



# HHS Public Access

Author manuscript

*Neuroimage*. Author manuscript; available in PMC 2017 July 11.

Published in final edited form as:

*Neuroimage*. 2015 April 01; 109: 118–129. doi:10.1016/j.neuroimage.2015.01.003.

## Cortical reinstatement and the confidence and accuracy of source memory

Preston P. Thakral, Tracy H. Wang, and Michael D. Rugg

Center for Vital Longevity and School of Behavioral and Brain Sciences, University of Texas at Dallas

### Abstract

Cortical reinstatement refers to the overlap between neural activity elicited during the encoding and the subsequent retrieval of an episode, and is held to reflect retrieved mnemonic content. Previous findings have demonstrated that reinstatement effects reflect the quality of retrieved episodic information as this is operationalized by the accuracy of source memory judgments. The present functional magnetic resonance imaging (fMRI) study investigated whether reinstatement-related activity also co-varies with the confidence of accurate source judgments. Participants studied pictures of objects along with their visual or spoken names. At test, they first discriminated between studied and unstudied pictures and then, for each picture judged as studied, they also judged whether it had been paired with a visual or auditory name, using a three-point confidence scale. Accuracy of source memory judgments – and hence the quality of the source-specifying information – was greater for high than for low confidence judgments. Modality-selective retrieval-related activity (reinstatement effects) also co-varied with the confidence of the corresponding source memory judgment. The findings indicate that the quality of the information supporting accurate judgments of source memory is indexed by the relative magnitude of content-selective, retrieval-related neural activity.

### 1. Introduction

Episodic memory refers to the processes that support the retrieval of qualitative information about unique events, such as their temporal and spatial contexts (Tulving, 1983). The retrieval of qualitative information – also known as ‘recollection’ – can be distinguished from ‘familiarity’, a memory signal that supports judgments of prior occurrence in the absence of contextual information (Mandler, 1980; Yonelinas, 2002; Wixted and Mickes, 2010). It has been argued that, among other memory tests, both the ‘remember/know’ procedure and, critically for the aims of the present study, judgments of source memory, require retrieval of contextual information and hence depend heavily on encoding and retrieval processes supporting recollection (e.g. Yonelinas, 2002; Wixted and Mickes, 2010).

Corresponding author: Preston P. Thakral, Center for Vital Longevity and School of Behavioral and Brain Sciences, University of Texas at Dallas, Dallas, TX 75235, preston6667@gmail.com, Tel.: +1 972 883 3737, Fax: +1 972 883 3250.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

An account of episodic memory relevant to the present study proposes that the processes engaged when an episode is initially experienced and when it is later recollected are strongly related. The account is derived from the complementary frameworks of transfer-appropriate processing (TAP) and cortical reinstatement theory (Rugg et al., 2008). The principle of transfer-appropriate processing is based on the twin notions that memories are represented in terms of the cognitive processes engaged when an event is experienced, and that retrieval occurs when a retrieval cue initiates the recapitulation or reinstatement of those processes (e.g., Morris et al., 1977; for a review, see Roediger et al., 2002). The idea that successful recollection depends on the reinstatement of processes engaged during encoding is also found in a number of neurally-inspired models of episodic memory (e.g., Alvarez and Squire, 1994; McClelland et al., 1995; Rolls, 2000; Shastri, 2002; Norman and O'Reilly, 2003). According to the model proposed by Norman and O'Reilly (2003), for example, successful encoding occurs when a pattern of cortical activity elicited by an experience is stored by the hippocampus. Subsequently, when a retrieval cue elicits a partially overlapping pattern of activity, pattern-completion within the hippocampus leads to reinstatement of the whole pattern, providing access to the content of the stored memory representation. The cortical reinstatement framework leads to a testable prediction regarding the relationship between neural activity associated with encoding and retrieval: if reinstatement of the processes and representations active during encoding supports episodic retrieval, greater reinstatement should be observed when retrieval is successful than when it is unsuccessful.

Cortical reinstatement effects have been identified in numerous prior studies that have used a variety of test procedures to operationalize recollection (e.g., Kahn et al., 2004; Woodruff et al., 2005; Johnson and Rugg, 2007; Slotnick, 2009; McDuff et al., 2009; Johnson et al., 2009; Huijbers et al., 2011; Kuhl et al., 2011; Hofstetter et al., 2012; Staresina et al., 2012; Johnson et al., 2013; Kuhl et al., 2013; Ritchey et al., 2013; Gordon et al., in press; for reviews of early work, see Rugg et al., 2008; Danker and Anderson, 2010). In the study of Johnson and Rugg (2007), for example, participants were presented at study with words in association with either a scene or a sentence task. Participants were later presented with old and new words and undertook a remember/know test. Relative to test words judged as familiar only, words from each study task endorsed as recollected elicited activity that overlapped regions demonstrating enhanced study activity for the corresponding task. Whereas Johnson and Rugg (2007) operationalized reinstatement in terms of regional overlap in mean signal change, reinstatement has also been demonstrated using multivariate analysis of fMRI data (e.g., Johnson et al., 2009; McDuff et al., 2009; Kuhl et al., 2011, 2013; Staresina et al., 2012; Ritchey et al., 2013; Gordon et al., in press). For example, in Staresina et al. (2012) cross-voxel similarity between study and test trials in parahippocampal cortex was reported to be higher when a cue word elicited accurate rather than inaccurate retrieval of the scene it had been paired with at study.

These findings clearly demonstrate that reinstatement effects reflect, in part, the content of retrieved episodic information. Moreover, findings that reinstatement effects are greater for accurate than for inaccurate content-based (i.e. source) memory judgments (Kahn et al., 2004; Kuhl et al., 2011; Staresina et al., 2012; Gordon et al., in press; see also, Slotnick, 2009; Huijbers et al., 2011; Hofstetter et al., 2012; Kuhl et al., 2013) are consistent with the proposal that the reinstated information provides the basis for such judgments.

Until now, the relationship between reinstatement and the information supporting content-based (source) memory judgments has been investigated through contrasts between test trials attracting correct and incorrect judgments. There is however considerable evidence that the accuracy of a source memory judgment co-varies with the confidence expressed in the judgment, suggesting the existence of a recollection signal that varies continuously in its fidelity or 'strength' (Slotnick et al., 2000; Qin et al., 2001; Glanzer et al., 2004; Slotnick and Dodson, 2005; Wixted, 2007; Mickes et al., 2009; Slotnick, 2010; Yu et al., 2012a; but see, Parks and Yonelinas, 2007). Thus, if reinstatement effects provide the informational basis for judgments about retrieved episodic content (as in source memory tests), the effects should vary not only as a function of accuracy, but also between accurate judgments made with differing levels of confidence. Such a finding would demonstrate both that reinstatement tracks the fidelity or quality of retrieved content, and that source judgments are indeed based on a continuous memory signal. An alternative possibility is that reinstatement-related activity reflects the mere detection of recollected information, regardless of the quality of that information. Evidence in favor of this possibility would take the form of reinstatement-related activity that was sensitive to whether recollection was successful (as indexed by an accurate source judgment) but was unaffected by recollection strength (as indexed by source confidence; cf., Yu et al., 2012b). To our knowledge, this issue has yet to be addressed.

Here, we addressed the issue by employing a source memory procedure and assessing whether cortical reinstatement tracks the confidence and accuracy of source memory judgments (we expected that these two variables would co-vary, as has been reported previously; see above). Participants studied pictures of objects along with their visual or spoken names. At test, participants discriminated between studied and unstudied pictures and, for each picture judged studied, went on to judge whether the picture had been paired with a visual or a spoken name, using a three-point confidence scale.

Successful episodic retrieval is associated not only with content-selective reinstatement effects, but also with content-independent retrieval effects. Content-independent effects have been consistently observed in a characteristic network of regions, sometimes referred to as the 'core recollection network' (for reviews, see Kim, 2010; Rugg and Vilberg, 2013). It has been proposed that, in interaction with regions manifesting content-selective retrieval effects, this network supports the retrieval and maintenance of consciously accessible memory representations (Rugg and Vilberg, 2013). The network comprises the hippocampus and parahippocampal cortex, along with ventral parietal cortex, retrosplenial/posterior cingulate cortex, and medial prefrontal cortex. It has been reported that retrieval-related activity in these regions, most notably in ventral parietal cortex in the vicinity of the angular gyrus and in the hippocampus, co-varies with the confidence and accuracy of source memory judgments (Yu et al., 2012a, 2012b; but see, Wais et al., 2010; Slotnick and Thakral, 2013). For example, in the study of Yu et al. (2012a, 2012b), retrieval-related activity elicited by items endorsed as 'remembered' was compared across three levels of source confidence. Activity in the hippocampus and angular gyrus co-varied with confidence in a graded fashion (i.e., high confidence > moderate confidence, moderate confidence > low confidence). These and related findings prompted Rugg and Vilberg (2013) to suggest that activity in these regions co-varies with the amount of retrieved episodic information (see

also, Hayama et al., 2012; Rugg et al., 2012). We expected to obtain similar results in the present study.

## 2. Materials and Methods

### 2.1 Participants

Twenty participants with normal or corrected-to-normal vision completed the experiment. Participants reported themselves to be right-handed and free of neurological and psychiatric disease. They had no contraindications for MRI. Informed consent was obtained prior to participation. The experimental protocol was approved by the Institutional Review Boards of the University of Texas at Dallas and the University of Texas, Southwestern Medical Center at Dallas. Data from two participants were excluded from all analyses because of excessive head movement (> 3 mm). An additional participant was excluded from the analyses because of too few high source confidence judgments for visually studied items ( $n = 3$ ). The remaining 17 participants (8 females) had a mean age of 22.7 years (range 19–28).

### 2.2 Stimulus Materials

Experimental items comprised 291 stimulus triplets consisting of a colored picture of a common object and the visual and auditory name of the object. The objects were drawn from Hemera Photo Objects 50,000 Volume III (<http://www.hemera.com/index.html>). Their names ranged in length from 3 to 10 letters and in frequency from 1 to 100 counts/million (Kucera and Francis, 1967). Auditory versions of the names were spoken by a female voice. The recordings were edited to a constant sound pressure level and filtered to remove ambient noise (<http://audacity.sourceforge.net>). Auditory names were presented binaurally via MR compatible earphones (Sensimetrics Insert Earphones for fMRI Research Model S14, <http://www.sens.com/>) and did not exceed 1000 ms in duration (mean duration 650 ms). Volume was adjusted to a comfortable level for each individual participant during a preliminary MRI scan.

For each participant, the 291 triplets were randomly sorted. Eighteen triplets were used as buffers (2 at the beginning and end of each study session and 2 at the beginning of each test session). The remaining triplets were subdivided into those to be used during study and new items to be used for the test task. During study (Figure 1A), 144 pictures were paired with matching names (72 visual and 72 auditory ‘congruent’ trials) and 18 pictures were paired with mismatching names that denoted objects not otherwise employed in the experiment (9 visual and 9 auditory ‘incongruent’ trials - as described below, the congruent and incongruent trials were inherent to the study task). At test (Figure 1B), the pictures from 72 of the triplets were used as new items and the 144 pictures presented on congruent study trials served as old items. Twenty-one triplets were used as practice items for the study and test tasks.

A study list was created for each participant using the 144 congruent picture-name pairings and the 18 incongruent pairings. The study list was subdivided into three sub-lists, one for each fMRI study session. Each sub-list contained a pseudorandom ordering of 48 congruent pairings (24 visual and 24 auditory) and 6 incongruent pairings (3 visual and 3 auditory). No

more than three trials were consecutively presented in the same modality. One test list was created using the pictures from the 144 congruent picture-name pairings presented at study and 72 new pictures. The test list was subdivided into three test sub-lists, one for each fMRI test session. Each sub-list contained a pseudorandom ordering of 48 old pictures (24 auditory and 24 visual) and 24 new pictures. No more than 3 trials of the same study status were consecutively presented.

Pictures were presented in central vision on a mirror mounted above the scanner head-coil within a grey box that subtended a visual angle of  $6.7^\circ \times 6.7^\circ$ . During the study phase, visual names were presented directly below the grey box in lowercase Arial font in white (25 point font and subtending a maximum horizontal visual angle of  $3.6^\circ$ ). When the accompanying name was presented auditorily, only the picture in the grey box was presented.

### 2.3 Experimental Procedures

Participants completed one study-test cycle. Each study and test phase was broken into three fMRI scanning sessions, for a total of 6 fMRI sessions (3 study fMRI sessions followed by 3 test fMRI sessions). Following the study phase, participants were instructed on the test task and practiced it while remaining in the scanner.

An incidental study task, which had been practiced outside the scanner, was employed (Figure 1A; see also, Gottlieb et al., 2010). The task was to judge whether each picture and its accompanying name was congruent or incongruent (representative trials are shown in Figure 1A left and right, respectively). Participants were instructed to maintain fixation and to respond as quickly as possible without sacrificing accuracy. Responses were made on a button-box using the middle and index fingers. The hand used for responding was counterbalanced across participants.

Each study trial began with the presentation of a red fixation cross at the center of the screen for 500 ms. This was replaced by the picture of an object for a duration of 1000 ms. Five hundred milliseconds into the presentation of the object, the visual or auditory name was presented. Following picture offset, a white fixation cross was displayed for a variable duration of 6.5, 8.5, or 10.5 s to conclude the trial. The fixation durations were equivalently distributed across the visual and auditory congruent trials within each study session (32 trials at 6.5 s, 12 trials at 8.5 s, and 4 trials at 10.5 s). The fixation duration for buffer trials and incongruent trials was 6.5s<sup>1</sup>.

The test task had two stages. First, participants judged whether each test item had been presented at study (old) or was new (Figure 1B) under the instruction to respond “old” only when confident the picture had been studied. Following an old response, the requirement was then to judge whether the test item had been paired at study with an auditory or visual word. The judgment could be made using one of three confidence levels; “definitely”, “probably” and “possibly”, giving a total of six response options. Instructions were to make

---

<sup>1</sup>As is apparent from the experimental design, the present study was intended to allow the fMRI data to be subjected to multi-voxel pattern analysis (MVPA) in addition to the univariate analyses described here. Contrary to what would have been expected from prior findings (e.g., Gordon et al., in press; Johnson et al., 2009), a pattern classification analysis failed to identify evidence of modality-selective cortical reinstatement effects, and hence added nothing to the results reported here.

all responses as quickly as possible without sacrificing accuracy. Old/New responses were made on a button-box using the left or right middle and index fingers. Source judgments were made using left and right middle, index, and ring fingers. The hand used for responding was counterbalanced across participants.

Each test trial began with the presentation of a red fixation cross at the center of the screen for 500 ms. The cross was replaced by the test picture, which remained in view for 500 ms. Following picture offset, an old/new prompt appeared above a white fixation cross for 2 s. If participants responded “old”, a source prompt appeared for 2 s (Figure 1B left). If participants responded “new”, a white central fixation cross appeared for 2 s (Figure 1B right). The trial concluded with a white fixation cross that was displayed for a variable duration of 6.5, 8.5, or 10.5s. Within each test session, the inter-trial interval was distributed equivalently across all old visual, old auditory and new trials (48 trials at 6.5 s, 18 trials at 8.5 s, and 6 trials at 10.5 s in each case)<sup>1</sup>. For the buffer trials at the beginning of each test session the inter-trial interval was 6.5 s. Each study and test session concluded with a white fixation period that was presented for 15 s.

## 2.4 Image Acquisition and Analysis

MR images were acquired with a 3 Tesla Philips Achieva MRI scanner (Philips Medical Systems) using a 32-channel head coil. Functional images were acquired using an echo-planar imaging sequence (EPI: SENSE factor of 1.5, flip angle 70°, 80 × 80 matrix, FOV = 24 cm, TR = 2000 ms, TE 30 ms, 34 slices, 3 mm slice thickness, 1 mm gap, 3 mm isotropic voxels). Slices were acquired in ascending order and were oriented parallel to the anterior-posterior commissure plane. Anatomical images were acquired using a magnetization rapidly acquired gradient echo sequence (MPRAGE: 240 × 240 matrix, 1 mm isotropic voxels). For each study session, 263 volumes were acquired (for one participant, 262, 207, and 267 volumes were acquired in the different study sessions because of technical error). For each test session, 337 volumes were acquired (for one participant, only 245 volumes were acquired in the third test session).

fMRI analysis was conducted using Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, London, UK). Functional image preprocessing included a two-pass spatial realignment (first, realigning all images to the first image of the first session and then realigning to the mean image), slice-time correction (using the middle slice as reference), reorientation, spatial smoothing with a 8 mm full-width-half-maximum (FWHM) Gaussian kernel, and spatial normalization to Montreal Neurological Institute (MNI) space using a standard echo-planar imaging template (Cocosco et al., 1997).

Anatomical images were normalized to MNI space for each individual participant (using the SPM8 T1 template) and then averaged across participants to create a mean image. The time series in each voxel was high-pass filtered at 1/128 Hz and scaled to a constant mean within session.

Statistical analysis was performed on the study and test data separately, in each case using a two-stage mixed effects model. In the first stage, neural activity associated with each picture at study and test was modeled for each participant by a delta function at the onset of the picture. The associated BOLD response was modeled by convolving the functions with a

canonical hemodynamic response (HDR) function and its temporal and dispersion derivative (Friston et al., 1998), yielding regressors in a General Linear Model (GLM) that modeled the BOLD response for each event type.

There were 6 events of interest, comprising trials associated with source judgments segregated according to confidence/accuracy. High visual and auditory source confidence corresponded to test pictures that attracted both an accurate old response and an accurate source response at the highest level of confidence. Low visual and auditory source confidence corresponded to test pictures that attracted an accurate old response and an accurate source response at the moderate or low level of confidence (the “probably” or “possibly” response categories). Limitations on trial numbers necessitated collapsing across “probably” and “possibly” accurate source responses to form this single response category. The proportions of trials attracting accurate probable and accurate possible judgments did not significantly differ between the two modalities (means (standard deviations), 0.58 (0.15) and 0.56 (0.15) for auditory and visual probable judgments, respectively). The final response category of interest comprised test pictures that later attracted an accurate old response and an inaccurate source response at any level of confidence. The relative proportions of inaccurate judgments for visual trials were 0.17 (0.18), 0.30 (0.13), and 0.53 (0.22) for “definitely auditory”, “probably auditory”, and “possibly auditory” responses, respectively. The relative proportions of inaccurate judgments for auditory trials were 0.11 (0.11), 0.34 (0.11), and 0.55 (0.20) for “definitely visual”, “probably visual”, and “possibly visual” responses, respectively. ANOVA (factors of modality and response) gave rise solely to a main effect of response ( $F(1.40, 22.37) = 26.77, p < 0.001$ ). Thus, for both modalities, only a small minority of inaccurate source judgments were made with high confidence.

Two further categories of test events comprised correct rejections (new pictures that attracted an accurate new response) and a single collapsed event of no-interest (comprising misses (studied pictures that later attracted a new response), false-alarms, trials with no response, and buffer trials). Six regressors representing movement-related variance (three for rotation and three for rigid-body translation) and regressors modeling each scan session were also entered into the study and test design matrices. An AR(1) model was used to estimate and correct for nonsphericity of the error covariance (Friston et al., 2002). Participant-wise parameter estimates for each event of interest derived from these within-participant GLMs were taken to the second analysis stage where they were entered into a repeated measures ANOVA with participants modeled as a random effect.

The ANOVA model employed within-participants factors of modality (visual and auditory) and response category (high confidence, low confidence, and inaccurate). Table 1 lists the proportions of old visual and auditory trials associated with each response category. Unless otherwise noted, the primary contrasts were thresholded at  $p < 0.001$  with a cluster extent threshold of 21 voxels. The extent threshold corresponds to a Type I error rate of  $p < 0.05$  as estimated by a Monte Carlo simulation of 10,000 iterations using the AlphaSIM tool in Analysis of Functional NeuroImages (AFNI) software ([http://afni.nimh.nih.gov/pub/dist/doc/program\\_help/AlphaSim.html](http://afni.nimh.nih.gov/pub/dist/doc/program_help/AlphaSim.html)).

### 3. Results

#### 3.1 Behavioral Results

Mean accuracy of the congruency judgments at study was slightly but significantly higher for the visual than the auditory modality (0.98 (0.02) and 0.95 (0.04) respectively;  $t(16) = 3.89$ ,  $p < 0.01$ ). Turning to the test data, item hit rate collapsed over all possible source responses was 0.81 (0.10) for pictures previously paired with a visual word and 0.84 (0.08) for pictures previously paired with an auditory word, against a false alarm rate of 0.06 (0.08). The hit rates significantly differed ( $t(16) = 2.14$ ,  $p < 0.05$ ).

Figure 2A illustrates the accuracy of source judgments (conditionalized on accurate item memory) at each level of source confidence. A two-way repeated measures ANOVA (factors of modality and confidence) gave rise solely to a main effect of confidence ( $F(1, 16) = 92.39$ ,  $p < 0.001$ ; this and all following degrees of freedom were corrected for non-sphericity with the Greenhouse-Geisser procedure). Follow-up pairwise comparisons revealed greater accuracy for high than for low confidence judgments ( $F(1, 34.43) = 104.89$ ,  $p < 0.001$ ; unless otherwise noted, this and all following post-hoc pairwise comparisons were conducted using the standard error of the difference derived from the error term of the relevant parent ANOVA). For both confidence levels accuracy exceeded the chance level of 0.5 (in both cases,  $t(16) = 2.12$ ,  $p < 0.001$ ).

It has been argued that neural correlates of source memory are confounded by differences in the accuracy, and hence the ‘strength’, of memory for the items themselves (Wixted et al., 2010). Thus, we addressed the question whether item memory accuracy co-varied with the confidence and accuracy of the associated source judgment. Figure 2B illustrates the item recognition accuracy (indexed by  $pHit/(pHit + pFA)$ ; Wixted et al., 2010) as a function of response category. Because accuracy approached ceiling, we used nonparametric tests to contrast the scores associated with the different response categories. Bradley’s procedure was used to test for the presence of a modality by response category interaction (Bradley, 1979; Sawilowsky, 1990). The interaction was not significant ( $\chi^2(2) = 2.42$ ,  $p > 0.10$ ). Follow-up pairwise comparisons, collapsed across modality, and conducted using a Wilcoxon Signed Ranked Test, revealed greater recognition accuracy for high than for low confidence judgments ( $p < 0.001$ ) and for low confidence than for inaccurate judgments ( $p < 0.05$ ).

**fMRI Results**—We first identified modality-independent source memory effects, and then went on to identify modality-selective effects (‘cortical reinstatement’). For both classes of effects, we assessed whether retrieval-related activity varied according to the quality of source-specifying information by conducting pairwise comparisons between the peak parameter estimates for each of the three categories of source judgment (high confidence, low confidence, and inaccurate).

#### 3.2 Modality-Independent Source Memory Effects

Regions demonstrating modality-independent source memory effects were identified by exclusively masking the main effect of response category, collapsed across modality, with



the modality by response category interaction (mask threshold of  $p < 0.05$  two-sided; note that the more liberal the threshold of an exclusive mask, the more conservative the procedure). As illustrated in Figure 3 and listed in Table 2, the analysis identified effects in, among other regions, bilateral posterior parietal cortex, bilateral medial prefrontal cortex, bilateral orbitofrontal cortex, left inferior temporal sulcus, and bilateral posterior cingulate/retrosplenial cortex (extending into extrastriate and striate cortex).

Parameter estimates for each response category were extracted from peak voxels of clusters localized within the ‘core recollection network’ (see Introduction and Rugg and Vilberg, 2013). The selected regions were left angular gyrus (MNI coordinates:  $-57 -67 22$ )<sup>2</sup>, retrosplenial/posterior cingulate cortex ( $-9 -52 34$ ), and two regions of medial prefrontal cortex ( $-6 59 16$  and  $-9 65 1$ ). Since the main effect of response category did not identify effects in the hippocampus or parahippocampal cortex, two other members of the putative recollection network, parameter estimates could not be obtained for these regions. Because voxels that demonstrated a modality by response category interaction were excluded from the main effect, extracted parameter estimates were collapsed across modality. A two-way repeated measures ANOVA (factors of region and response category) identified a main effect of region ( $F(2.42, 38.68) = 4.46, p < 0.05$ ), but failed to identify a region by response category interaction ( $F(3.96, 63.28) = 2.42$ ). Therefore, the parameter estimates were collapsed across region. Consistent with the impression given in Figure 3, pairwise comparisons revealed that high confidence judgments were associated with significantly greater activity than either of the other response categories ( $F$ 's(1, 52.44)  $> 65.70, p$ 's  $< 0.001$ ), which did not differ from one another<sup>3</sup>.

### 3.3 Hippocampal Source Memory Effects

In contrast to several prior studies in which source memory effects were identified in the hippocampus (for reviews, see Diana et al. 2007; Mitchell and Johnson 2009; Rugg and Vilberg, 2013), the whole-brain analysis described above failed to reveal such effects. When the threshold for the unmasked ANOVA main effect was lowered to  $p < 0.01$ , however, a cluster including the hippocampus was identified ( $-30 -19 -20$ , peak  $Z = 2.66$ , see Figure 4). The effect survived small volume correction (Worsley et al., 1996) within a 3 mm radius sphere centered on the left hippocampal source memory effect reported by Yu et al. (2012a;  $-30 -16 -20$ ). A two-way repeated measures ANOVA (factors of modality and response category) conducted on the peak parameter estimates (see Figure 4) identified a significant modality by response category interaction ( $F(2.15, 30.7) = 8.45, p < 0.001$ ). Separate ANOVAs conducted on the data for each modality revealed significant main effects of response category in each case (visual:  $F(1.81, 28.87) = 8.19, p < 0.01$ ; auditory:  $F(1.98, 31.73) = 4.71, p < 0.05$ ). Follow-up pairwise comparisons revealed that high confidence

<sup>2</sup>For the purposes of this analysis, the sub-peak of the left lateral parietal cluster closest to the peak of the recollection effect identified in the meta-analysis of Kim et al (2010) was selected. The findings did not differ when the same analysis was repeated using the left lateral parietal coordinate listed in Table 2.

<sup>3</sup>We conducted a whole-brain analysis to confirm that the pattern of effects identified by the parameter estimates was not restricted to the peak voxels identified by the main effect. This analysis entailed inclusively masking each of the three pairwise comparisons of high confidence  $>$  low confidence, low confidence  $>$  inaccurate, and high confidence  $>$  inaccurate (each thresholded at  $p < 0.025$ ) with the main effect of response category. The first and third of these masks identified clusters ( $> 21$  voxels) that were centered around the same peak co-ordinates from which parameter estimates were extracted from. The second mask did not identify any significant voxels. Thus this secondary analysis revealed similar findings to those identified by the analysis of the peak parameter estimates.

visual judgments were associated with significantly greater activity than either of the other response categories for that modality ( $F(1, 56.56) > 10.67$ , both  $p$ 's  $< 0.001$ ), which did not differ from one another. Activity associated with high confidence auditory judgments was also significantly greater than that for low confidence judgments ( $F(1, 56.56) = 11.52$ ,  $p < 0.001$ ). In contrast to the visual condition, however, activity associated with high confidence judgments did not differ from that associated with inaccurate judgments. Moreover, activity was greater for inaccurate than for low confidence auditory judgments ( $F(1, 56.56) = 7.75$ ,  $p < 0.01$ ).

### 3.4 Modality-Selective Source Memory Reinstatement Effects

A three-step procedure was used to identify modality-selective source memory effects. First, regions differentially active on visual and auditory study trials were identified by directional contrasts between the two classes of trial (visual  $>$  auditory and vice-versa, in each case thresholded at  $p < 0.001$ ; see Figures 5A and 5B). Second, the main effect of response category was identified for each modality, using a threshold  $p < 0.01$  (cf., Gottlieb et al., 2010). To identify voxels demonstrating a modality-selective memory effect, each of these main effects was exclusively masked with the alternate effect (mask threshold of  $p < 0.05$ ). The final step was to inclusively mask each modality-selective main effect with the relevant study contrast (giving an estimated conjoint significance of  $p < 0.0001$ ; Fisher, 1950; Lazar et al., 2002). To ensure against Type I error, a cluster extent threshold (to give  $p < 0.05$ ) was estimated and applied to the modality-selective contrasts across response category. The threshold was estimated for the regions identified by each encoding contrast (visual  $>$  auditory: 3196 voxels, auditory  $>$  visual: 2594 voxels). The minimum cluster extents for the retrieval contrasts (thresholded at  $p < 0.01$ ) were 26 and 27 voxels for the visually- and auditorally-selective regions respectively.

As is illustrated in Figure 6A and listed in Table 2, visual source reinstatement effects were identified in a single 86-voxel cluster with a maximum in left middle occipital gyrus. Follow-up pairwise comparisons on the peak parameter estimates revealed that activity was significantly greater for high than for low confidence judgments, as well as for low confidence than for inaccurate judgments ( $F(1, 28.89) > 12.44$ ,  $p$ 's  $< 0.001$ ).

Auditory source reinstatement effects are illustrated in Figure 6B and listed in Table 2. The effects were took the form of a single 106-voxel cluster with a maximum in right superior temporal sulcus. Pairwise comparisons of the peak parameter estimates revealed that activity associated with high confidence judgments was significantly greater than that for low confidence and inaccurate judgments ( $F(1, 31.36) > 4.84$ ,  $p$ 's  $< 0.05$ ). In addition, parameter estimates for inaccurate judgments were significantly greater than those for low confidence judgments ( $F(1, 31.36) = 10.30$ ,  $p < 0.01$ )<sup>4</sup>.

---

<sup>4</sup>As was the case for the modality-independent effects, whole brain analysis of the visual and auditory reinstatement effects led to findings similar to those reported for the extracted parameter estimates.

## 4. Discussion

The primary aim of the present study was to address the question whether content-selective retrieval-related activity co-varies not just with the accuracy, but also with the confidence, of an associated source memory judgment. Content-selective activity elicited by test items attracting accurate, highly confident source judgments was greater than the activity elicited by items associated with accurate judgments made with low confidence. These results are inconsistent with the possibility that reinstatement-related activity reflects detection of source-specifying information, but not its strength or quality. As was expected, the activity elicited by test items attracting inaccurate judgments – that is, items presumed to have elicited little or no information diagnostic of source – was further decremented for objects paired with visual words. Contrary to expectation, however, this pattern was not evident for objects paired with auditory words. Rather, activity elicited by test items studied with auditory words that attracted an incorrect source judgment was enhanced relative to that elicited by items endorsed with an accurate low confidence judgment. We discuss the implications of these findings, along with those pertaining to modality-independent source retrieval effects, below.

### 4.1 Behavioral findings

Replicating previous findings (e.g., Slotnick et al., 2000; Qin et al., 2001; Glanzer et al., 2004; Slotnick and Dodson, 2005; Mickes et al., 2009), the accuracy of source memory judgments co-varied with confidence. As was noted in the Introduction, this is consistent with the proposal that the memory signal supporting source memory is graded (e.g., Slotnick and Dodson, 2005; Mickes et al., 2009; Slotnick, 2010).

Also replicating prior results (e.g., Glanzer et al., 2004; Wixted et al., 2010), item memory accuracy co-varied with source confidence (Figure 2B). This finding is important, as it raises the possibility that the fMRI findings discussed below are reflections of the strength of a generic, undifferentiated memory signal rather than a signal that selectively supports source recollection (e.g., Wixted et al., 2010). However, there are two reasons for doubting this possibility. First, the magnitude of the activity associated with the different categories of source judgment did not invariably track accuracy of item memory. For example, whereas item accuracy differed reliably between low confidence and inaccurate source judgments, retrieval-related activity generally failed to discriminate between these two response categories and, in two cases, demonstrated reliably *greater* activity for the inaccurate judgments (Figures 4 and 6B). Second, it is unclear how the strength of an undifferentiated memory signal could have neural correlates that varied according to the study modality of the test item (Figure 6). Therefore it seems unlikely that the confidence-sensitive fMRI reinstatement effects discussed below can be accounted for solely in terms of generic memory strength.

### 4.2 fMRI findings

**4.2.1 Modality-Independent Source Memory Effects**—Modality-independent source memory effects were identified in cortical regions previously identified as members of a putative ‘core recollection network’, adding to the evidence that the network is insensitive to

the content of recollection (e.g., Duarte et al., 2011; Hayama et al., 2012; Johnson et al., 2013; for a review, see Rugg and Vilberg, 2013). These regions included the angular gyrus, posterior cingulate/retrosplenial cortex, medial prefrontal cortex and, at a reduced threshold, the hippocampus. Source memory effects within these regions varied with the quality of the source-specifying signal, in that the activity associated with highly confident judgments was greater than that for items attracting low confident judgments. This finding is consistent with prior reports that retrieval-related activity in the angular gyrus and hippocampus co-varies with source confidence (Yu et al., 2012a, 2012b; but see, Wais et al., 2010; Slotnick and Thakral, 2013).

A caveat to the conclusion that the activity elicited by accurate source judgments varies as a function of the strength of a continuously varying, source-specifying memory signal arises from the possibility that a larger proportion of low than high confidence judgments were based on ‘lucky guesses’. By this argument, the diminished retrieval-related activity for low than for high confidence judgments reflects the diluting effects of these guesses, in the absence of which the two classes of judgment would be associated with equivalent levels of activity. Given that guesses are necessarily confined to the lowest confidence judgments (i.e., the “possible” response category; see Materials and Methods), we performed an additional analysis on the fMRI data of 15 participants who had sufficient trials to permit contrasts between accurate source judgments at all three levels of confidence (definite, probable, and possible). In each of the regions identified in our main analysis, activity for high confidence judgments reliably exceeded that for the remaining two response categories. In no region did the activity elicited by probable and possible judgments significantly differ. It seems unlikely, therefore, that the sensitivity of retrieval-related activity to source confidence reflects the differential mixing of two classes of activity, associated with accurate judgments and lucky guesses respectively.

Modality-independent retrieval-related activity in cortical regions belonging to the recollection network did not differ between low confidence and inaccurate source judgments (see Figure 3). It has been previously argued that retrieval-related activity in this network co-varies with the total amount of retrieved episodic information, rather than just the information relevant to a specific memory judgment (Hayama et al., 2012; Rugg and Vilberg, 2013; see also, Rugg et al., 2012). If this argument is correct, retrieval-related activity should co-vary with the accuracy and confidence of source judgments only when the different judgments are associated with retrieval of different amounts of episodic information. In the present case, the difference between the amount of source-specifying information necessary to support low confidence rather than inaccurate judgments (or moderate rather than low confidence judgements – see foregoing paragraph) may simply have been too little to lead to a detectable difference in the corresponding fMRI signals. A similar argument was advanced to explain why, in studies employing the remember/know procedure, retrieval-related hippocampal activity did not differ between test items endorsed as familiar only and items given a ‘remember’ judgment when the judgment was associated with retrieval of only a small amount of episodic information (Rugg et al., 2012).

Alternatively, the failure to identify a difference in the activity between low confidence and inaccurate source judgments might be indicative of a non-linear relationship between source

memory strength and the fMRI BOLD response (cf., Squire et al., 2007). By this argument, the form of the hemodynamic response function relating the BOLD signal to source (or, perhaps, item) memory strength is such that the signal is enhanced only when strength is especially high. Hence, the signal is insensitive to differences between items eliciting no (inaccurate judgments) as opposed to weak (low confidence judgments) source specifying information. Although it is not possible to reject this account on the basis of the present data, it is hard to reconcile it with the findings of Yu et al (2012a, 2012b; see also, Diana et al., 2009).

A final possibility is that the absence of a graded relationship between source confidence and modality-independent retrieval effects is a reflection of the nature of the memory signal supporting highly confident source judgments (see Introduction). According to this account, highly confident judgments are based upon an above-threshold recollection signal, whereas lower confidence judgments do not benefit from this signal, and are instead supported by a lower-fidelity, continuous ‘familiarity’ signal with a different neural substrate (Parks and Yonelinas, 2007). This account is however inconsistent with a wealth of behavioral evidence indicating that source judgments are supported by a continuous rather than a discontinuous recollection signal (see above and Introduction).

**4.2.2 Hippocampal Source Memory Effects**—In contrast to the modality-independent cortical retrieval effects discussed above, the profile of retrieval-related activity in the hippocampus differed according to study modality. Activity elicited by test items studied in the visual condition demonstrated a similar profile to that of the modality-independent cortical regions (greater activity for high confidence than for low confidence or inaccurate source judgments). By contrast, the activity elicited by items paired with auditory study words demonstrated a U-shaped profile, such that activity associated with inaccurate judgments was greater than that for low confidence judgments, and did not differ from that elicited by items endorsed with high confidence (see Figure 4). The greater hippocampal activity for inaccurate than for low confidence accurate judgments is reminiscent of the pattern defining hippocampal ‘associative novelty’ effects - the finding that novel conjunctions of familiar items elicit greater hippocampal activity than familiar conjunctions (e.g., Düzel et al., 2003; Köhler et al., 2005; Chen et al., 2011). Thus, the elevated hippocampal activity evident for inaccurate source judgments on items studied in the auditory condition may be a response to the unfamiliarity of the conjunction of the item and the ‘image’ of its spoken name (see below). We admit, however, that this account is both *ad hoc* and offers no explanation for why elevated activity for inaccurate judgments was not also evident for items from the visual study condition.

**4.2.3 Modality-Selective Source Reinstatement Effects**—Source reinstatement effects associated with visual study trials were identified in left middle occipital gyrus, whereas corresponding effects for auditory information were identified in the right superior temporal sulcus. The auditory reinstatement effects add to the rather limited prior evidence (Nyberg et al., 2000; Wheeler et al., 2000; Huijbers et al., 2011) that retrieval-related activity can be identified in auditorily-responsive temporal cortex. Importantly, for both modalities, the magnitude of modality-selective activity was greater for high than for low confidence

source judgments. Thus, the findings extend those of prior studies (Kahn et al., 2004; Kuhl et al., 2011; Staresina et al., 2012; Gordon et al., in press; see also, Slotnick, 2009; Huijbers et al., 2011; Hofstetter et al., 2012; Kuhl et al., 2013) by demonstrating that reinstatement effects can vary not only according to source accuracy, but also with the strength or quality of the memory signal supporting an accurate source judgment.

Consistent with the proposal that the strength of reinstatement indexes the fidelity of the information supporting source judgments (see Introduction), the parameter estimates associated with visual reinstatement not only differed according to source confidence, but were also lower for inaccurate than for low confidence judgments (see Figure 6A). Thus, these results replicate prior findings that reinstatement-related activity is stronger for accurate than inaccurate source judgments (Kahn et al., 2004; Kuhl et al., 2011; Staresina et al., 2012; Gordon et al., in press). In striking contrast, the response profile of the parameter estimates associated with auditory reinstatement demonstrated a U-shaped profile similar to that observed for the hippocampus: items attracting highly confident or inaccurate judgments exceeded the activity associated with low confidence judgments (see Figure 6B). Why test items from the auditory study condition should have elicited this response profile is unclear and, given that the finding was unpredicted and conflicts with prior findings (see above), it should be treated with caution. One highly speculative possibility is that prior to wrongly rejecting items as having been studied in the auditory condition, participants employed something akin to a ‘recall to reject’ strategy. According to this account, items eliciting only very weak evidence of auditory study elicited the effortful generation of the spoken name of the test item – and hence activation of auditorily responsive cortex – which went on to be rejected as a match with the item’s study condition.

The present findings clearly demonstrate that cortical reinstatement effects, as operationalized by univariate modulation of the BOLD signal, differ in their magnitude according to the strength or quality of the information supporting accurate source memory judgments. Thus, the findings support prior proposals that reinstatement effects represent recollected content (e.g., Kahn et al., 2004; Rugg et al., 2008; Danker and Anderson, 2010; Staresina et al., 2012; Kuhl et al., 2013; Rugg and Vilberg, 2013;). The findings do not, however, shed light on more fine-grained questions about the functional significance of reinstatement effects. For example, it remains uncertain whether the effects reflect the retrieval process itself as opposed to the downstream ‘post-retrieval’ representation of the retrieved information (for prior discussion of this possibility, see for example, Kahn et al., 2004; Woodruff et al., 2005; Johnson and Rugg, 2007). It is also unclear whether the same regions that demonstrate reinstatement effects participate in the maintenance of the recollected information over time, or whether maintenance depends upon the ‘transfer’ of the retrieved information to other brain regions (cf., Baddeley, 2000; Vilberg and Rugg, 2012, 2014). Addressing these and related questions will benefit both from further research with fMRI, and also from the employment of methodologies with markedly greater temporal resolution (e.g. Johnson et al., 2008; Wimber et al., 2012).

## Acknowledgments

This research was supported by NIMH Grant 5R01MH072966.

## References

- Alvarez P, Squire LR. Memory consolidation the medial temporal lobe: a simple network model. *Proc Natl Acad Sci USA*. 1994; 91:7041–7045. [PubMed: 8041742]
- Baddeley A. The episodic buffer: A new component of working memory? *Trends Cogn Sci*. 2000; 4:417–423. [PubMed: 11058819]
- Bradley JV. A nonparametric test for interactions of any order. *J Qual Technol*. 1979; 11:177–184.
- Chen J, Olsen RK, Preston AR, Glover GH. Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1. *Learn Mem*. 2011; 18:523–528. [PubMed: 21775513]
- Cocosco, CA., Kollokian, V., Kwan, RS., Evans, AC. BrainWeb: Online interface to a 3D MRI simulated brain database; Proceedings of the third international conference on function mapping of the human brain; Copenhagen, Denmark. 1997. p. S235
- Danker JF, Anderson JR. The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychol Bull*. 2010; 136:87–102. [PubMed: 20063927]
- Daselaar SM, Fleck MS, Cabeza R. Triple dissociation in the medial temporal lobes: Recollection familiarity and novelty. *J Neurophysiol*. 2006; 96:1902–1911. [PubMed: 16738210]
- Diana RA, Yonelinas AP, Ranganath C. Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends Cogn Sci*. 2007; 11:379–386. [PubMed: 17707683]
- Diana RA, Yonelinas AP, Ranganath C. Medial temporal lobe activity during source retrieval reflects information type not memory strength. *J Cogn Neurosci*. 2009; 22:1808–1818.
- Duarte A, Henson RN, Graham KS. Stimulus content and the neural correlates of source memory. *Brain Res*. 2011; 1373:110–123. [PubMed: 21145314]
- Düzel E, Habib R, Rotte M, Guderian S, Tulving E, Heinze H-J. Human hippocampal and parahippocampal activity during visual associative recognition memory for spatial and nonspatial stimulus configurations. *J Neurosci*. 2003; 23:9439–9444. [PubMed: 14561873]
- Fisher, RA. *Statistical methods for research workers*. London: Oliver and Boyd; 1950.
- Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R. Event-related fMRI: Characterizing differential responses. *Neuroimage*. 1998; 7:30–40. [PubMed: 9500830]
- Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G, Ashburner J. Classical and Bayesian inference in neuroimaging: Theory. *Neuroimage*. 2002; 16:465–483. [PubMed: 12030832]
- Glanzer M, Hilford A, Kim K. Six regularities of source recognition. *J Exp Psychol Learn Mem Cogn*. 2004; 30:1176–1195. [PubMed: 15521797]
- Gordon MA, Rissman J, Khan R, Wagner AD. Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cereb Cortex*. in press.
- Gottlieb LJ, Uncapher MR, Rugg MD. Dissociation of the neural correlates of visual and auditory contextual encoding. *Neuropsychologia*. 2010; 48:137–144. [PubMed: 19720071]
- Hayama HR, Vilberg KL, Rugg MD. Overlap between the neural correlates of cued recall and source memory: Evidence for a generic recollection network? *J Cogn Neurosci*. 2012; 24:1127–1137. [PubMed: 22288393]
- Hofstetter C, Achaïbou A, Vuilleumier P. Reactivation of visual cortex during memory retrieval: Content specificity and emotional modulation. *Neuroimage*. 2012; 60:1734–1745. [PubMed: 22305991]
- Huijbers W, Pennartz CMA, Rubin DC, Daselaar SM. Imagery and retrieval of auditory and visual information: Neural correlates of successful and unsuccessful performance. *Neuropsychologia*. 2011; 49:1730–1740. [PubMed: 21396384]
- Johnson JD, McDuff SG, Rugg MD, Norman KA. Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*. 2009; 63:697–708. [PubMed: 19755111]
- Johnson JD, Minton BR, Rugg MD. Content dependence of the electrophysiological correlates of recollection. *Neuroimage*. 2008; 39:406–416. [PubMed: 17933555]
- Johnson JD, Rugg MD. Recollection the reinstatement of encoding-related cortical activity. *Cereb Cortex*. 2007; 17:2507–2515. [PubMed: 17204822]

- Johnson JD, Suzuki S, Rugg MD. Recollection familiarity and content-sensitivity in lateral parietal cortex: a high-resolution study. *Front Hum Neurosci.* 2013; 7:1–15. [PubMed: 23355817]
- Josephs O, Henson RNA. Event-related functional magnetic resonance imaging: Modeling, inference, and optimization. *Philos Trans R Soc Lond B Biol Sci.* 1999; 354:1215–1228. [PubMed: 10466147]
- Kahn I, Davachi L, Wagner AD. Functional-neuroanatomic correlates of recollection: Implications for models of recognition memory. *J Neurosci.* 2004; 24:4172–4180. [PubMed: 15115812]
- Kim H. Dissociating the roles of the default-mode dorsal and ventral networks in episodic memory retrieval. *Neuroimage.* 2010; 50:1648–1657. [PubMed: 20097295]
- Köhler S, Danckert S, Gati JS, Menon RS. Novelty responses to relational and non-relational information in the hippocampus and the parahippocampal region: A comparison based on event-related fMRI. *Hippocampus.* 2005; 15:763–774. [PubMed: 15999342]
- Kucera, H., Francis, WN. Computational analysis of present-day American English. Brown University Press; Providence RI: 1967.
- Kuhl BA, Johnson MK, Chun MM. Dissociable neural mechanisms for goal-directed versus incidental memory reactivation. *J Neurosci.* 2013; 33:16099–16109. [PubMed: 24107943]
- Kuhl BA, Rissman J, Chun MM, Wagner AD. Fidelity of neural activation reveals competition between memories. *Proc Natl Acad Sci USA.* 2011; 108:5903–5908. [PubMed: 21436044]
- Kumaran D, Maguire EA. Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus.* 2007; 17:735–748. [PubMed: 17598148]
- Lazar NA, Luna B, Sweeney JA, Eddy WF. Combining brains: A survey of methods for statistical pooling of information. *Neuroimage.* 2002; 16:538–550. [PubMed: 12030836]
- Loftus GR, Masson MEJ. Using confidence intervals in within-subject designs. *Psychon Bull Rev.* 1994; 1:476–490. [PubMed: 24203555]
- McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev.* 1995; 102:419–457. [PubMed: 7624455]
- McDuff SGR, Frankel HC, Norman KA. Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. *J Neurosci.* 2009; 29:508–516. [PubMed: 19144851]
- Mickes L, Wais PE, Wixted JT. Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychol Sci.* 2009; 20:509–515. [PubMed: 19320859]
- Mitchell KJ, Johnson MK. Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychol Bull.* 2009; 135:638–677. [PubMed: 19586165]
- Morris CD, Bransford JD, Franks JJ. Levels of processing versus transfer appropriate processing. *J Verbal Learn Verbal Behav.* 1977; 16:519–533.
- Norman KA, O'Reilly RC. Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychol Rev.* 2003; 110:611–646. [PubMed: 14599236]
- Nyberg L. Any novelty in hippocampal formation and memory? *Curr Opin Neurol.* 2005; 18:424–428. [PubMed: 16003119]
- Nyberg L, Habib R, McIntosh AR, Tulving E. Reactivation of encoding-related brain activity during memory retrieval. *Proc Natl Acad Sci USA.* 2000; 97:11120–11124. [PubMed: 11005878]
- Parks CM, Yonelinas AP. Moving beyond pure signal-detection models: Comment on Wixted 2007. *Psychol Rev.* 2007; 114:188–202. [PubMed: 17227187]
- Qin J, Raye CL, Johnson MK, Mitchell KJ. Source memory ROCs are typically curvilinear: Comment on Yonelinas. *J Exp Psychol Learn Mem Cogn.* 2001; 27:1110–1115. 1999. [PubMed: 11486922]
- Ritchey M, Wing EA, LaBar KS, Cabeza R. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cereb Cortex.* 2013; 23:2818–2128. [PubMed: 22967731]
- Roediger HL, Gallo DA, Geraci L. Processing approaches to cognition: The impetus from the levels of processing framework. *Memory.* 2002; 10:319–332. [PubMed: 12396644]

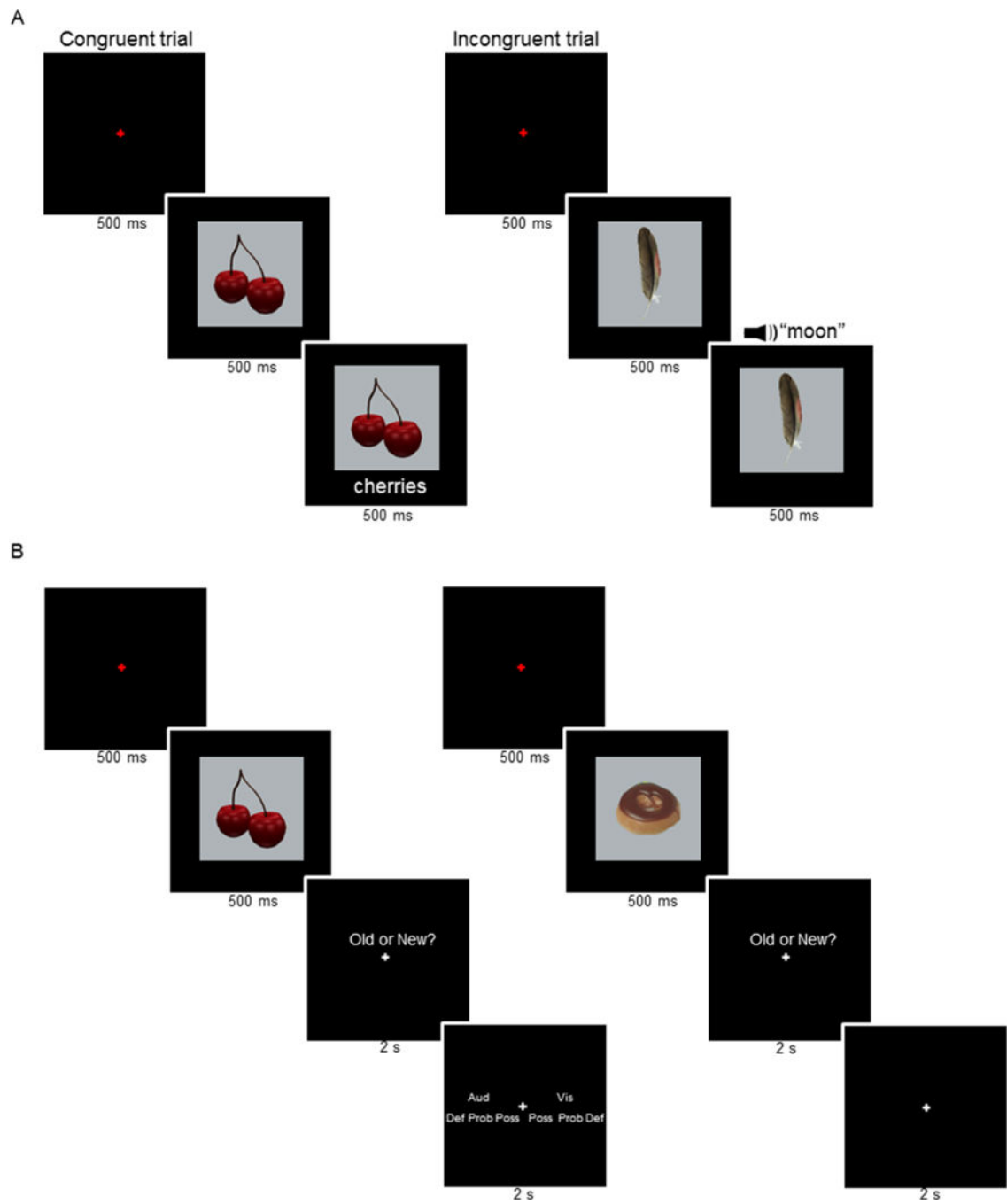


- Rolls ET. Memory systems in the brain. *Annu Rev Psychol.* 2000; 51:599–630. [PubMed: 10751982]
- Ross RS, Slotnick SD. The hippocampus is preferentially associated with memory for spatial context. *J Cogn Neurosci.* 2008; 20:432–446. [PubMed: 18004952]
- Rugg MD, Johnson JD, Park H, Uncapher MR. Encoding-retrieval overlap in human episodic memory: A functional neuroimaging perspective. *Prog Brain Res.* 2008; 169:339–352. [PubMed: 18394485]
- Rugg MD, Vilberg KL. Brain networks underlying episodic memory retrieval. *Curr Opin Neurobiol.* 2013; 23:255–260. [PubMed: 23206590]
- Rugg MD, Vilberg KL, Mattson JT, Yu SS, Johnson JD, Suzuki M. Item memory context memory and the hippocampus: fMRI evidence. *Neuropsychologia.* 2012; 50:3070–3079. [PubMed: 22732490]
- Sawilowsky SS. Nonparametric tests of interaction in experimental design. *Rev Educ Res.* 1990; 60:91–126.
- Shastri L. Episodic memory and cortico-hippocampal interactions. *Trends Cogn Sci.* 2002; 6:162–168. [PubMed: 11912039]
- Slotnick SD. Rapid retinotopic reactivation during spatial memory. *Brain Res.* 2009; 1268:97–111. [PubMed: 19272364]
- Slotnick SD. “Remember” source memory ROCs indicate recollection is a continuous process. *Memory.* 2010; 18:27–39. [PubMed: 19937493]
- Slotnick SD, Dodson CS. Support for a continuous single-process model of recognition memory and source memory. *Mem Cogn.* 2005; 33:151–170.
- Slotnick SD, Klein SA, Dodson CS, Shimamura AP. An analysis of signal detection and threshold models of source memory. *J Exp Psychol Learn Mem Cogn.* 2000; 26:1499–1517. [PubMed: 11185779]
- Slotnick SD, Thakral PP. The hippocampus operates in a threshold manner during spatial source memory. *Neuroreport.* 2013; 24:265–269. [PubMed: 23407277]
- Staresina BP, Henson RNA, Kriegeskorte N, Alink A. Episodic reinstatement in the medial temporal lobe. *J Neurosci.* 2012; 32:18150–18156. [PubMed: 23238729]
- Squire LR, Wixted JT, Clark RE. Recognition memory and the medial temporal lobe: A new perspective. *Nat Rev Neurosci.* 2007; 8:872–883. [PubMed: 17948032]
- Suzuki M, Johnson JD, Rugg MD. Decrements in hippocampal activity with item repetition during continuous recognition: An fMRI study. *J Cogn Neurosci.* 2011; 23:1522–1532. [PubMed: 20617887]
- Tulving, E. *Elements of Episodic Memory.* Oxford University Press; New York: 1983.
- Van Essen DC. A population-average landmark- and surface-based PALS atlas of human cerebral cortex. *Neuroimage.* 2005; 28:635–662. [PubMed: 16172003]
- Vilberg KL, Rugg MD. Dissociation of the neural correlates of recognition memory according to familiarity, recollection, and amount of recollected information. *Neuropsychologia.* 2007; 45:2216–2225.
- Vilberg KL, Rugg MD. Memory retrieval and the parietal cortex: A review of evidence from a dual-process perspective. *Neuropsychologia.* 2008; 46:1787–1799.
- Vilberg KL, Rugg MD. The neural correlates of recollection: transient versus sustained FMRI effects. *J Neurosci.* 2012; 32:15679–15687. [PubMed: 23136408]
- Vilberg KL, Rugg MD. Temporal dissociations within the core recollection network. *Cogn Neurosci.* 2014; 5:77–84. [PubMed: 24283400]
- Wais PE, Squire LR, Wixted JT. In search of recollection and familiarity signals in the hippocampus. *J Cogn Neurosci.* 2009; 22:109–123.
- Wheeler ME, Petersen SE, Buckner RL. Memory’s echo: Vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci USA.* 2000; 97:11125–11129. [PubMed: 11005879]
- Wimber M, Maaß A, Staudigl T, Richardson-Klavehn A, Hanslmayr S. Rapid memory reactivation revealed by oscillatory entrainment. *Curr Biol.* 2012; 22:1482–1486. [PubMed: 22795695]
- Wixted JT. Dual-process theory and signal-detection theory of recognition memory. *Psychol Rev.* 2007; 114:152–176. [PubMed: 17227185]
- Wixted JT, Mickes L, Squire LR. Measuring recollection familiarity in the medial temporal lobe. *Hippocampus.* 2010; 20:1195–1205. [PubMed: 20848603]

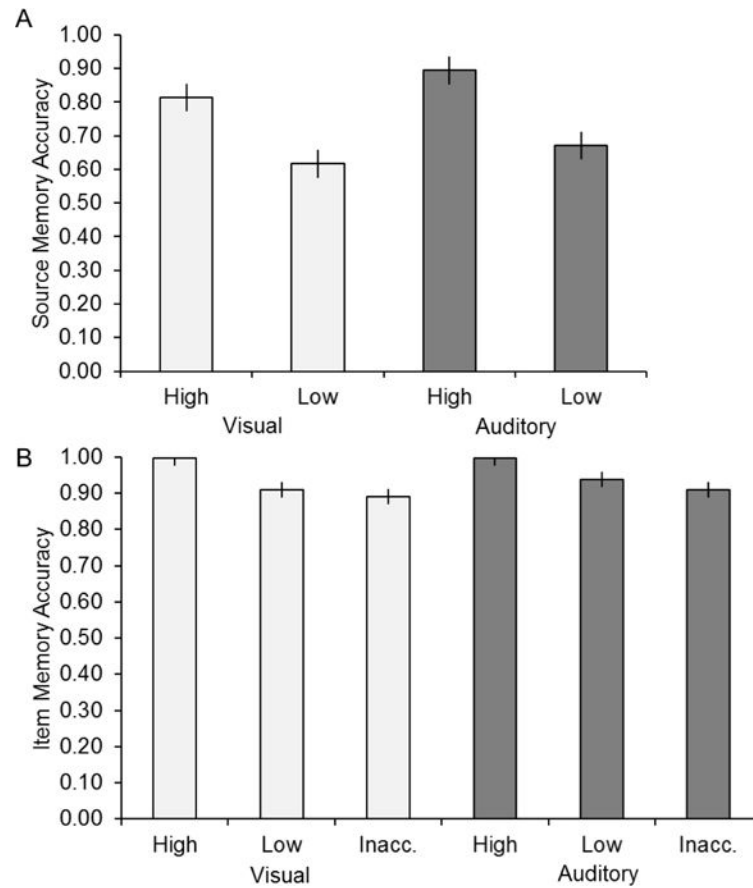
- Woodruff CC, Johnson JD, Uncapher MR, Rugg MD. Content-specificity of the neural correlates of recollection. *Neuropsychologia*. 2005; 43:1022–1032. [PubMed: 15769488]
- Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp*. 1996; 4:58–73. [PubMed: 20408186]
- Yu SS, Johnson JD, Rugg MD. Hippocampal activity during recognition memory co-varies with the accuracy confidence of source memory judgments. *Hippocampus*. 2012a; 22:1429–1437. [PubMed: 22076964]
- Yu SS, Johnson JD, Rugg MD. Dissociation of recollection-related neural activity in ventral lateral parietal cortex. *Cogn Neurosci*. 2012b; 3:142–149. [PubMed: 23049621]

### Highlights

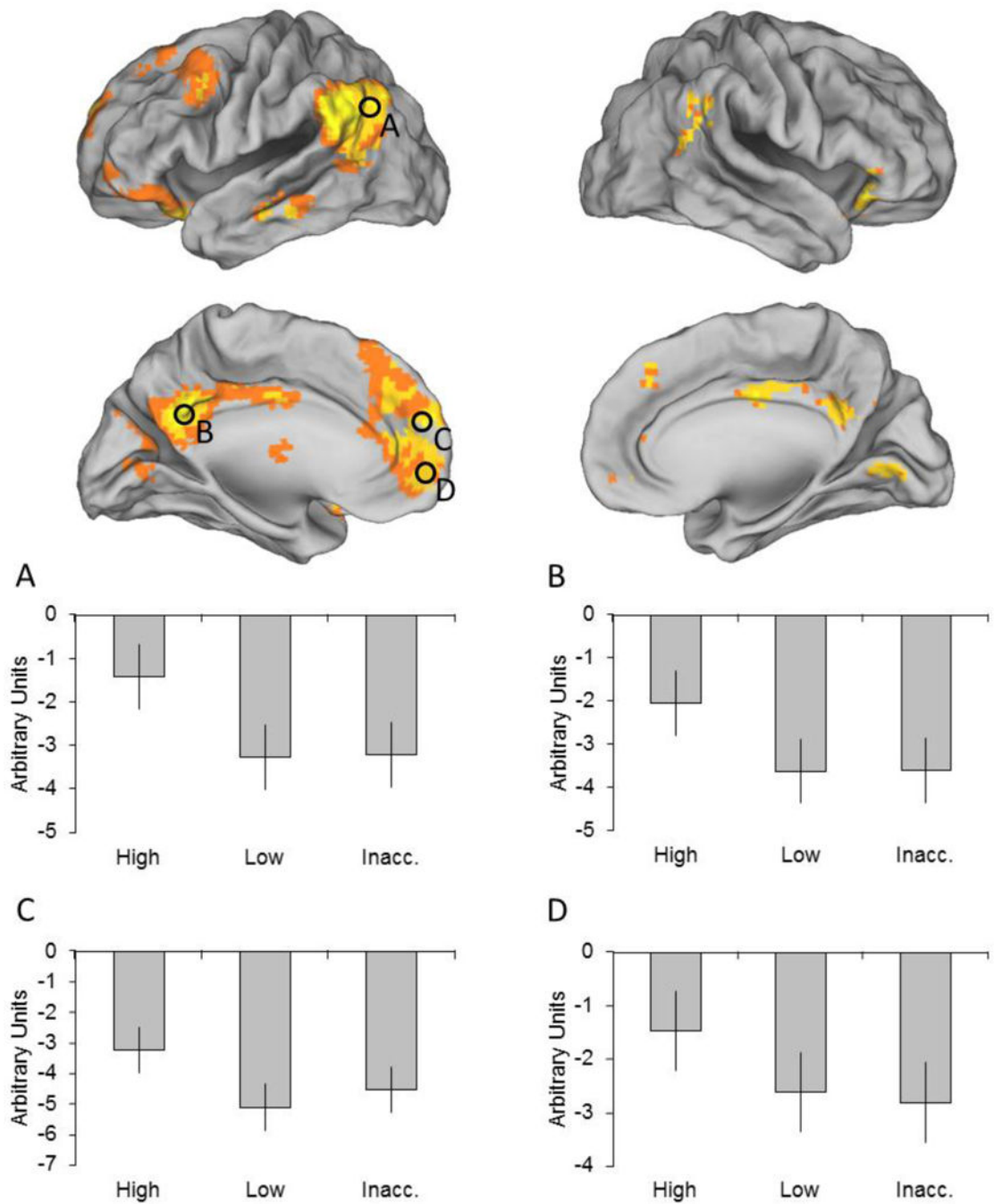
- ▶ Reinstatement refers to overlapping neural activity between encoding and retrieval
- ▶ We tested whether reinstatement co-varies with the confidence of source memory
- ▶ Reinstatement-related activity co-varied with the confidence of source memory
- ▶ Reinstatement strength indexes the quality of information supporting source memory



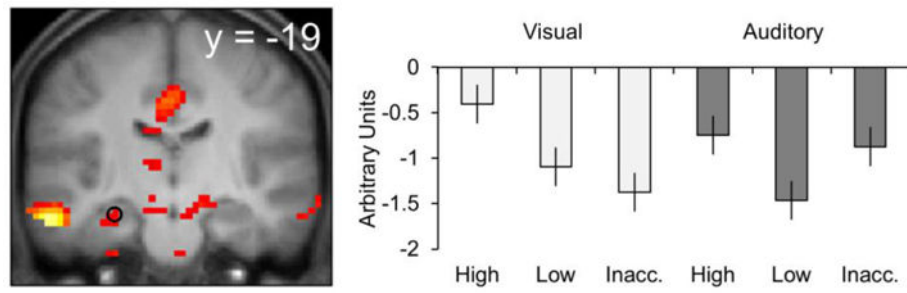
**Figure 1.**  
 A. Study task (representative congruent and incongruent trials are shown to the left and right, respectively). B. Test task. Durations are shown beneath each frame.



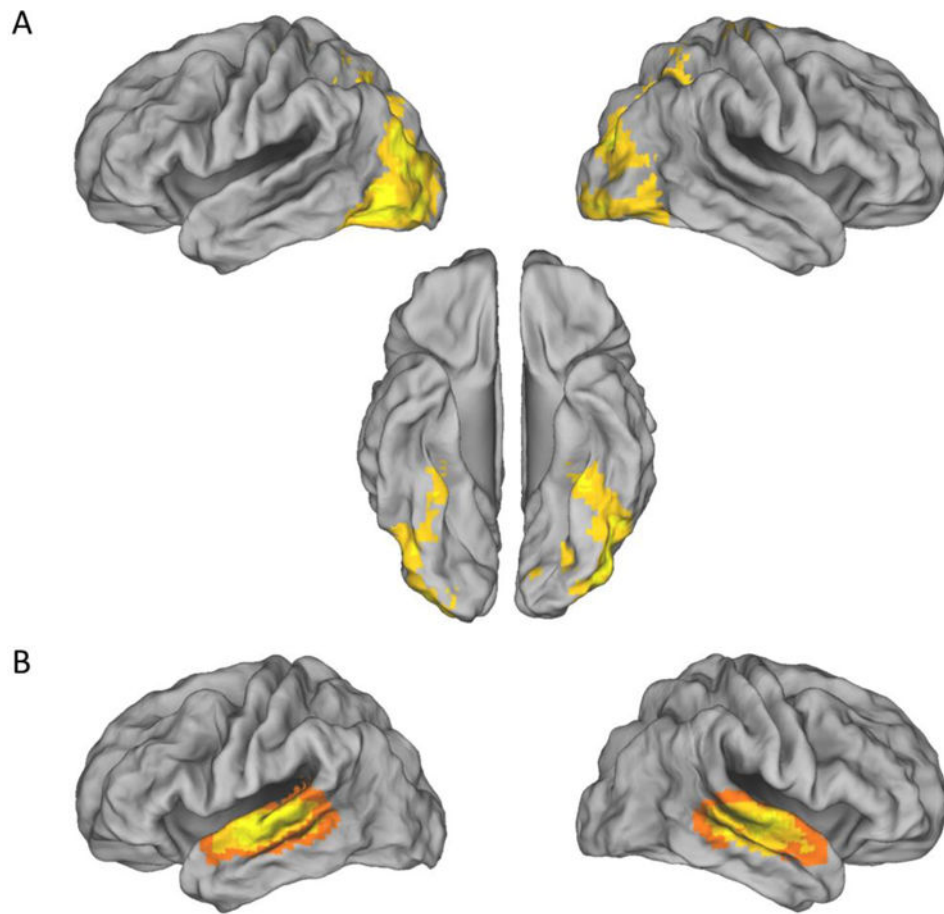
**Figure 2.** Source and item memory accuracy. A. The proportion of old pictures from each study condition that attracted accurate high and low confidence source judgments (source accuracy conditionalized on accurate item memory). B. Item accuracy ( $p_{Hit}/(p_{Hit}+p_{False\ Alarm})$ ) as a function of response category. Error bars in this and subsequent figures signify the standard error derived from the error term of the one-way ANOVA (see Results and Loftus and Masson, 1994).



**Figure 3.** Modality-independent source memory effects. Results are overlaid on the standardized brain of the PALS-B12 atlas implemented in Caret5 (VanEssen, 2005). In this and subsequent figures, regions where parameter estimates were extracted from are denoted by black circles. Mean across-modality parameter estimates are shown for each response category and region (A. left angular gyrus, B. left posterior cingulate/retrosplenial cortex, C. left medial prefrontal cortex, and D. left ventromedial prefrontal cortex).

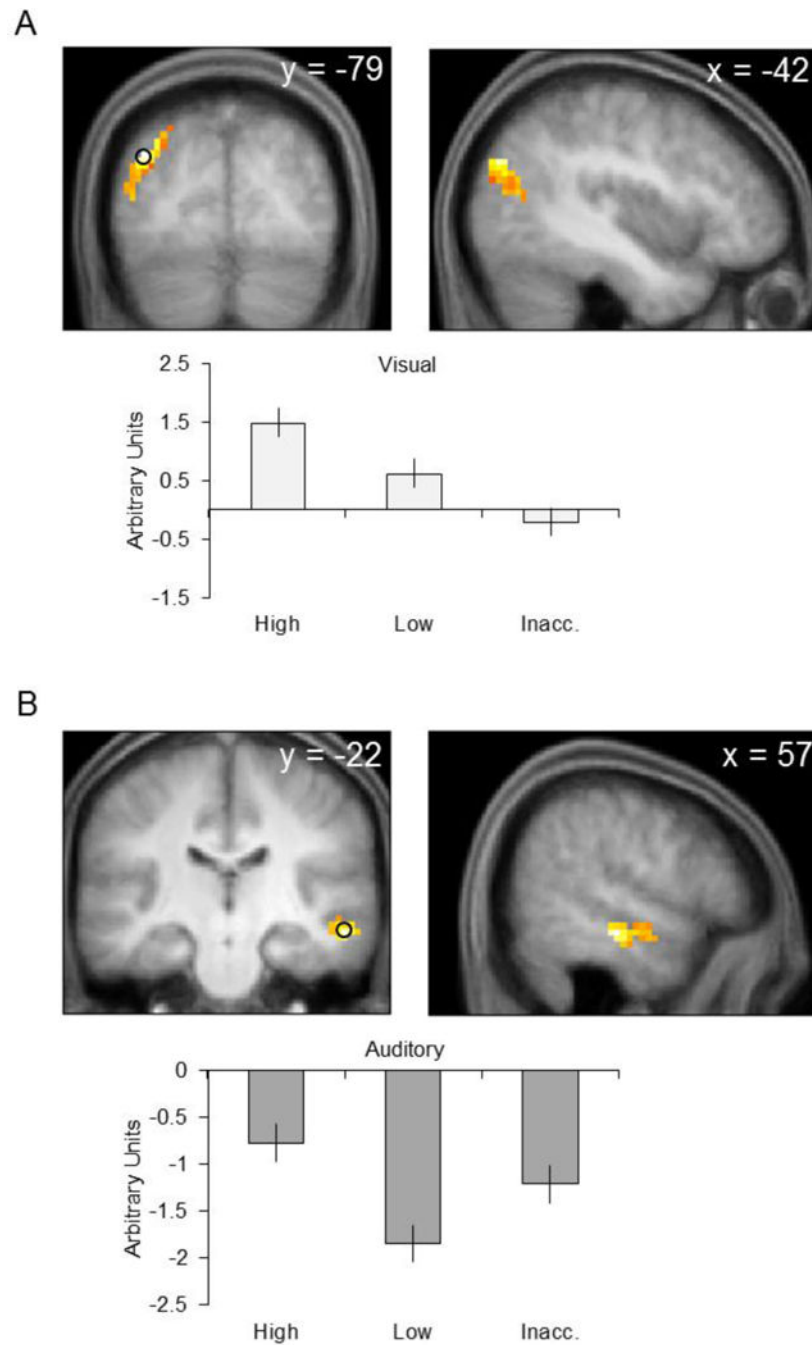


**Figure 4.** Hippocampal source memory effects. Results are overlaid onto a coronal section of the across-participants mean T1-weighted anatomical image (in this and subsequent figures, the mean image is derived from only 16 of the 17 participants). Parameter estimates are shown for each response category.



**Figure 5.** Modality-selective effects during study. Regions where activity elicited on visual study trials exceeded that on auditory trials (A), and vice-versa (B).





**Figure 6.** Source reinstatement effects for each modality. A. Effects associated with the visual condition. B. Effects associated with the auditory condition. In each case, parameter estimates from the peak voxel are illustrated. Results are overlaid onto the coronal and sagittal sections of the across-participants mean T1-weighted anatomical image.

**Table 1**

Mean ( $\pm$  standard deviation) response proportions for old pictures from each study condition for each response category

	<b>High</b>	<b>Low</b>	<b>Inaccurate</b>
Visual	0.20 (0.13)	0.35 (0.11)	0.27 (0.12)
Auditory	0.24 (0.10)	0.39 (0.13)	0.22 (0.09)

The values for each study condition do not sum to 1 as the proportions of misses (i.e., new responses to old pictures) are not included.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

## Loci of source memory effects

MNI Coordinates			Peak Z	Number of above-threshold voxels	Region
x	y	z			
<i>Modality-independent source memory effects</i>					
-57	-52	40	5.83	700	L angular/supramarginal gyrus
-54	-67	25	5.63		L superior temporal gyrus
60	-49	34	5.32	149	R angular gyrus
66	-46	25	3.83		R superior temporal gyrus
-9	-52	34	5.63	741	L posterior cingulate/retrosplenial cortex
-3	-13	37	4.51		L middle cingulate cortex
-57	-22	-17	6.15	149	L inferior temporal sulcus
-69	-31	-5	3.39		L middle temporal gyrus
-6	-10	7	3.97	26	L thalamus
-27	17	-23	5.49	362	L orbitofrontal cortex
45	26	-20	4.48	132	R orbitofrontal cortex
-6	50	10	5.35	1015	L medial prefrontal cortex
-9	65	1	5.21		L ventromedial prefrontal cortex
<i>Modality-selective source memory effects for the visual condition</i>					
-39	-79	34	4.15	86	L middle occipital gyrus
-30	-85	37	3.83		L superior occipital gyrus
<i>Modality-selective source memory effects for the auditory condition</i>					
57	-22	-11	3.74	106	R superior temporal sulcus

Coordinates for cluster sub-peaks which lied in distinct cortical regions are listed directly below relevant peak cluster.