Maratos et al. 2001; Smith et al. 2004, 2005; Sergerie et al. 2006). The amygdala has also been linked to recollection during episodic retrieval (Dolcos et al. 2005; Fenker et al. 2005). The outdoor scenes used in our study were not emotionally salient, suggesting that the amygdala may play a general role in memory retrieval. Consistent with this idea, the amygdala has been implicated in attentional modulation of memory (Gallagher and Holland 1994; Gallagher and Chiba 1996; Holland and Gallagher 1999; Holland et al. 2000), associative learning (Hatfield et al. 1996; Holland and Gallagher 1999), and memory consolidation (Cahill and McGaugh 1998; Malin and McGaugh 2006). The basolateral nucleus of the amygdala has direct connections with the hippocampal formation (Saunders et al. 1988; Ishikawa and Nakamura 2006) and with the retrosplenial cortex in rats and monkeys (Buckwalter et al. 2008), providing a consistent neuroanatomical framework for our results. In addition, consistent with our observation that a WM load effect during retrieval was absent in the lateral PFC, direct connections between the amygdala and lateral PFC have been shown to be sparse (Ghashghaei and Barbas 2002).

In summary, the current fMRI study demonstrates that interconnected brain structures that have previously been implicated in episodic retrieval, including the hippocampal formation, the amygdala, and the retrosplenial area, are also recruited to retrieve complex visual scenes during WM. Retrieval-related activity in these areas was modulated by WM load during Nonmatch trials, which required the correct rejection of lures. This finding is consistent with predictions derived from recent computational modeling work that suggest a WM load effect in the MTL, an effect that has previously been attributed mainly to the PFC. This work extends recent WM studies in demonstrating a role for the MTLs not only for long-term encoding but also for episodic retrieval during WM and supports models of WM predicting that activity in the MTL should be modulated by WM load.

Funding
National Science Foundation grant (SBE-0354278); National Institutes of Health grants (P50 MH071702-01, National Center for Research Resources P41RR14075); Mental Illness and Neuroscience Discovery Institute.

Notes
We would like to thank Dr Mike Kahana for discussions about the experimental design. The results described in the manuscript were previously presented as a poster at the Annual Meeting of the Society for Neuroscience in San Diego, CA, 2007. Conflict of Interest: None declared.

Address correspondence to Karin Schon, Center for Memory and Brain, Boston University, 2 Cummingston Street, Suite 109, Boston, MA 02215, USA. Email: kschon@bu.edu.

References


