Dissociation of the neural correlates of visual and auditory contextual encoding

Lauren J. Gottlieb [a,*], Melina R. Uncapher [b], Michael D. Rugg [a]

[a] Center for the Neurobiology of Learning and Memory, Department of Neurobiology and Behavior, University of California at Irvine, Irvine, CA 92697-3800, United States
[b] Stanford Memory Laboratory, Department of Psychology, Stanford University, Stanford, CA 94305-2130, United States

ARTICLE INFO

ABSTRACT

The present study contrasted the neural correlates of encoding item-context associations according to whether the contextual information was visual or auditory. Subjects ($N = 20$) underwent fMRI scanning while studying a series of visually presented pictures, each of which co-occurred with either a visually or an auditorily presented name. The task requirement was to judge whether the name corresponded to the presented object. In a subsequent memory test subjects judged whether test pictures were studied or unstudied and, for items judged as studied, indicated the presentation modality of the associated name. Dissociable cortical regions demonstrating increased activity for visual vs. auditory trials (and vice versa) were identified. A subset of these modality-selective regions also showed modality-selective subsequent source memory effects, that is, enhanced responses on trials associated with correct modality judgments relative to those for which modality or item memory later failed. These findings constitute direct evidence for the proposal that successful encoding of a contextual feature is associated with enhanced activity in the cortical regions engaged during the on-line processing of that feature. In addition, successful encoding of visual objects within auditory contexts was associated with more extensive engagement of the hippocampus and adjacent medial temporal cortex than was the encoding of such objects within visual contexts. This raises the possibility that the encoding of across-modality item-context associations places more demands on the hippocampus than does the encoding of within-modality associations.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Episodic memory – memory for unique events – depends on the ability to bind the different elements of an event, such as a study item and its spatiotemporal context, into a cohesive memory representation (Tulving, 1983). According to one widely held proposal (e.g. Norman & O'Reilly, 2003) mnemonic binding is achieved through the establishment in the hippocampus of a representation of the patterns of cortical activity elicited by an episode as it is processed in real-time.[1] Retrieval occurs when the hippocampal representation is reactivated by a retrieval cue, which in turn leads to the reinstatement of the pattern of cortical activity encoded by the representation (Alvarez & Squire, 1994; Norman & O'Reilly, 2003; Nyberg, Habib, McIntosh, & Tulving, 2000; Nyberg et al., 2001; Persson & Nyberg, 2000; Rolls, 2000; Shastri, 2002). Over the past few years functional neuroimaging studies, employing designs

of increasing sophistication, have provided strong support for the notion of retrieval-related 'cortical reinstatement' (e.g. Gottfried, Smith, Rugg, & Dolan, 2004; Johnson & Rugg, 2007; Kahn, Davachi, & Wagner, 2004; Khader, Burke, Bien, Ranganath, & Rosler, 2005; Vaidya, Zhao, Desmond, & Gabrieli, 2002; Wheeler & Buckner, 2003, 2004; Wheeler, Petersen, & Buckner, 2000; Wheeler et al., 2006; Woodruff, Johnson, Uncapher, & Rugg, 2005). Here, we focus on the related but less well-investigated question whether the neural correlates of successful episodic encoding are also content-sensitive, as would be expected in the context of the above theoretical framework (Rugg et al., 2008).

Starting with Brewer, Zhao, Desmond, Glover, and Gabrieli (1998) and Wagner et al. (1998), numerous studies have investigated the neural correlates of successful memory encoding with the 'subsequent memory procedure'. In this procedure, event-related functional magnetic resonance imaging (fMRI) is used to contrast the neural activity elicited by study items according to how the items are endorsed on a subsequent memory test (e.g. whether the items are correctly recognized or misclassified as 'new'). Studies employing this procedure have provided strong evidence in favor of the proposal outlined above that the hippocampus plays a key role in the formation of cohesive episodic memory representations. Thus, there are numerous reports that items associated with later successful recollection of contextual details of a study episode elicit greater levels of study activity in the hippocampus and sur-

---

* Corresponding author.
  E-mail address: gottlieb@uci.edu (L.J. Gottlieb).

[1] In an extension of these proposals (Rugg, Johnson, Park, & Uncapher, 2008; Uncapher, Otten, & Rugg, 2006; Uncapher & Rugg, 2009; see also Moscovitch, 1992), the hippocampus 'captures' only the most behaviorally relevant or salient aspects of the activity elicited by a study event within different cortical regions, rather than representing the entirety of the activity.

rounding medial temporal lobe (MTL) than do items for which such details cannot be recollected on a later memory test (e.g. Davachi, Mitchell, & Wagner, 2003; Kensinger & Schacter, 2006; Ranganath et al., 2004; Staresina & Davachi, 2006; Staresina & Davachi, 2008; Uncapher & Rugg, 2009; for reviews see Davachi, 2006; Diana, Yonelinas, & Ranganath, 2007; see Gold et al., 2006 and Kirwan, Wixted, & Squire, 2008 for contradictory findings).

More germane here are findings that support the proposal that episodic memories are encoded in terms of the processes engaged as events are experienced in real-time. For example, several studies have reported that the cortical loci of subsequent memory effects differ according to the nature of the study task, and that the loci of these task-selective effects are found in regions selectively engaged by the task itself (Mitchell, Macrae, & Banaji, 2004; Otten & Rugg, 2001; Otten, Henson, & Rugg, 2002; Park, Uncapher, & Rugg, 2008; Rugg, Otten, & Henson, 2002). Additionally, and of more relevance to the present study, it has also been reported that the localization of the cortical activity supporting the encoding of a contextual feature differs according to the nature of the feature (Uncapher et al., 2006; Uncapher & Rugg, 2009). In the study of Uncapher et al. (2006), two orthogonally varying contextual features (font color and location) were associated with each study item. Subsequent memory effects predicting later memory for item-color associations were localized to a different cortical region than the regions where activity predicted memory for item-location associations. Moreover, the regions demonstrating these color- and location-selective encoding effects overlapped regions implicated by prior studies in the processing of color and location information respectively. This finding is consistent with the proposal that cortical subsequent memory effects reflect modulation of activity supporting the on-line processing of the study episode (see above). Uncapher et al. (2006) proposed that feature-selective subsequent memory effects are a consequence of trial-wise fluctuations in the allocation of attentional resources to the feature. They argued that as the allocation of attentional resources to a particular feature increases, so does activity in the cortical regions engaged by the feature. This enhanced activity strengthens the feature's on-line representation, and increases the probability that the associated pattern of cortical activity will be bound into a (hippocampally mediated) episodic memory representation. This proposal receives further support from a subsequent study that explicitly manipulated attention to color and location, and which reported dissociable subsequent memory effects for the two features that were selectively enhanced by feature-specific attention (Uncapher & Rugg, 2009).

In the present experiment, we further address the question whether the encoding of different contextual features is associated with dissociable subsequent memory effects. The principal aim of the experiment was to obtain direct evidence that feature-selective subsequent memory effects are localized to the same cortical regions that are selectively engaged during the processing of the feature. Such evidence would both replicate and significantly extend the findings of Uncapher et al. (2006) and Uncapher and Rugg (2009). In both of those studies, the conclusion that feature-selective subsequent memory effects overlapped cortical regions engaged during the processing of color and location information depended on findings obtained from prior studies (i.e. on 'reverse inference'; Poldrack, 2006) rather than on data acquired in the course of the experiment. By contrast, the design of the present experiment permitted a direct assessment of the extent of the overlap between subsequent memory effects predictive of later memory for a contextual feature, and the cortical regions selectively activated by the processing of the feature. This was accomplished by employing sensory modality as a contextual feature. Subjects studied a series of pictures that were accompanied by a word that was presented either visually or auditorily. At

test, they were required to identify each studied picture and to recollect the modality of the associated word. We expected that subsequent memory effects predictive of successful memory for the visual modality would be found in the regions where activity was greater on visual vs. auditory study trials, whereas auditory subsequent memory effects would overlap auditorily selective cortical regions. We further predicted that, regardless of modality, successful context memory would be associated with enhanced activity in hippocampus and adjacent MTL cortex (as was found by Uncapher & Rugg, 2009, albeit for different visual features rather than sensory modalities). In addition, although the dearth of prior subsequent memory studies employing auditory stimuli (though see Peters, Suchan, Köster, & Daum, 2007; Poppenk, Walia, McIntosh, Joanisse, & Kohler, 2008) precludes formulation of a specific prediction, the present study affords the opportunity to address the question whether MTL encoding-related activity is modality-sensitive.

2. Materials and methods

2.1. Subjects

Twenty-two subjects consented to participate in the study. All subjects reported themselves to be in good general health, right-handed, with no history of neurological disease or other contra-indications for MR imaging, and to have learned English as their first language. They were recruited from the University of California, Irvine (UCI) community and remunerated for their participation, in accordance with the human subjects procedures approved by the Institutional Review Board of UCI. The data from two subjects were excluded from all analyses because they had fewer than twelve trials in the 'miss' experimental condition (Ns of 7 and 8 respectively). Data are reported from the remaining 20 subjects (nine females) ranging in age from 19 years to 29 years (mean = 20 years).

2.2. Stimulus materials

Three hundred and six stimulus triplets were used in the experiment. Each triplet consisted of a picture of a commonly encountered object and both a visually and an auditorily presented word naming the object represented in the picture. The colored pictures were drawn from Hemera Photo Objects 50,000 Volume III (http://www.hemera.com/index.html). The names of the pictures were between 3 and 10 letters long, with a mean written frequency between 1 and 100 counts per million (Kucera & Francis, 1967). Visually presented words were displayed in black uppercase 30 point Helvetica font on a gray background. Auditory words were recorded by a male voice in the laboratory, edited to a constant sound pressure level and filtered to remove ambient noise (http://audacity.sourcefourge.net). Auditory stimuli were presented binaurally via MR compatible headphones and did not exceed 1000 ms in duration (mean duration 650 ms). Presentation volume was adjusted in the scanner to a comfortable listening level for each volunteer prior to scanning.

Of the 306 item triplets, 10 served as buffers (two at the beginning and end of each study list, and two at the beginning of the test list), and 16 additional triplets were used in the practice phases preceding the study and test sessions (see below). For every subject, 160 pictures and their corresponding names were randomly selected from the remaining 280 stimulus triplets to serve as the critical stimuli in the study session. These stimuli were further randomly divided into two groups: 80 pictures associated with their corresponding visual name and 80 pictures associated with their corresponding auditory name (congruent trials).

Twenty additional pictures were randomly selected for the study phase. These pictures were not paired with their names; instead, names from unused pictures in the pool were selected, such that for these 20 pictures the subsequently presented name did not correspond to the object (incongruent trials). Out of these 20 pictures, 10 of the mismatching names were presented visually, while the other 10 were presented auditorily.

Two study lists were created from the 180 study pictures (160 pictures with corresponding names, 20 pictures with mismatching names) for each subject. Each list contained a pseudo-random ordering of 40 pictures associated with matching visual words, 40 pictures associated with matching auditory words, five pictures associated with mismatching visual words, and five pictures associated with mismatching auditory words. The study task required a decision as to whether or not the words corresponded to the pictured objects.

Test items consisted of the 160 pictures from the congruent study trials and 80 new pictures. The test requirement was to judge whether the item had been presented at study and, if so, to indicate whether the associated word had been presented visually or auditorily. Both study and test items were presented in a subject-unique pseudo-random order, such that there were no more than three consecutive presentations of items belonging to any one experimental condition.

Study items were back-projected onto a screen and viewed via a mirror mounted on the scanner headcoil. Pictures were presented in central vision within a continuously displayed solid gray frame, and subtended maximum visual angles of 5.7° × 5.7°. Words were centered 1° below the pictures and subtended a maximum horizontal visual angle of 5.7° and a maximum vertical visual angle of 1.15°. Pictures and words displayed together subtended a maximum visual angle of 5.7° × 8° (width by height).

Test items were presented outside of the scanner on a computer screen. The pictures and associated cues were presented in central vision within a solid gray frame (subtending 6.8° × 6.8° visual angle at the 1 m viewing distance) that was continuously presented.

2.3. Experimental tasks and procedures

The experiment comprised a single study-test cycle.

2.3.1. Study procedure

Instructions and practice were administered outside the scanner. The study phase of the experiment proper consisted of the presentation of two blocks of items, separated by a brief rest period (approximately 1 min). Each study trial began with the presentation of a red fixation character in the center of the display frame for 500 ms. This was replaced by a picture that was presented for 1500 ms. Five hundred ms after the onset of the picture, a second item was presented concurrently. This was either a visual word (visual condition) or an auditory word (auditory condition). Following picture offset, a centrally presented black fixation character was displayed for a further 1500 ms, completing the trial. Subjects were informed they would receive no warning as to the modality of the word on each trial. Congruency judgments were signaled by button press of left or right index fingers, and response mapping was counterbalanced across subjects. Instructions placed equal emphasis on speed and accuracy.

The stimulus onset asynchrony (SOA) of study trials was stochastically distributed with a minimum of 3500 ms modulated by the addition of 40 randomly intermixed null trials (Josephs & Henson, 1999). Trials were presented in pseudo-random order, with no more than three trials of one item-type (visual condition, auditory condition, or null) occurring consecutively. Each block consisted of 94 trials, comprising 80 critical study items (congruent trials), 10 noncritical items (incongruent trials), and four buffer items for a total of 160 critical study items across two blocks.

2.3.2. Test procedure

Following the completion of the second study block, volunteers were removed from the scanner and taken to a neighboring testing room. Only then were they informed of the source memory test and given instructions and a short practice test. Approximately 30 min elapsed between the completion of the second study block and the beginning of the memory test. Each test trial began with a red fixation character presented in the center of a gray frame for 500 ms, followed by the presentation for 500 ms of a centrally presented picture within a solid gray frame.

The test items consisted of the 160 critical study items (i.e. from the congruent trials) and 80 randomly interspersed unstudied (new) pictures (no more than three items of one type were presented consecutively). Instructions were to judge whether each picture was old or new, and to indicate the decision with the right index (old) or left index (new) finger. If uncertain whether an item was old or new, volunteers were instructed to indicate 'new' so as to maximize the likelihood that subsequent source memory judgments (see below) would be confined to confidently recognized items. If a picture was judged 'new,' the test advanced to the next trial (with a 1 s inter-trial interval, during which a black fixation character was presented). If the picture was judged 'old' the prompt "Heard, Seen, Unsure?" appeared in black uppercase letters. Subjects were required to recall the modality of the name associated with the picture at study. The modality judgment was signaled by one of two button presses: right index finger for 'heard', and right middle finger for 'seen'. Subjects were further instructed that if they were unable to retrieve the modality they should respond 'unsure' with their right ring finger. The test was self-paced, with instructions to complete the test as quickly as possible without sacrificing speed for accuracy. The test was presented as a single block, lasting approximately 20 min.

2.4. fMRI data acquisition

A Philips Achieva 3T MR scanner (Philips Medical System, Andover, MA, USA) was used to acquire both T_1-weighted anatomical volume images data (240 × 240 matrix, 1 mm^3 voxels, 160 slices, sagittal acquisition, 3D MP-RAGE sequence) and T_2^*-weighted echoplanar images (EPI) [80 × 79 matrix, 3 mm × 3 mm in-plane resolution, axial acquisition, flip angle 70°, echo time (TE) 30 ms] with blood-oxygenation level dependent (BOLD) contrast. The data were acquired using a sensitivity encoding (SENSE) reduction factor of 1.5 on an eight-channel parallel imaging headcoil. Each EPI volume comprised 30 3 mm-thick axial slices separated by 1 mm, oriented parallel to the AC-PC plane, and positioned to give full coverage of the cerebrum and most of the cerebellum. Data were acquired in two sessions of 260 volumes each, with a repetition time (TR) of 2 s/vol. Volumes within sessions were acquired continuously in an ascending sequential order. The first five volumes of each session were discarded to allow equilibration of tissue magnetization.

2.5. fMRI data analysis

Data were analyzed with Statistical Parametric Mapping (SPM5, Wellcome Department of Cognitive Neurology, London, UK; Friston et al., 1995) implemented under Matlab2007a (The Mathworks Inc., USA). Functional images were subjected to a two-pass spatial realignment. Images were realigned to the first image, generating a mean image of the sessions. In the second pass the raw images were realigned to the generated mean image. The images were then subjected to reorientation, spatial normalization to a standard EPI template (based on the Montreal Neurological Institute (MNI) reference brain; Cocosco, Kollokian, Kwan, & Evans, 1997) and smoothing with an 8 mm FWHM Gaussian kernel. Functional time series were concatenated across sessions.

Statistical analyses were performed on the study phase data in two stages of a mixed effects model. In the first stage, neural activity elicited by the study pictures was modeled by delta functions (impulse event) that coincided with the onset of each picture. The ensuing BOLD response was modeled by convolving the neural functions with a canonical hemodynamic response function (HRF) and its temporal and dispersion derivatives (Friston et al., 1998) to yield regressors in a General Linear Model (GLM) that modeled the BOLD response to each event-type.

For the reasons discussed in the results section, the principal analyses were confined to four events of interest: studied pictures that were later recognized and correctly endorsed as having been paired with a visual or auditory word (visual source hits and auditory source hits respectively), and studied pictures that, on the later memory tests, were associated either with inaccurate source judgments (incorrect or unsure) or were misclassified as new. A fifth category of trials comprised events of no interest, namely, incongruent trials, buffer trials, and trials associated with incorrect or omitted study responses. Six regressors modeling concatenated movement-related variance (three rigid-body translations and three rotations determined from the realignment stage) and session-specific constant terms modeling the mean over scans in each session were also entered into the design matrix.

For each voxel, the functional time series was highpass-filtered to 1/128 Hz and scaled within session to yield a grand mean of 100 across voxels and scans. Parameter estimates for events of interest were estimated using a General Linear Model. Non-sphericity of the error covariance was accommodated by an AR(1) model, in which the temporal autocorrelation was estimated by pooling over suprathreshold voxels (Friston et al., 2002). The parameters for each covariate and the hyperparameters governing the error covariance were estimated using Restricted Maximum Likelihood (ReML). Effects of interest were tested using linear contrasts of the parameter estimates. These contrasts were carried forward to a second stage in which subjects were treated as a random effect. Unless otherwise specified, only effects surviving an uncorrected threshold of $p < .001$ and including nine or more contiguous voxels were interpreted. The peak voxels of clusters exhibiting reliable effects are reported in MNI coordinates.

Regions of overlap between the outcomes of two contrasts were identified by inclusive masking of the relevant SPMs. When the two contrasts were independent, the statistical significance of the resulting SPM was computed using Fisher's method for estimating the conjoint significance of independent tests (Fisher, 1950; Lazar, Luna, Sweeney, & Eddy, 2002). In all cases, the SPM to be masked was thresholded at $p < .01$ (again with a nine voxels extent threshold) and the SPM that constituted the mask was thresholded at $p < .001$, giving a conjoint significance level of $p < 10^{-4}$. Exclusive masking was used to identify voxels where effects were not shared between two contrasts. Contrasts to be masked were thresholded at $p < .001$ and the SPM constituting the exclusive mask was thresholded at $p < .05$ ($p < .1$ for bi-directional F contrasts). Note that the more liberal the threshold of an exclusive mask, the more conservative is the masking procedure.

3. Results

3.1. Behavioral performance

3.1.1. Study task

Accuracy of congruency judgments was high (.98 correct in both conditions). Mean reaction time (RT) for judgments on visual words was 1199 ms (SD = 226) compared with 1349 ms (SD = 186) for auditory words. Study RTs are shown in Table 1 segregated according to later memory performance. A 2 × 3 ANOVA [factors of modality and later memory (source hit, source miss, item miss)], revealed a main effect of modality ($F(1,19) = 105.63$, $p < .001$), but no effect of memory, and no memory by modality interaction. This analysis was repeated after collapsing RTs across the three classes of incorrect trial, analogous to the fMRI analyses reported below. A 2 × 2 ANOVA [factors of modality and memory (source hit vs. forgotten)] revealed only a main effect of modality ($F(1,19) = 115.39$, $p < .001$).

Table 1
Mean reaction times (ms) for congruency decisions at study segregated by subsequent memory and mean accuracy rates at test (SD in parentheses).

	Visual	Auditory
Reaction time		
Source hit	1257 (208)	1450 (204)
Source unsure	1257 (288)	1336 (199)
Source miss	1337 (229)	1417 (265)
Item miss	1207 (215)	1423 (200)
Accuracy		
Source hit	.65 (.13)	.70 (.16)
Source unsure	.16 (.11)	.17 (.12)
Source miss	.19 (.11)	.13 (.09)

3.1.2. Retrieval task

Hit rate collapsed over source accuracy was .79 (SD = .12) for items presented in the visual condition and .80 (SD = .10) for items from the auditory condition, against a false alarm rate of .03. Unsurprisingly, these two hit rates did not significantly differ. Source memory performance conditionalized on items receiving a correct recognition judgment is given in Table 1, where it can be seen that the proportions of correct source judgments for the two modalities are very similar and, as in the case of the hit rates, did not differ significantly ($t(19) < 1$). An overall measure of source recollection (Psr) was estimated using an index derived from a single high threshold model (Snodgrass & Corwin, 1988), in which the probability of recollection was computed as: $[p(\text{source hit}) - .5(1 - p(\text{source unsure}))]/[1 - (.5(1 - p(\text{source unsure})))]$, where 'Source Hit' refers to studied items that were recognized and assigned to their correct encoding context, and 'Source Unsure' refers to recognized items followed by an "unsure" modality response. The estimate of source recollection was .44 (SD = .13), which was significantly different from the chance value of 0 ($t(19) = 15.19$, $p < .001$).

3.2. fMRI results

The distribution of responses in the test phase meant that whereas all subjects contributed sufficient study trials to the source correct condition (means (ranges) of 45 (17–64) and 45 (22–68) for visual and auditory conditions respectively), the great majority of subjects had too few (<12) trials in one or more of the remaining trial types (source miss, source unsure, and item miss) to allow stable estimates of the activity elicited by these different trial types. Therefore these trials were collapsed to form a single category of 'source forgotten' trials containing all study items for which source-specifying information was unavailable. Consequently, the analyses reported below identify the neural correlates of successful visual and auditory source memory but do not speak to the question of the neural correlates of item recognition in the absence of source-specifying information. Mean (range) trial numbers for the forgotten trials were 38 (21–62) and 35 (12–58) for the visual and auditory conditions respectively.

We first identified regions demonstrating subsequent source memory effects common to the two modality conditions. We then performed hypothesis-driven analyses to assess whether, as predicted, modality-selective subsequent source memory effects (that is, subsequent source memory effects that were reliable for items studied in one of the two encoding conditions, but not the other) overlapped regions demonstrating generic modality effects (see Section 1). Lastly, we identified regions where subsequent source memory effects were selective for only one of the two modalities when the analyses were unconstrained by the generic effects of modality.

Table 2
Regions demonstrating modality-independent and auditorily selective subsequent memory effects.

Coordinates (x, y, z)	Z (#voxels)	Region	BA
Modality-independent			
−48 6 33	4.83 (84)	L inferior frontal gyrus	6/9
−21 −3 −21	3.72 (9)	L amygdala	
−45 −48 −24	3.49 (12)	L fusiform gyrus	37
−48 −69 −15	3.96 (86)	L middle occipital gyrus	19
−24 −75 51	4.26 (48)	L posterior intra-parietal sulcus	7/19
48 9 27	3.55 (17)	R inferior frontal gyrus	6/9
45 −54 −18	4.52 (304)	R fusiform gyrus	37
48 −81 0	3.53 (25)	R middle occipital gyrus	19
Auditorily selective			
−27 −3 −21	4.40 (44)	L anterior medial temporal lobe/amygdala	
−36 −36 −27	3.44 (9)	L cerebellum	
36 −15 −36	4.13 (10)	R entorhinal cortex	20
15 −15 −24	4.34 (59)	R hippocampus	
21 −36 −3	3.74 (9)	R posterior hippocampus/parahippocampal gyrus	

3.2.1. Modality-independent subsequent memory effects

Regions where subsequent source memory effects were insensitive to modality were identified by exclusively masking the main effect of subsequent memory (source hit > forgotten, collapsed over modality) with the bi-directional (F) contrast for the subsequent memory × modality interaction. As detailed in Table 2 and illustrated in Fig. 1, this analysis identified effects in bilateral inferior frontal gyrus (greater in extent on the left), bilateral fusiform cortex, left posterior intra-parietal sulcus (IPS) and left anterior MTL in the vicinity of the amygdala.

3.2.2. Overlap between subsequent source memory and modality effects

Visual effects: Visually selective subsequent source memory effects that overlapped with regions selectively responsive on visual vs. auditory study trials were identified by inclusively masking the visual subsequent memory contrast with the main effect of modality (visual > auditory; see Section 2). To ensure that any resulting effects were selective for the visual modality, voxels where analogous auditory subsequent memory effects were also significant were removed by using the auditory subsequent memory contrast as an exclusive mask ($p < .05$). This procedure identified several clusters (see Table 3 and Fig. 2b), localized predominantly to regions within bilateral fusiform cortex and intra-parietal sulcus.

Auditory effects: An analogous approach was used to identify overlap between auditory subsequent memory effects and auditorily selective regions. A cluster was identified in right middle superior temporal sulcus (see Table 3 and Fig. 2c), along with a

Table 3
Regions where visually and auditorily selective subsequent memory effects overlapped with global modality effects.

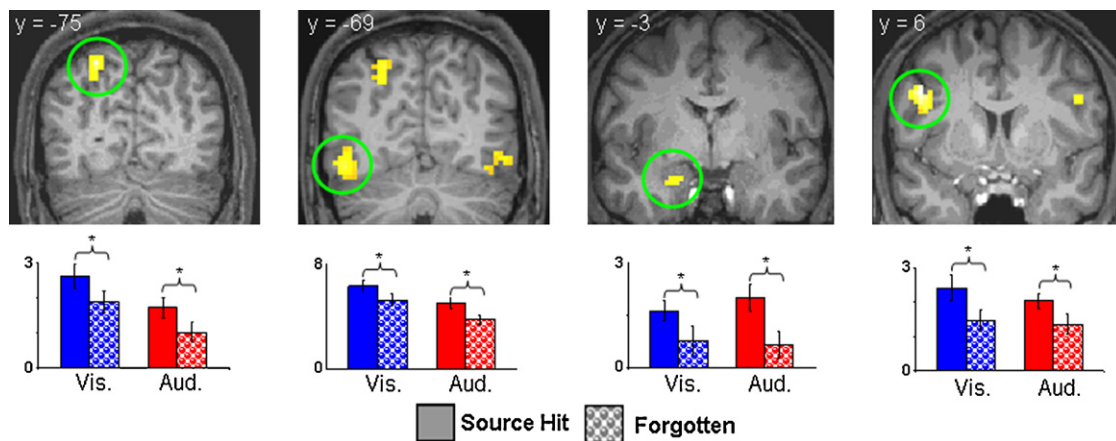| Coordinates (x, y, z)	Z (#voxels)	Region	BA
Visual effects			
−27 −72 54	3.90 (108)	L intra-parietal sulcus	7
27 −54 45	3.46 (53)	R intra-parietal sulcus	7
36 −63 −24	3.49 (20)	R fusiform gyrus	19
30 −63 −12	3.27 (23)	R fusiform gyrus	37
36 −84 0	2.73 (19)	R middle occipital gyrus	19
Auditory effects			
60 −9 −15	3.08 (25)	R middle superior temporal sulcus	21
45 −42 −3	2.69 (9)	R posterior temporal lobe (white matter)	

Fig. 1. Modality-independent subsequent source memory effects (displayed at p < .001, nine voxel extent threshold) projected onto sections of a representative subject's normalized structural image. Bar plots show (left to right) parameter estimates (in arbitrary units) for source hit and forgotten trials at peak voxels in left intra-parietal sulcus (−24 −75 51), left middle occipital gyrus (−48 −69 −15), left amygdala (−21 −3 −21), and left inferior frontal gyrus (−48 6 33). *p < .05.

small cluster localized to white matter adjacent to posterior STS (see Table 3).

To determine if the subsequent source memory effects elicited by visual and auditory study trials in each of the regions exhibiting modality-selective effects differed significantly in their magnitudes, the peak parameter estimates representing the visual and auditory subsequent memory effects in each region were directly contrasted (the equivalent of a 2 (source correct vs. incorrect) × 2 (auditory vs. visual) interaction contrast). In one region associated with subsequent visual source memory – right fusiform cortex (30 −63 −12) – the visual effects were reliably greater than the auditory effects (t(19) = 1.93, p < .05 1-tailed; see Fig. 2b). A reliable difference in the opposite direction was evident for the right STS region that exhibited an auditorily selective subsequent source memory effect (t(19) = 4.37, p < .001 1-tailed; see Fig. 2c).

3.2.3. Modality-selective subsequent memory effects unconstrained by global modality effects

Visually selective subsequent source memory effects were identified by the contrast between visual source hits and visual misses (p < .001), exclusively masked by the analogous contrast for the auditory condition (p < .05). This procedure revealed a single effect in right IPS (nine voxels, 27 −54 45, peak Z = 3.46, BA 7) that overlapped one of the regions identified in the analysis of visually selective effects described above. The analogous procedure identified five MTL clusters where subsequent source memory effects were auditorily selective (Table 2 and Fig. 3). Three clusters were located in bilateral MTL in the vicinity of hippocampus and amygdala, while the remaining clusters were located in right entorhinal cortex and left cerebellum. With the exception of the left anterior MTL (−27 −3 −21), contrasts of the peak parameter estimates revealed that in each case the auditory subsequent memory effects were significantly larger than the visual effects (t(19) = 1.78–3.79, p's < .05 to < .001 1-tailed).

4. Discussion

We contrasted the neural correlates of the incidental encoding of associations between pictures and the visual or auditory contextual information carried by a co-presented word. In addition to regions where subsequent source memory effects were common to the two types of association, we identified other regions where source effects were modality-selective—that is, reliable for only one of the two modalities. As predicted, some of these modality-selective effects were evident in regions that also responded preferentially to the corresponding class of study trials. To the best of our knowledge, the present findings constitute the first direct evidence that successful encoding of a contextual feature of a study episode is associated with enhanced activity in cortical regions selectively engaged during its on-line processing (cf. Uncapher et al., 2006; Uncapher & Rugg, 2009).

4.1. Behavioral findings

RTs during study were significantly slower for the auditory trials compared to the visual trials, an effect that likely reflects the more temporally extended delivery of the auditory information. Crucially, however, there was no evidence of a subsequent memory effect in RTs or judgment accuracy for either class of study trial. Thus, neither the visual nor the auditory fMRI subsequent memory effects discussed below can be attributed to differential study performance associated with later remembered vs. later forgotten study trials.

4.2. fMRI findings

A number of regions were identified where activity was predictive of accurate source memory for both visual and auditory trials, including bilateral fusiform cortex, bilateral dorsal IFG, and left anterior MTL in the vicinity of the amygdala. Subsequent memory effects predictive of successful episodic retrieval have been identified previously in all of these regions (see Cansino, Maquet, Dolan, & Rugg, 2002; Gold et al., 2006; Ranganath et al., 2004 for e.g. of prior fusiform, dorsal IFG and amygdala effects respectively). We assume that these modality-independent effects reflect the successful encoding of episodic information (such as the identity of the picture) that was independent of the modality of the associated word, and to which modality-specific information was bound. Arguably, such effects are to be expected, since failure to incorporate information about picture identity into the memory representation of the study episode would likely lead to failure on the subsequent memory test. In addition, it is possible that some of these modality-independent effects (notably, those in posterior prefrontal cortex and IPS—part of the 'dorsal frontoparietal attentional network'; Corbetta & Shulman, 2002) may reflect the benefit to modality memory of efficient redeployment of attention from the centrally presented picture to the later-presented word.

One seeming inconsistency with prior findings is the absence of modality-independent subsequent source memory effects in the hippocampus (cf. Davachi et al., 2003; Ranganath et al., 2004;

Uncapher & Rugg, 2009). This inconsistency may however be more apparent than real. When the threshold of the contrast employed to identify modality-insensitive effects was lowered to $p < .01$, the right anterior MTL effect was joined by a more posterior cluster in right anterior hippocampus ($27 - 12 - 21$, $Z = 2.97$). This remained significant ($p < .05$) when subjected to a small volume correction based on a 5 mm radius sphere centered on the right hippocampal subsequent source memory effect reported by Uncapher et al. (2006; 27 -15 -15).

Analysis of modality-selective subsequent memory effects focused on the prediction that source effects should overlap with regions that responded preferentially to the corresponding class of study trials (see Section 1). This prediction was confirmed for both context modalities. Visually selective subsequent source memory effects were evident in several posterior regions, whereas auditorily selective source memory effects were identified in right
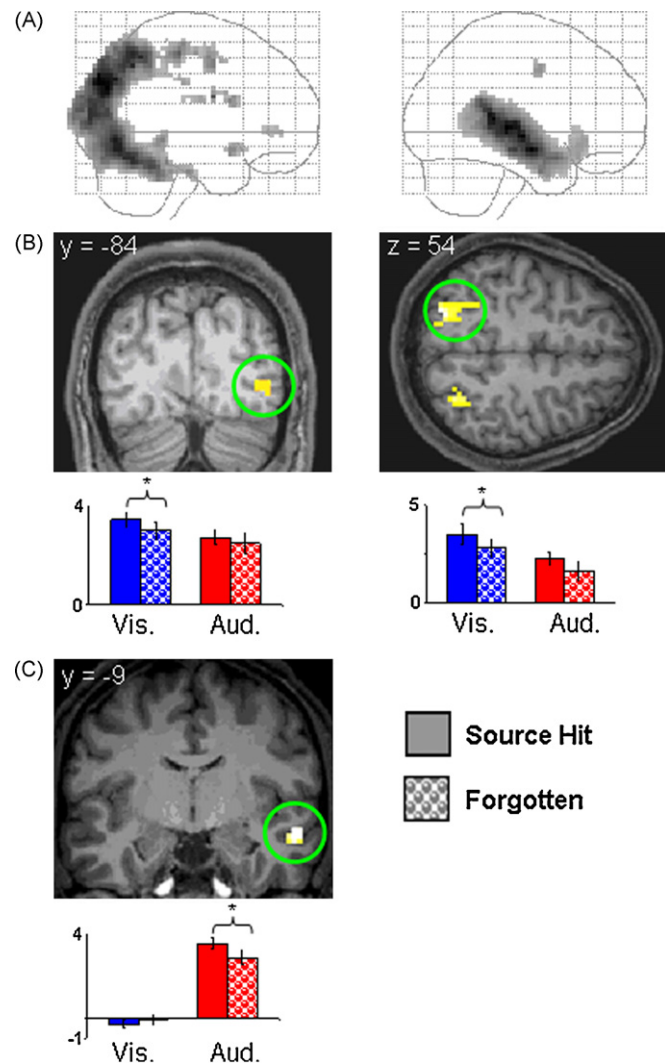
Fig. 2. (A) Maximum intensity projections showing visually (left) and auditorily (right) selective regions. (B) Upper: visually selective subsequent memory effects ($p < .01$) that overlap regions demonstrating visually selective activity projected onto sections of a representative subject's normalized structural image. Lower: peak parameter estimates (in arbitrary units) for source hit and forgotten trials at peak voxels in (left to right) right middle occipital gyrus (36 −84 0) and intra-parietal sulcus (−27 −72 54). (C) Upper: auditorily selective subsequent memory effects ($p < .01$) that overlap auditorily selective regions projected onto sections of a representative subject's normalized structural image. Lower: peak parameter estimates (in arbitrary units) for source hit and forgotten trials at the peak voxel in right middle superior temporal sulcus (60 −9 −15). *$p < .05$. All subsequent memory effects displayed at $p < .001$, nine voxels extent threshold.
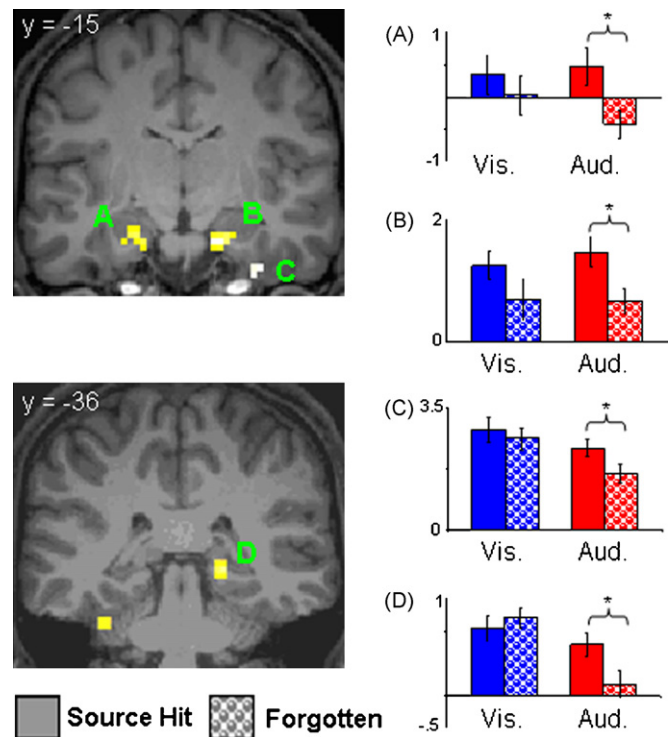
Fig. 3. Auditorily selective subsequent source memory effects unconstrained by the global modality effect (displayed at $p < .001$, nine voxels extent threshold) projected onto sections of a representative subject's normalized structural image. Bar plots (from top to bottom) illustrate peak parameter estimates (in arbitrary units) for source hit and forgotten trials at peak voxels in left amygdala (A, $-27 - 3 - 21$), right hippocampus (B, $15 - 15 - 24$), right entorhinal cortex (C, $36 - 15 - 36$), and right posterior hippocampus/parahippocampal gyrus (D, $21 - 36 - 3$). *$p < .05$.

mid-STS (Fig. 2). It should be noted, however, that in the case of visually selective effects, in only one case did the effect significantly exceed the magnitude of the corresponding (albeit non-significant) auditory effect, and then at only a modest level of statistical significance. Thus, the evidence for overlap between modality-selective subsequent memory effects and modality-sensitive cortical activity is arguably somewhat stronger for the auditory than the visual condition. That said, these results add significantly to our prior finding that the cortical regions engaged during episodic encoding differ according to the nature of the accompanying contextual information (Uncapher et al., 2006).

The findings suggest that successful encoding of the modality of the study words was associated with enhanced processing in only a subset of the cortical regions engaged by each class of study trial. The localization of the visual subsequent source memory effect to right ventral fusiform and bilateral IPS suggests that memory for visual contextual information was facilitated when the context word elicited a strong visual representation of its referent (the posterior IPS has been shown to represent object-level information; Konen & Kastner, 2008). In addition, since successful encoding in the visual condition required a shift of visual-spatial attention (from the picture to the word presented below it), the visually selective IPS subsequent memory effects might, like the modality-independent effects discussed above, reflect the mnemonic benefit of efficient attentional re-orienting. Analogously, the finding that the auditory source memory effect was localized to right mid-STS – a region selectively activated during perception of speech relative to other auditory input (Hein & Knight, 2008; Hickok & Poeppel, 2007; Scott, Rosen, Lang, & Wise, 2006) – suggests that memory for auditory information benefited when processing of the phonological attributes of the context word was emphasized.

In addition to modality-selective subsequent source memory effects that overlapped the main effects of modality, additional source effects were also identified in analyses that were unconstrained by generic modality effects. Of most interest is the finding of an auditorily selective effect in bilateral MTL, including the hippocampus (Fig. 3). This finding is of course consistent with the numerous prior reports of hippocampal subsequent source memory effects (see above and Section 1). The question arises however why memory for auditory information should have been associated with more extensive hippocampal effects than memory for visual contexts. In light of the equivalent levels of visual and auditory source memory on the later memory test, it is unlikely this finding reflects differences in the difficulty or efficacy of visual vs. auditory encoding. An intriguing possibility is that successful encoding of across-modality item-context associations places more demands on hippocampal resources than does encoding of within-modality associations, consistent perhaps with the proposal that 'across-domain' associations are more dependent upon the hippocampus than are 'within-domain associations' (Mayes, Montaldi, & Migo, 2007). Alternatively, the findings might reflect a simple modality effect, indicating that memory for auditory information places more demands on hippocampally supported encoding processes than does visual information (a consequence, perhaps, of the temporally extended nature of the auditory items, leading to heavier or more sustained attentional demands than in the case of the visual words). The lack of prior studies investigating the neural correlates of the encoding of auditory episodic information (though see Peters et al., 2007; Poppenk et al., 2008) means that resolution of this issue will have to await further research.

In conclusion, the present findings extend prior research on the neural correlates of episodic encoding in three principal ways: they provide direct evidence that encoding of a contextual feature is associated with enhanced activity in cortical regions engaged during its on-line processing; they demonstrate that such feature-selective encoding effects extend to the auditory modality; and they suggest that encoding of across-modality associations, or, perhaps, episodes that contain auditory contextual information, places greater demands on the hippocampus than does the encoding of exclusively visual episodic information.

Acknowledgments

References

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 7041–7045.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. E. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, 281, 1185–1187.

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex*, 12, 1048–1056.

Cocosco, C. A., Kollokian, V., Kwan, R. S., & Evans, A. C. (1997). Brainweb: Online interface to a 3D MRI simulated brain database. In *Proceedings of the 3rd international conference on functional mapping of the human brain* Copenhagen, Denmark, (p. S245).

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, 3, 201–215.

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, 16, 693–700.

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

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences*, 11, 379–386.

Fisher, R. (1950). *Statistical methods for research workers* (11 ed.). London: Oliver & Boyd.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *Neuroimage*, 7, 30–40.

Friston, K. J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Theory. *Neuroimage*, 16, 465–483.

Friston, K., Holmes, A. P., Worsely, K., Poline, J., Frith, C., & Frackowiak, R. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 189–210.

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

Gottfried, J. A., Smith, A. P. R., Rugg, M. D., & Dolan, R. J. (2004). Remembrance of odors past: Human olfactory cortex in cross-modal recognition memory. *Neuron*, 42, 687–695.

Hein, G., & Knight, R. T. (2008). Superior temporal sulcus-it's my area: Or is it? *Journal of Cognitive Neuroscience*, 20, 2125–2136.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.

Johnson, J. D., & Rugg, M. D. (2007). Recollection and the reinstatement of encoding-related cortical activity. *Cerebral Cortex*, 17, 2507–2515.

Josephs, O., & Henson, R. N. (1999). Event-related functional magnetic resonance imaging: Modelling, inference and optimization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 354, 1215–1228.

Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-neuroanatomic correlates of recollection: Implications for models of recognition memory. *Journal of Neuroscience*, 28, 4172–4180.

Khader, P., Burke, M., Bien, S., Ranganath, C., & Rosler, F. (2005). Content-specific activation during associative long-term memory retrieval. *Neuroimage*, 27, 805–816.

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

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

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, 11, 224–231.

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown UP.

Lazar, N. A., Luna, B., Sweeney, J. A., & Eddy, W. F. (2002). Combining brains: A survey of methods for statistical pooling of information. *Neuroimage*, 16, 538–550.

Mayes, A., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, 11, 126–135.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *The Journal of Neuroscience*, 24, 4912–4917.

Moscovitch, M. (1992). Memory and working with memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience*, 4, 257–267.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, 110, 611–646.

Nyberg, L., Habib, R., McIntosh, A. R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11120–11124.

Nyberg, L., Petersson, K. M., Nilsson, L. G., Sandblom, J., Aberg, C., & Ingvar, M. (2001). Reactivation of motor brain areas during explicit memory for actions. *Neuroimage*, 14, 521–528.

Otten, L. J., & Rugg, M. D. (2001). Task-dependency of the neural correlates of episodic encoding as measured by fMRI. *Cerebral Cortex*, 11, 1150–1160.

Otten, L. J., Henson, R. N. A., & Rugg, M. D. (2002). State-related and item-related neural correlates of successful memory encoding. *Nature Neuroscience*, 5, 1339–1344.

Park, H., Uncapher, M. R., & Rugg, M. D. (2008). Effects of study task on the neural correlates of source encoding. *Learning & Memory*, 15, 417–425.

Persson, J., & Nyberg, L. (2000). Conjunction analyses of cortical activations common to encoding and retrieval. *Microscopy Research and Technique*, 51, 39–44.

Peters, J., Suchan, B., Köster, O., & Daum, I. (2007). Domain-specific retrieval of source information in the medial temporal lobe. *European Journal of Neuroscience*, 26, 1333–1343.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, 10, 59–63.

Poppenk, J., Walia, G., McIntosh, A. R., Joanisse, M. F., & Kohler, S. (2008). Why is the meaning of a sentence better remembered than its form? An fMRI study on the role of novelty-encoding processes. *Hippocampus*, 18, 909–918.

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

Rolls, E. T. (2000). Memory systems in the brain. *Annual Review of Psychology*, 51, 599–630.

Rugg, M. D., Johnson, J. D., Park, H., & Uncapher, M. R. (2008). Chapter 21 encoding-retrieval overlap in human episodic memory: A functional neuroimaging perspective. *Progress in Brain Research*, *169*, 339–352.

Rugg, M. D., Otten, L. J., & Henson, R. N. A. (2002). The neural basis of episodic memory: Evidence from functional neuroimaging. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*, 1097–1110.

Scott, S. K., Rosen, S., Lang, H., & Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech—A positron emission tomography study. *Journal of the Acoustical Society of America*, *120*, 1075–1083.

Shastri, L. (2002). Epidodic memory and cortico-hippocampal interactions. *Trends in Cognitive Sciences*, *6*, 162–168.

Snodgrass, J., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology*, *117*, 34–50.

Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *The Journal of Neuroscience*, *26*, 9162–9172.

Staresina, B. P., & Davachi, L. (2008). Selective and shared contributions of the hippocampus and perirhinal cortex to episodic item and associative encoding. *Journal of Cognitive Neuroscience*, *20*, 1478–1489.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Oxford University Press.

Uncapher, M. R., Otten, L. J., & Rugg, M. D. (2006). Episodic encoding is more than the sum of its parts: An fMRI investigation of multifeatural contextual encoding. *Neuron*, *52*, 547–556.

Uncapher, M. R., & Rugg, M. D. (2009). Selecting for memory? The influence of selective attention on the mnemonic binding of contextual information. *The Journal of Neuroscience*, *29*, 8270–8279.

Vaidya, C. J., Zhao, M., Desmond, J. E., & Gabrieli, J. D. E. (2002). Evidence for cortical encoding specificity in episodic memory: Memory-induced re-activation of picture processing areas. *Neuropsychologia*, *40*, 2136–2143.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., et al. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*, 1188–1191.

Wheeler, M. E., & Buckner, R. L. (2003). Functional dissociation among components of remembering: Control, perceived oldness, and content. *Journal of Neuroscience*, *23*, 3869–3880.

Wheeler, M. E., & Buckner, R. L. (2004). Functional-anatomic correlates of remembering and knowing. *Neuroimage*, *21*, 1337–1349.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 11125–11129.

Wheeler, M. E., Shulman, G. L., Buckner, R. L., Miezin, F. M., Velanova, K., & Petersen, S. E. (2006). Evidence for separate perceptual reactivation and search processes during remembering. *Cerebral Cortex*, *16*, 949–959.

Woodruff, C. C., Johnson, J. D., Uncapher, M. R., & Rugg, M. D. (2005). Content-specificity of the neural correlates of recollection. *Neuropsychologia*, *43*, 1022–1032.