

Dissociating Source Memory Decisions in the Prefrontal Cortex: fMRI of Diagnostic and Disqualifying Monitoring

David A. Gallo¹, Ian M. McDonough¹, and Jason Scimeca²

Abstract

■ We used event-related fMRI to study two types of retrieval monitoring that regulate episodic memory accuracy: diagnostic and disqualifying monitoring. Diagnostic monitoring relies on expectations, whereby the failure to retrieve expected recollections prevents source memory misattributions (sometimes called *the distinctiveness heuristic*). Disqualifying monitoring relies on corroborative evidence, whereby the successful recollection of accurate source information prevents misattribution to an alternative source (sometimes called *recall to reject*). Using criterial recollection tests, we found that orienting retrieval toward distinctive recollections (colored pictures) reduced source memory misattributions compared with a control test in

which retrieval was oriented toward less distinctive recollections (colored font). However, the corresponding neural activity depended on the type of monitoring engaged on these tests. Rejecting items based on the absence of picture recollections (i.e., the distinctiveness heuristic) decreased activity in dorsolateral prefrontal cortex relative to the control test, whereas rejecting items based on successful picture recollections (i.e., a recall-to-reject strategy) increased activity in dorsolateral prefrontal cortex. There also was some evidence that these effects were differentially lateralized. This study provides the first neuroimaging comparison of these two recollection-based monitoring processes and advances theories of prefrontal involvement in memory retrieval. ■

INTRODUCTION

Episodic memory accuracy is determined by the quality of the originally encoded information as well as the effectiveness of the monitoring processes engaged during retrieval. Retrieval monitoring is a general concept, referring to the various search and decision processes people use when reconstructing events from memory. According to the influential source-monitoring framework, numerous factors influence retrieval monitoring (Johnson, 2006). These factors include the types and amounts of retrieved information, retrieval expectations, and relationships between past events. Retrieval monitoring is a high-level cognitive process, and evidence from patient and neuroimaging studies indicates that prefrontal brain regions are critically involved (for reviews, see Pannu & Kaszniak, 2005; Fletcher & Henson, 2001).

Dozens of fMRI studies have found that regions within dorsolateral prefrontal cortex (DLPFC), including bilateral regions along the midfrontal gyrus such as Brodmann's areas (BA) 9, 46, and sometimes 8, are more active for complex or effortful memory decisions that likely require retrieval monitoring (Achim & Lepage, 2005; Velanova et al., 2003; Wheeler & Buckner, 2003; Cansino, Maquet, Dolan, & Rugg, 2002; Cabeza, Rao, Wagner, Mayer, & Schacter, 2001; McDermott, Jones, Petersen, Lageman, & Roediger, 2000;

Henson, Shallice, & Dolan, 1999). Moreover, activity within these DLPFC regions is not specific to memory or to the materials used, suggesting that these regions support domain-general monitoring or decisions (e.g., Dobbins & Han, 2006; Fleck, Daselaar, Dobbins, & Cabeza, 2005).

Some theories maintain that different types of retrieval monitoring are subserved by different prefrontal regions. For example, it has been argued that left prefrontal regions subservise the use of systematic source monitoring processes or strategically using specific recollections or contextual information to make a decision. In contrast, it has been argued that right prefrontal regions subservise heuristic source monitoring processes or the use of less differentiated information or more vague feelings of familiarity to make a decision (Nolde, Johnson, & Raye, 1989; see also Mitchell et al., 2008; Dudukovic & Wagner, 2007; Dobbins, Simons, & Schacter, 2004; Mitchell, Johnson, Raye, & Greene, 2004; Ranganath, 2004). A different account emphasizes right DLPFC in postretrieval monitoring or the engagement of search and decision processes after an initial retrieval attempt yields insufficient information (Lepage, 2004; Henson et al., 1999; for a review, see Rugg, 2004). To the extent that systematic processes involve postretrieval monitoring, this emphasis on right DLPFC conflicts with the idea that systematic processes depend on left DLPFC.

An important aspect in many of these earlier neuroimaging studies is that they compared conditions that likely

¹University of Chicago, ²Northwestern University

relied on recollection and familiarity to different degrees. Whereas recollection refers to the conscious recall of information that was previously associated with a test item, familiarity refers to a decontextualized feeling that the test item was earlier encountered (see Yonelinas, 2002). The main evidence for prefrontal lateralization within the systematic/heuristic distinction comes from comparisons of source memory tests, which engage recollection, and item recognition or recency tests, which can rely more heavily on familiarity (Mitchell et al., 2004, 2008; Rugg, Fletcher, Chua, & Dolan, 1999; Nolde, Johnson, & D'Esposito, 1998). As a result of these comparisons, these lateralization effects may have been caused by the differential retrieval of recollection and familiarity across tests (Wheeler & Buckner, 2004; Kensinger, Clarke, & Corkin, 2003), the different decision processes that might correspond to recollection and familiarity (Yonelinas, 2002), or other monitoring processes that may contribute to source memory tests.

An fMRI study by Dobbins and Han (2006) attempted to disentangle some of these factors, holding test items constant but varying the decision to be made on these items. They found that bilateral DLPFC and frontopolar activity increased with the number of subprocesses required to make the test decision, but not with the level of supporting evidence in memory (also see Hayama, Johnson, & Rugg, 2008; Rajah, Ames, & D'Esposito, 2008). The exact nature of the decision rules used in these tasks as well as the degree that they may have relied on recollection and familiarity is unclear. Nevertheless, these studies raise the intriguing possibility that DLPFC regions are mostly sensitive to the nature and complexity of the decision rules used during memory tests, as opposed to the actual information retrieved. Building on this idea, the current experiment investigated two well-characterized recollection-based monitoring processes.

Diagnostic and Disqualifying Monitoring

Diagnostic and disqualifying monitoring refer to two types of recollection-based retrieval monitoring that reduce false recognition and source misattributions across a variety of memory tasks (Gallo, 2004; for a review see Gallo, 2006). This dichotomy is intended to highlight qualitative differences in the logic of the underlying decision process. *Diagnostic monitoring* involves searching memory for recollections that are expected to be characteristic of a particular type of event, context, or target source (e.g., studying items as words or as pictures). If the test item passes these recollective criteria, then it is attributed to the target source; otherwise, it is rejected as having occurred in that source. This process underlies many source monitoring biases, such as the "it-had-to-be-you" effect (e.g., Marsh & Hicks, 1998; Johnson, Raye, Foley, & Foley, 1981) as well as the distinctiveness heuristic that has been demonstrated in false recognition tasks (e.g., Ghetti, 2003; Dodson, Koutstaal, & Schacter,

2000; Schacter, Israel, & Racine, 1999). A subject rejecting a test item based on the distinctiveness heuristic might reason, "This item is familiar, but I don't have a distinctive recollection. Because the target source elicits distinctive recollections, it probably wasn't studied in that source." Note that diagnostic monitoring does not always rely on distinctive features, but expecting distinctive features can make the monitoring process more accurate, thereby reducing memory distortion.

Disqualifying monitoring involves searching memory for recollections that are not diagnostic of the target source but nevertheless can strategically inform or corroborate whether the item had occurred in that target source. This process often occurs when the participant believes that the different types of studied items or sources are mutually exclusive. In these situations, if the test item is recollected as having occurred in the nontarget source, then it is rejected as having occurred in the target source. This process underlies many source-based exclusion processes (e.g., Jacoby, Jones, & Dolan, 1998) as well as many recall-to-reject processes that have been found in false recognition tasks (e.g., Lampinen, Odegard, & Neuschatz, 2004; Brainerd, Reyna, Wright, & Mojardin, 2003; Rotello & Heit, 2000; Hintzman & Curran, 1994). A subject using a recall-to-reject strategy might reason, "This item is familiar, but I recollect that it was studied in the nontarget source. Because the sources were mutually exclusive, it couldn't have been studied in the target source." Disqualifying monitoring can occur for multiple types of information, including the source of an item as well as other features associated with an item, as long as the studied events are structured in such a way that the retrieval of one type of information logically excludes another from having occurred.

Gallo, Cotel, Moore, and Schacter (2007) demonstrated both diagnostic and disqualifying monitoring using the criterial recollection task. Subjects studied a list of common object labels in black font. Each label was paired with the same word in red font or with a colored picture of the corresponding object. Memory then was tested using black labels as retrieval cues. On the red word test, subjects responded "yes" if they recollected a red word, whereas on the picture test, they responded "yes" if they recollected a picture. Importantly, some test items had been paired with both a red word and a picture at study. Because these study formats were not mutually exclusive, noncriterial recollections were relatively uninformative (e.g., picture items on the red word test). Subjects instead had to selectively search their memory for the to-be-recollected (or criterial) information (e.g., red word items on the red word test) using diagnostic monitoring processes. Under these conditions, source memory confusions were lower on the picture test than on the red word test. Subjects expected more distinctive recollections on the picture test, and these retrieval expectations enhanced diagnostic monitoring accuracy (the distinctiveness heuristic). Gallo et al. also included

a condition where the study formats were mutually exclusive, thereby allowing the use of a disqualifying monitoring process (e.g., rejecting items on the red word test by recollecting a picture). As predicted, source memory confusions were further reduced in this exclusion condition.

A handful of fMRI studies are relevant to these specific monitoring processes. Gallo, Kensinger, and Schacter (2006) used the nonexclusive conditions of the criterial recollection task to investigate the neural correlates of diagnostic monitoring in the absence of a recall-to-reject strategy. They found that studied items on the red word test were more likely to activate regions in bilateral DLPFC relative to these same items on the picture test. Because red word recollections were relatively less distinctive than picture recollections, it was argued that subjects needed to engage more effortful retrieval monitoring on the red word test, thereby increasing activity in DLPFC (for related ERP results, see Budson et al., 2005; for related patient results, see Hwang et al., 2007). Said differently, the DLPFC was less likely to be recruited when subjects monitored memory for distinctive picture recollections (i.e., the distinctiveness heuristic). Woodruff, Uncapher, and Rugg (2006) also found greater activity in bilateral DLPFC when subjects were tested for words compared with pictures. These effects were sustained across the retrieval blocks, potentially reflecting the adoption of different retrieval orientations (analogous to a distinctiveness heuristic). However, the study formats were mutually exclusive in this study, potentially implicating a recall-to-reject strategy as well.

A few fMRI studies are more directly relevant to disqualifying monitoring. Rugg, Henson, and Robb (2003) and Henson et al. (1999) both found elevated bilateral DLPFC activity when the test required a source-based exclusion strategy compared with item recognition tests in which all studied items were to be accepted (for relevant ERP findings, see Fraser, Brisdon, & Wilding, 2007; Herron & Rugg, 2003). These results suggest that DLPFC is recruited when subjects use a recall-to-reject process, potentially reflecting the application of a rule-based rejection strategy. However, these differences may have reflected a differential reliance on recollection and familiarity across the tests, as opposed to the monitoring process itself. Using a conjunction-word task, McDermott et al. (2000) found greater activity in bilateral DLPFC when subjects rejected lures that could have benefited from a recall-to-reject strategy, relative to control lures or studied targets (for analogous results in an associative-recognition task, see Achim & Lepage, 2005). Unlike the targets, though, each lure corresponded to the rearrangement of two different studied items in these tasks, so these activity differences may have been due to differences in retrieval success as opposed to monitoring processes.

To summarize, only a handful of neuroimaging studies have investigated different types of retrieval monitoring. With respect to disqualifying monitoring, fMRI studies have shown elevated DLPFC activity (as well as

other PFC regions) when subjects reject items based on a recall-to-reject strategy. These findings are consistent with the idea that recollection-based retrieval monitoring is a type of cognitive control that relies on DLPFC. In contrast, only one fMRI study has attempted to isolate diagnostic monitoring (Gallo et al., 2006), and this study found decreased DLPFC activity associated with the use of a distinctiveness heuristic. To the extent that the distinctiveness heuristic is considered a recollection-based monitoring process or even a metacognitive strategy that can be turned “on” or “off” (cf. Schacter & Wiseman, 2006; Dodson et al., 2000), the finding that this process reduced DLPFC activity may seem at odds with the idea that recollection-based monitoring depends on DLPFC. However, if the distinctiveness heuristic is instead conceptualized as one instance of a more general diagnostic monitoring process, a process that occurs whenever one searches memory for recollections, these findings can be reconciled with prior research. From this perspective, diagnostic monitoring is critically dependent on the DLPFC, but the degree to which diagnostic monitoring is recruited on any test depends on the relative distinctiveness of the to-be-retrieved information. By definition, distinctive representations are more vivid and more easily discriminated in memory, requiring less effortful diagnostic monitoring and leading to less recruitment of DLPFC.

Current Experiment

For the current experiment, we modified the task used by Gallo et al. (2007) to directly compare diagnostic and disqualifying monitoring. No prior fMRI study has compared these two monitoring processes, but doing so provides a more direct test of the theoretical framework described above. Prior work investigating these processes has relied on different types of tasks and different types of to-be-remembered materials. Thus, the degree that the different patterns of neural activity observed reflects the use of two different retrieval monitoring processes, as opposed to other differences between the experiments, cannot be ascertained. The current task avoids these issues by using the same type of to-be-remembered materials (i.e., pictures) for each of the recollection-based monitoring processes (i.e., the distinctiveness heuristic and the recall-to-reject processes) and also by controlling for possible familiarity confounds. By comparing these processes in a single task, the current experiment also afforded a test of the functional lateralization ideas that have been proposed within the source monitoring framework and the postretrieval monitoring theory, as described below.

To investigate diagnostic monitoring, we compared the red word test and picture test under nonexclusive conditions. These conditions required subjects to selectively recollect a single source of information, differing only in the relative distinctiveness of the to-be-recollected

information (pictures > red words). Based on Gallo et al. (2006), we expected that DLPFC would be more active on the red word test relative to the picture test, indicating more effortful diagnostic monitoring. This earlier study found activity in bilateral DLPFC, but unlike that study, we more precisely equated familiarity across the stimuli in the current study. Thus, any observed laterality effects would be more interpretable in terms of the underlying recollection-based monitoring process. According to the postretrieval monitoring theory, effortful diagnostic monitoring situations should recruit right DLPFC more than left DLPFC (consistent with the earlier finding). Predictions based on the systematic/heuristic distinction are less clear in this case because this dichotomy is more general and descriptive. The simplest prediction is that the picture test should recruit right DLPFC more than the red word test because the picture test relies more heavily on heuristic-based responding (the distinctiveness heuristic), although other interpretations are possible (discussed later).

To investigate disqualifying monitoring, we compared the red word test to an exclusion test. Like the red word test, subjects responded “yes” to red words on the exclusion test. Unlike the red word test, we did not include test items that were studied in both formats so that subjects could use an exclusion-based rejection rule (i.e., rejecting test items that elicit picture recollections). We expected that DLPFC would be more active when subjects engaged in such a recall-to-reject process, but lateralization theories result in contrasting predictions. The systematic/heuristic distinction predicts that the exclusion strategy would elicit more activity in left DLPFC than in right DLPFC because this is a more systematic rule-based process. In contrast, the postretrieval monitoring perspective predicts greater activity in right DLPFC. Only the exclusion test required subjects to monitor memory for multiple types of recollections, thereby requiring the most postretrieval monitoring.

METHODS

Subjects

For the imaging experiment, 27 students at the University of Chicago participated for \$40 (18–36 years, 16 women, all right-handed and fluent in English). Behavioral and imaging data from six subjects were excluded, owing to excessive head movement or other technicalities or an insufficient number of behavioral responses. Before the imaging study, 18 subjects participated in a behavioral pilot (not reported). For the manipulation check experiment, 13 students participated for course credit or \$10, with data from one student replaced for failure to pay attention at encoding. MRI subjects were prescreened using standard safety procedures, and all subjects gave informed consent.

Materials

Stimuli were drawn from Gallo et al. (2006) and included 360 colored pictures of common objects (e.g., *lemon*, *toaster*) on a white background and corresponding verbal labels in red font. Each study trial began with a black word (500 msec), immediately followed (100 msec) by the same word in larger red letters (1200 msec) or the corresponding picture (1200 msec), with a 150-msec ISI. At test, a 6-sec prompt cued the retrieval demands of the upcoming test block (Red Word, Picture, or Exclusion). Verbal labels (black font on white background) were used as retrieval cues, along with a test prompt to keep subjects on task (“red word?” for the Red Word test, “picture?” for the Picture test, and “red word only?” for the Exclusion test). Twelve scripts were created to counterbalance the stimuli, across subjects, through the item conditions (studied as red word, picture, both, or non-studied) and test conditions (Red Word, Picture, Exclusion). After the initial counterbalancing was met, the remaining subjects were arbitrarily assigned. For ease of viewing, stimuli were projected onto a mirror above the head coil.

Criteria Recollection Procedure

Subjects first completed a practice version of the task, using separate stimuli, to ensure that they understood the instructions (approximately 6 min). For the main experiment, the study phase occurred outside the scanner (approximately 25 min). Subjects studied 240 red words and pictures for the upcoming tests (90 red words, 90 pictures, and 60 both items, presented as both red words and pictures, nonconsecutively). Each red word and picture was presented twice, nonconsecutively (including those corresponding to both items). Subjects were instructed to make a semantic judgment for each red word (“Is this item made in a factory?”) and focus on the perceptual features of each picture. To avoid carryover effects of these orienting tasks, subjects studied alternating blocks of red words and pictures, with stimuli randomized within each block and order counterbalanced across subjects. Pilot work indicated that these study presentation procedures would equate the familiarity of the test words across the red word and picture conditions while maintaining differences in recollective distinctiveness (pictures > red words).

The test phase occurred in the scanner, following approximately 20 min of preparation. The tests were divided into three runs of functional scans (approximately 10 min per run). Each run was subdivided into three test blocks, corresponding to each of the three types of test, separating each block by 21 sec of fixation. Test block order was varied across runs and counterbalanced across subjects. During each test block, subjects saw 10 test words corresponding to each type of studied item (red word, picture, both, or new), except that the both items were replaced with 10 additional new items on the exclusion test blocks. Test

words were presented for 3 sec and separated by a central fixation cross of jittered duration (3, 6, or 9 sec, mean SOA = 3.83 sec). The order of items and fixations was arbitrarily mixed using a program to maximize the MR signal (e.g., Dale, 1999). In total, there were 30 items of each critical type (red words, pictures, new) on each of the three tests, with 30 filler items also included to manipulate the exclusion demands (30 both items on the Red Word and Picture tests, 30 additional new items on the Exclusion test).

Test responses were made with the index (“yes”) and middle (“no”) fingers of the right hand, whereas the test word was on the screen. On the Red Word test, subjects pressed “yes” if they remembered studying a corresponding red word (i.e., red word and both items) and “no” if not, regardless of whether they remembered a corresponding picture (i.e., picture and new items). On the Picture test, subjects pressed “yes” if they remembered studying a corresponding picture (i.e., picture and both items) and “no” if not, regardless of whether they remembered a corresponding red word (i.e., red word and new items). For these nonexclusion tests, it was emphasized that some items were studied in both formats so that the recollection of one format (e.g., a picture) did not preclude presentation in the other format (e.g., a red word). Instead, subjects were instructed to focus only on whether they could recollect the to-be-remembered format (a diagnostic monitoring process). On the Exclusion test, they were to press “yes” if they remembered studying a corresponding red word and “no” if not. It was emphasized that both items would not be included on this test so that red word and picture items were mutually exclusive. Thus, if they recollected a picture on the exclusion test, they could be sure that the item was not associated with a red word at study (a disqualifying process).

Manipulation Check Procedure

The manipulation check experiment was conducted independently from the fMRI study and used several procedures to measure recollection and familiarity. The study phase for the manipulation check experiment was identical to the fMRI experiments except there was no practice phase. The test phase immediately followed and was divided into three test blocks. Each block contained a subset of the red words, pictures, and new items. Both items were not relevant to the purpose of this study, but because we used the same materials as in the fMRI study, we arbitrarily included them on the first and last test blocks.

The first block contained a speeded recognition test, on which subjects responded “yes” to studied items and “no” to nonstudied items, independent of whether they recollected a red word or a picture. Responses were speeded using prompts to establish a tempo (Balota, Burgess, Cortese, & Adams, 2002), yielding an average response latency of 688 msec ($SD = 24.6$ msec) that

should primarily reflect familiarity-based recognition (cf. Yonelinas, 2002). The second block was a self-paced subjective test, on which subjects responded “actually recollect” if they remembered a red word or picture from study, “very familiar” if they thought the item was studied but could not recollect the format, and “new” if they thought the item was nonstudied. The third block also was a self-paced subjective test, on which subjects rated the level of strength and details that they could recollect for each test item (using a 0–7 scale). The recollection strength index was designed to capture overall differences in memory strength (ranging from “no recollection” to “strong or vivid recollection”). The recollection details index was designed to capture the amount of unique or distinctive details that could be recollected (ranging from “no details” to “many details”). We were primarily interested in recollective distinctiveness, but we had subjects make a separate “strength” and “details” judgment to help clarify the difference (see McDonough & Gallo, 2008).

Neuroimaging Procedure

Images were acquired using a 3-T GE Signa scanner at the University of Chicago Brain Research Imaging Center. Functional images were acquired using a T2*-weighted gradient-echo spiral in/out pulse sequence (repetition time = 3 sec, echo time = 28 msec, field of view = 240 mm; flip angle = 80 degrees, matrix size = 64 × 64, in-plane resolution = 3.75 mm). For whole-brain coverage, 30 interleaved sagittal slices (4.8-mm thickness, 0.5-mm skip) were acquired. Three event-related functional runs were used for the memory tests, followed by two functional localizer runs (not reported here), and an anatomical scan using a T1-weighted multiplanar rapidly acquired gradient-echo sequence. All preprocessing and data analysis were carried out using SPM5 (Wellcome Department of Cognitive Imaging Neuroscience, London), as implemented in MATLAB 7.4.0 (The MathWorks Inc., Natick, MA). Standard preprocessing was performed on the functional data, including slice-timing correction relative to the second (middle) slice, realignment using rigid body motion correction, anatomical coregistration, normalization to the MNI template (resampling at 3-mm cubic voxels), and spatial smoothing (using an 8-mm FWHM isotropic Gaussian kernel).

For each participant, an event-related analysis was first conducted on a voxel-by-voxel basis, in which all instances of each event type were modeled through convolution with a canonical hemodynamic response function locked to event onset times. Event types reflected a combination of the test condition (Red Word, Picture, Exclusion), the item type (both, red word, picture, new), and the participant’s response (yes, no). All participants had at least 10 instances of each event type included in the analyses. Modeling proceeded in two levels, an individual subject

analysis using the general linear model (including the canonical HRF and the temporal derivatives, a 128-sec high-pass filter, and session effects) followed by a pooled analysis with subjects as the random effect. The most significant voxel within a cluster is reported in Talairach coordinates, along with the approximate BA from the Talairach atlas (Talairach & Tournoux, 1998) and the Talairach Daemon (Lancaster et al., 2000).

BEHAVIORAL RESULTS

We first report the results from the manipulation check experiment, followed by the criterial recollection task used during the fMRI sessions. All behavioral comparisons were based on prior work, and unless otherwise specified, all results were considered significant at $p < .05$, two-tailed. Effect sizes for significant comparisons were calculated with Cohen's d .

Manipulation Check

The critical results from the manipulation check are presented in Table 1. Consistent with our efforts to equate familiarity across test items, hits to red word items (0.73) and picture items (0.71) were equated on the speeded test, $t(11) < 1$, whereas these items were recognized more often than new items (0.36), both p 's $< .001$. On the recollect/familiar test, we estimated familiarity using the independent-remember-know adjustment (see Yonelinas, 2002). Like the speeded test, the difference between red word items (0.55) and picture items (0.47) was not significant with this estimate of familiarity, $t(11) = 1.25$, $SEM = 0.063$, $p = .24$, but each was greater than new items (0.19), both p 's $< .01$. In contrast, the proportion of "actually recollect" judgments was significantly greater for picture items (0.65) than red word items (0.55), $t(11) = 2.98$, $SEM = 0.033$, $d = .58$, and so too were ratings of unique or distinctive recollective details, 3.19 versus 2.10, $t(11) = 2.45$, $SEM = 0.446$, $d = .34$. Overall, these results confirm that test words that were associated with pictures at study led to more distinctive recollections than those associated with red words, with no differences in familiarity.

Criterial Recollection

The results from the criterial recollection task are presented in Table 2 and replicate prior work using this task (e.g., Gallo et al., 2007). Consider the results from the non-exclusive conditions first. On the red word test, subjects were more likely to endorse red word items than picture items (0.62 vs. 0.41), $t(20) = 4.07$, $SEM = 0.052$, $d = 1.49$, whereas on the picture test, subjects were more likely to endorse picture items than red word items (0.61 vs. 0.19), $t(20) = 8.25$, $SEM = 0.051$, $d = 2.71$. This crossover pattern indicates that subjects had relied on red word recollections on the red word test and picture recollections on the picture test. Recollection was not perfect, though, and subjects also made source memory confusions to items that were studied in the noncriterial format. Picture items were more likely to be falsely recognized than new items on the red word test (0.41 and 0.27), $t(20) = 4.94$, $SEM = 0.028$, $d = .81$, and red word items were more likely to be falsely recognized than new items on the picture test (0.19 and 0.11), $t(20) = 3.85$, $SEM = 0.020$, $d = .52$. Such familiarity influences also can explain why items studied in both formats were more likely to be recognized than items studied in only one format on the red word test (0.70 vs. 0.62), $t(20) = 3.02$, $SEM = 0.026$, $d = .61$, and the picture test (0.75 vs. 0.61), $t(20) = 5.49$, $SEM = 0.025$, $d = .90$.

Evidence that subjects had engaged in diagnostic monitoring comes from a planned comparison of false recognition across the red word and picture tests. Because pictures elicited more distinctive recollections than red words, diagnostic monitoring should have been more effective on the picture test, reducing false recognition for items that failed to elicit picture recollections (i.e., a distinctiveness heuristic). Consistent with this prediction, false recognition of noncriterial items was greater on the red word test than on the picture test (0.41 and 0.19), $t(20) = 8.03$, $SEM = 0.028$, $d = 1.49$, and so too was false recognition of new items (0.27 and 0.11), $t(20) = 6.13$, $SEM = 0.027$, $d = .93$. Because familiarity was equated across stimuli, these differences indicate that recollection-based monitoring was more effective on the picture test than on the red word test. The fact that these effects were found for false recognition of red words and new items on the picture test is consistent with this interpretation

Table 1. Speeded Recognition and Subjective Responses in the Manipulation Check Experiment

Item Type	Speeded Test		Recollect/Familiar Test		Recollection Quality Test	
	p , "Yes"	p , "AR"	p , "VF"	IRK	Strength	Details
Red words	.73 (0.04)	.55 (0.05)	.24 (0.03)	.55 (0.08)	4.05 (0.37)	2.10 (0.32)
Pictures	.71 (0.02)	.65 (0.04)	.16 (0.02)	.47 (0.08)	4.55 (0.31)	3.19 (0.30)
New	.36 (0.06)	.08 (0.02)	.16 (0.03)	.19 (0.04)	1.33 (0.28)	0.67 (0.22)

Standard errors of each mean are in parenthesis. Ratings for the recollection quality judgments were on a 0–7 scale. AR = actually recollect, VF = very familiar, IRK = familiarity estimate from independent-recollection-familiarity adjustment.

Table 2. Mean Recognition of Each Item Type on the Critical Recollection Tests and Response Latencies for Correct Responses

	<i>p</i> , "Yes"	Latency "Yes"	Latency "No"
<i>Red Word Test</i>			
Both	.70 (0.02)	1422	–
Red words	.62 (0.03)	1399	–
Pictures	.41 (0.03)	–	1481
New	.27 (0.04)	–	1343
<i>Picture Test</i>			
Both	.75 (0.03)	1209	–
Red words	.19 (0.03)	–	1406
Pictures	.61 (0.03)	1218	–
New	.11 (0.03)	–	1335
<i>Exclusion Test</i>			
Red words	.65 (0.03)	1552	–
Pictures	.30 (0.03)	–	1482
New	.29 (0.04)	–	1442

Standard errors of each mean are in parenthesis. Red words were targets on the red word test and exclusion test, and lures on the picture test. Pictures were targets on the picture test, but lures on the red word test and exclusion test.

because neither of these items should have elicited distinctive picture recollections.

Evidence that subjects had engaged a disqualifying monitoring strategy comes from a planned comparison between the red word test and the exclusion test. The exclusion test required subjects to endorse items that elicited red word recollections, but unlike the red word test, we did not include test items that were studied in both formats on the exclusion test. As a result, if subjects could recollect pictures then they could be sure that the item had not been studied as a red word, thereby reducing false recognition. Unlike the distinctiveness heuristic described above, such a recall-to-reject strategy should selectively reduce false recognition to picture items because only these items could elicit picture recollections. Consistent with this prediction, false recognition of picture items was significantly reduced on the exclusion test (0.30) compared with the red word test (0.41), $t(20) = 4.05$, $SEM = 0.028$, $d = .83$, with no corresponding differences in hits to red word items (0.65 and 0.62, $t < 1$) or false recognition of new items (0.29 vs. 0.27, $t < 1$). Additional evidence that subjects had used a recall-to-reject strategy is that false recognition of picture items and new items was equated on the exclusion test (0.30 vs. 0.29, $t < 1$), although a significant difference was observed on the red word test. On the exclusion test, sub-

jects overcame this false recognition effect by using a recall-to-reject strategy.

Response Latencies

On the basis of prior work, we expected latencies to be slower on the red word test than on the picture test because diagnostic monitoring should have been more effortful on the red word test. These differences were significant for hits to both items, $t(20) = 4.85$, $SEM = 44.01$, $d = .92$, and hits to criterial items, $t(20) = 5.84$, $SEM = 31.11$, $d = 1.21$, but failed to reach significance for correct rejections of noncriterial items, $t(20) = 1.79$, $SEM = 41.73$, $p = .09$, $d = .32$. We also compared latencies across the red word test and the exclusion test. Predictions for this comparison were less clear because the recall-to-reject strategy facilitated rejections (potentially making responses faster) but also should have required the search for multiple recollections (potentially making responses slower). Consistent with the latter, responses to red words were slower on the exclusion test than on the red word test, $t(20) = 3.85$, $SEM = 400$, $d = .76$, and similarly for new items, $t(20) = 3.05$, $SEM = 32.41$, $d = .49$. Subjects may have attempted an unsuccessful recall-to-reject process for these items, thereby slowing the decision. In contrast, there was no difference in response latencies for pictures on the red word test and exclusion tests, $t < 1$, illustrating that response latencies were a poor indicator of underlying retrieval monitoring processes. Although correct rejections were somewhat faster on the picture test (cf. Gallo et al., 2006), they did not differ between the exclusion test and the red word test.

NEUROIMAGING RESULTS

We report two primary sets of neuroimaging analyses. First, we used simple contrasts to compare activity between the memory tests, using unbiased whole-brain analyses ($p < .001$, uncorrected, five contiguous voxel threshold). These contrasts compared the red word test to the picture test, as in our behavioral analysis of diagnostic monitoring, and the red word test to the exclusion test, as in our behavioral analysis of disqualifying monitoring. Second, we conducted conjunction analyses that were designed to more precisely isolate the two types of recollection-based retrieval monitoring while controlling for the potentially confounding influence of retrieval success.

Cross-test Comparisons

For the first set of analyses, we pooled correct responses to red word items and picture items on each test (yielding an average of 40 observations per test, per subject). Pooling these items increased our statistical power for cross-test comparisons and was theoretically justified

because retrieval monitoring processes should have occurred for all studied items (cf. Gallo et al., 2006). This analysis equated item history and response types across the tests (correct hits and correct rejections), so that any resulting differences would be due to the different types of information that were recollected and/or corresponding monitoring processes. Analyses of items studied in both formats are not reported because they were more familiar than the other items (and only occurred on two of the tests). We also compared correct rejections of new items across the tests.

As can be seen from Table 3, two right frontal regions were more active for studied items on the red word test than the picture test, including a region in right DLPFC (near BA 8/9). In contrast, no regions were more active for studied items on the picture test compared with the red word test. This pattern is consistent with the idea that diagnostic retrieval monitoring was more effortful on the red word test than the picture test, thereby recruiting DLPFC more heavily. On the picture test, subjects were able to use the distinctiveness heuristic to suppress false recognition, and basing their decisions on relatively more distinctive recollections placed fewer demands on DLPFC. No DLPFC regions were active for the corresponding new item contrasts, although a more anterior and medial frontal region was more active on the red word test relative to the picture test (see Table 4). New items were less familiar than studied items and so should not have engaged recollection-based retrieval monitoring to the same extent, although false recognition to new items was greater on the red word test than on the picture test, suggesting at least some degree of retrieval monitoring.

Comparing the exclusion test and the red word test revealed that studied items were more likely to activate

frontal regions on the exclusion test, including a cluster in left anterior DLPFC (near BA 10/46). In contrast, no frontal regions were more active on the red word test than on the exclusion test. This pattern suggests that the picture-based recall-to-reject strategy increased DLPFC activity relative to the red word test, in contrast to the picture-based distinctiveness heuristic, which decreased DLPFC activity relative to the red word test. The comparison with new items yielded no activity in the left hemisphere, although several right frontal regions were more active on the exclusion test than on the red word test (including a DLPFC region near BA 8). This activity may reflect additional monitoring demands on the exclusion test, a possibility we discuss more in the General Discussion.

Monitoring Conjunctions

The aforementioned analyses of studied items revealed that frontal regions were less active on the picture test compared with the red word test but more active on the exclusion test compared with the red word test. Moreover, these effects appeared to be lateralized, with the former emerging in right DLPFC and the latter in the left. As discussed, these effects may reflect contributions of diagnostic and disqualifying monitoring, respectively, but different types of information may have been recollected across the tests. To further isolate retrieval monitoring, we conducted two conjunction analyses, controlling for retrieval success. Conjunction analyses identify regions of overlap between two simple contrasts, making them more conservative than the contributing contrasts. Although the simple contrasts in these analyses were not independent, we used a more liberal threshold for each contributing contrast to

Table 3. All Active Clusters from Comparing Studied Items across the Recollection Tests

<i>Talairach</i>	<i>No. Voxels</i>	<i>Approximate Region</i>	<i>BA</i>
Red Word Test > Picture Test			
33, 2, 44	16	R frontal/middle gyrus	6
27, 37, 40	12	R frontal/middle gyrus	8/9
30, -74, 29	14	R occipital/superior gyrus	19
-65, -22, 20	7	L parietal/postcentral gyrus	40
Picture Test > Red Word Test			
No suprathreshold regions			
Exclusion Test > Red Word Test			
3, 32, 48	5	R frontal/medial gyrus	8
-33, 45, 23	12	L frontal/middle gyrus	10/46
Red Word Test > Exclusion Test			
-62, -16, 20	10	L parietal/postcentral gyrus	43/40

Talairach coordinates (x, y, z) are the peak activation within a cluster, arranged anterior to posterior and laterally (R = right, L = left). BA = approximate Brodmann's areas.

Table 4. All Active Clusters from Comparing New Items across the Recollection Tests

<i>Talairach</i>	<i>No. Voxels</i>	<i>Approximate Region</i>	<i>BA</i>
Red Word Test > Picture Test			
12, 47, 0	5	R frontal/medial gyrus	10
-21, -81, 21	19	L occipital/cuneus	18
Picture Test > Red Word Test			
24, -40, 13	10	R sublobular/caudate nucleus	NA
Exclusion Test > Red Word Test			
21, 46, -10	5	R frontal/orbital gyrus	11
45, 22, 40	11	R frontal/middle gyrus	8
30, 20, 52	9	R frontal/middle gyrus	6/8
42, -10, 31	6	R frontal/precentral gyrus	6
48, -45, 30	20	R parietal/supramarginal gyrus	40
6, -57, 30	14	R parietal/precuneus	7
Red Word Test > Exclusion Test		No suprathreshold regions	

Talairach coordinates (x, y, z) are the peak activation within a cluster, arranged anterior to posterior and laterally (R = right, L = left). BA = approximate Brodmann's areas.

characterize the regions of overlap and avoid Type II error ($p < .01$, uncorrected, five contiguous voxel threshold).

To isolate diagnostic monitoring, we performed a conjunction analysis to find regions that were more active when subjects rejected picture items on the red word test compared with rejecting red word items on the picture test, as well as rejecting picture items on the red word test compared with accepting picture items on the picture test. The first contrast controls for the type of response across tests (correct rejection of equally familiar studied items) but varies item history and potential retrieval success (pictures and red words). The second contrast controls for item history and potential retrieval success effects (i.e., the recollection of pictures) while varying the type of response across tests (correct rejection and correct acceptance). The resulting overlap should reflect the diagnostic monitoring process that was more effortful on the red word test than on the picture test.

Figure 1 (top panel) illustrates prefrontal activity observed for this diagnostic monitoring conjunction. There were two prominent clusters in right DLPFC (Talairach coordinates for the centers = 27, 37, 41, near BA 8/9 on the middle frontal gyrus, and 31, 48, 30, near BA 9 on the superior frontal gyrus) as well as a more posterior right PFC region (34, 3, 47, near BA 6). There was no prominent activity in analogous left DLPFC regions, although there was overlapping activity near cingulate gyrus (-20, 8, 40, near BA 32) and relatively smaller clusters near BA 8 (-27, 23, 38) and BA 44 (-50, 14, 16). To show the extent of these and other regions of activity, Figure 2 presents whole brain activity for this conjunction. Of additional interest were clusters in bilateral inferior parietal cortex

(near BA 40) as well as more posterior regions in right occipital cortex, that is, fusiform (BA 19), lingual gyrus (BA 18), and a more superior region near BA 39/19. These latter regions have been associated with retrieval success in several memory studies, a point that we discuss more in the General Discussion section.

We next conducted an ROI analysis as a direct test of the lateralization observed in DLPFC activity. Using MarsBar software (Brett, Anton, Valabregue, & Poline, 2002), we created two ROIs based on a 5-mm sphere centered around the peak voxel of the two DLPFC clusters found in the conjunction analysis. We then created two ROIs that were based on the same coordinates but in the opposite hemisphere. For each ROI, we extracted the percent signal change (relative to the ROI mean signal) for red word items and picture items and then conducted a 2 (Hemisphere: left, right) \times 2 (Response Types: hits, correct rejections) \times 2 (Test: red word test, picture test) ANOVA on the extracted signal.¹ Analysis of the cluster near BA 9 (31, 48, 30 on the right) confirmed an interaction between test type and hemisphere, $F(1,20) = 4.29$, $MSE = 0.698$, $p = .05$, $\eta_p^2 = .177$, with no other main effects or interactions. Follow-up t tests revealed that this interaction was driven by greater right DLPFC activity for the rejection of lures on the red word test compared with the picture test, $t(20) = 2.38$, $SEM = 0.201$, $d = .49$, with no other significant effects. A similar analysis of the cluster near BA 8/9 (27, 37, 41) revealed only a marginal effect of test (red word test > picture test, $p = .10$) with no significant interactions. Thus, although we observed two right DLPFC clusters in the whole-brain conjunction, only one of these regions survived the direct test of laterality.

Figure 1. Axial slices illustrating prefrontal activity observed in the diagnostic monitoring conjunction (top) and in the disqualifying monitoring conjunction (bottom).

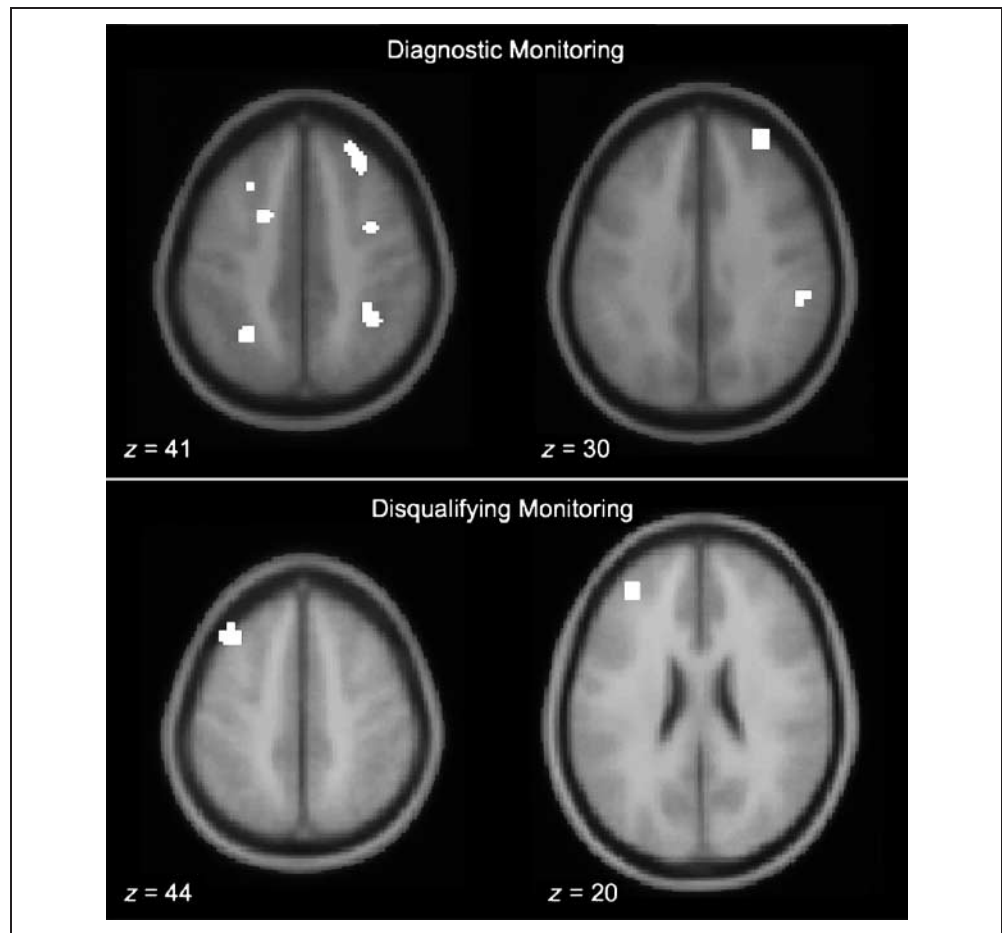
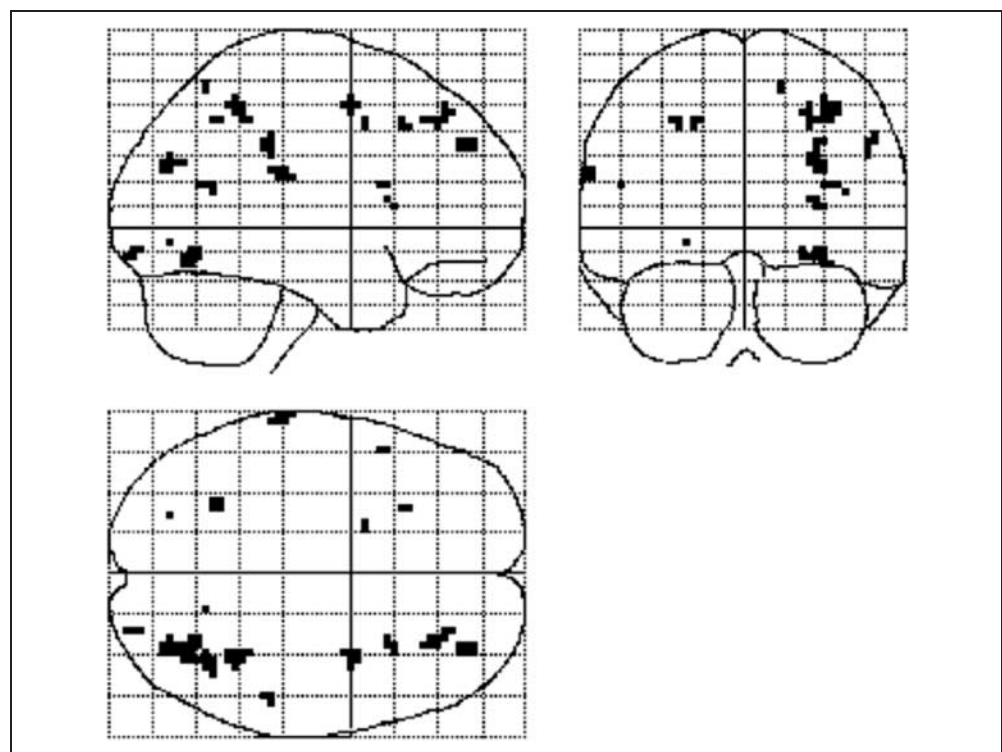


Figure 2. All active clusters in the diagnostic monitoring conjunction projected onto transparent brain templates.



To isolate disqualifying monitoring, we performed a conjunction analysis to find regions that were more active when subjects rejected picture items on the exclusion test compared with rejecting picture items on the red word test as well as rejecting picture items on the exclusion test compared with accepting picture items on the picture test. The first contrast controls for the type of response across tests (correct rejection of picture items) but potentially varies retrieval success (i.e., being greater when subjects were explicitly attempting to recollect pictures in the rejection strategy). The second contrast controls for retrieval success while varying the type of response across tests (correct rejection and correct acceptance). The resulting overlap should represent the disqualifying monitoring process that we assume was used only on the exclusion test.

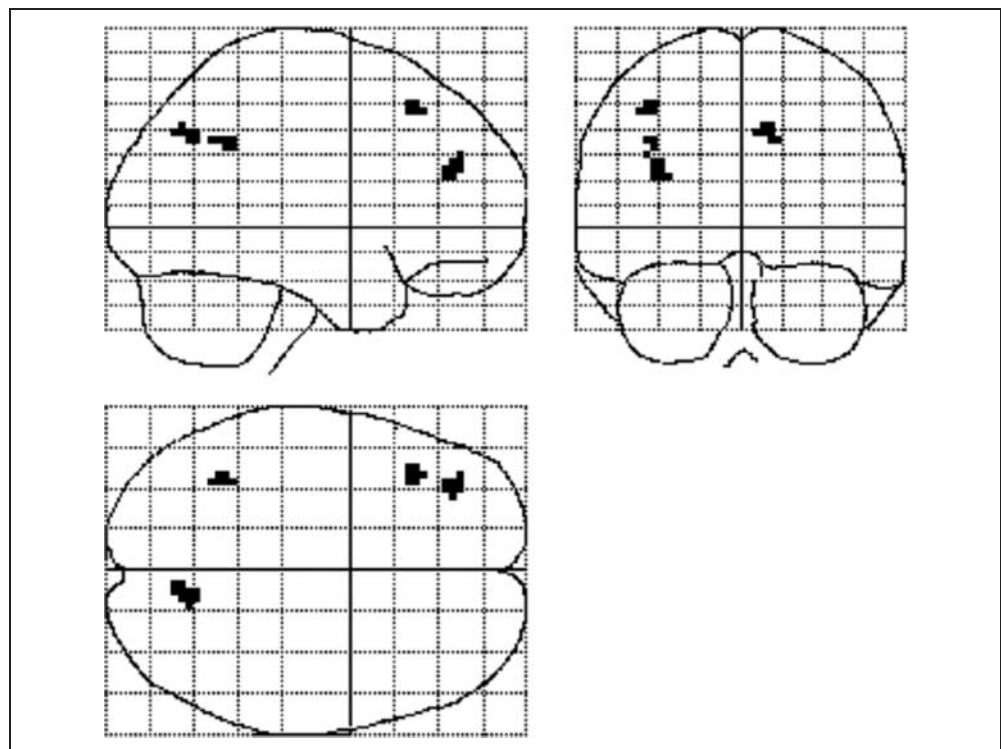
Figure 1 (bottom panel) illustrates prefrontal activity observed for this disqualifying monitoring conjunction. There were two clusters in left DLPFC, one near BA 10/46 on the middle frontal gyrus ($-34, 42, 20$) and one near BA 8 on the middle frontal gyrus ($-38, 28, 44$), with no corresponding activity on the right. Figure 3 presents whole brain activity for this conjunction. The two posterior clusters were in left inferior parietal cortex ($-36, -49, 35$, near BA 40, supramarginal gyrus) and precuneus ($10, -64, 39$, near BA 7). To explore the lateralization of the observed DLPFC effects, we again conducted an ANOVA on the signal extracted from ROIs centered on the obtained clusters. Analysis of the cluster near BA 10/46 ($-34, 42, 20$ on the left) revealed a main effect of test (exclusion test > red word test), $F(1,20) = 9.86$, $MSE = 0.468$, $\eta_p^2 = .330$, with

no other effects or interactions. Analysis of the cluster near BA 8 ($-38, 28, 44$ on the left) also revealed a main effect of test (exclusion test > red word test), $F(1,20) = 6.94$, $MSE = 1.23$, $\eta_p^2 = .258$, and further there was an effect of response (correct rejections > hits), $F(1,20) = 5.20$, $MSE = 0.408$, $\eta_p^2 = .206$, with no other effects or interactions. Thus, although the exclusion test elicited greater activity in these DLPFC regions than the red word test, these effects were not significantly greater in the left hemisphere than in the right.

GENERAL DISCUSSION

We measured the neural correlates of two fundamental types of retrieval monitoring. Subjects studied red words and pictures under conditions where (a) these stimuli were equated on familiarity and (b) pictures afforded more distinctive recollections than red words. We found that source memory confusions were reduced on the picture test relative to the red word test, indicating that expecting more detailed recollections for pictures enhanced diagnostic monitoring accuracy (i.e., a distinctiveness heuristic; Dodson et al., 2000; Schacter et al., 1999). The use of this diagnostic monitoring process on the picture test was associated with less DLPFC activity compared with the red word test. We also found that source memory confusions were reduced in a mutually exclusive condition compared with a nonexclusive condition, indicating that the exclusivity relationship facilitated disqualifying monitoring accuracy (i.e., a recall-to-reject strategy; Jacoby et al., 1998; Hintzman & Curran, 1994). The use of this

Figure 3. All active clusters in the disqualifying monitoring conjunction projected onto transparent brain templates.



disqualifying monitoring strategy on the exclusion test was associated with increased DLPFC activity compared with the red word test.

Our results indicate that different types of recollection-based retrieval monitoring correspond to different patterns of prefrontal activity. This activity cannot be mapped easily onto single-process notions such as retrieval effort. For example, one might argue that the red word test required the most retrieval effort (hence prefrontal resources) relative to the other two tests because it elicited the greatest number of source memory confusions. This idea can explain why the red word test elicited more prefrontal activity than the picture test (which benefitted from distinctive picture recollections), but it cannot explain why there was less activity compared with the exclusion test (which also benefitted from picture recollections). The argument that all recollection-based rejection processes recruit prefrontal regions also falters because the picture test clearly benefitted from the use of a recollection-based monitoring process but was less likely to activate DLPFC than the red word test. More generally, all three of our tests required the recollection and the monitoring of specific details.

Our results are best understood by considering the two different types of retrieval monitoring processes isolated by our task. When subjects used the absence of picture recollections to reject familiar items (the distinctiveness heuristic), the memory decision was associated with less effortful diagnostic monitoring and led to reduced DLPFC activity. Subjects engaged in diagnostic monitoring on both the red word test and the picture test, but such monitoring was more heavily recruited on the red word test (i.e., they had to spend more effort recollecting the studied context). In contrast, when subjects used the presence of picture recollections to reject familiar lures (recall-to-reject), the memory decision was associated with an additional disqualifying monitoring strategy and led to increased DLPFC activity. Unlike diagnostic monitoring, which theoretically occurred to some degree on all of our tests, subjects were only justified in using disqualifying monitoring on the exclusion test (i.e., recollecting pictures to avoid source confusions). Considered as a whole, these results implicate DLPFC in both types of retrieval monitoring, with the relative level of activity tracking those memory decisions that more heavily relied on a given type of monitoring.

Retrieval Monitoring Theories

The current results help to constrain the more general theories of retrieval monitoring that were described in the Introduction. Consider the systematic/heuristic distinction of the source monitoring framework (Johnson, 2006). Heuristic processes involve relatively fast or simple decisions, such as when a memory decision is based on a single type of information like familiarity. Similar to a signal-detection process, subjects expect to retrieve a certain

degree of information along the relevant memory dimension, and a test item is compared with this criterion to make a decision (cf. Johnson & Raye, 1981). In contrast, systematic processes take advantage of additional information to help make a memory decision. Systematic processing may also involve criterion setting, but it goes a step further in terms of the strategic use of contextual information that also can inform the decision. Heuristic processes have been attributed more to the right PFC and systematic processes have been attributed more to the left PFC (e.g., Nolde et al., 1998).

The disqualifying monitoring process that occurred on our exclusion test can be considered a type of systematic process. On this test, in addition to searching memory for the to-be-recollected information, subjects also benefitted from strategically recollecting noncritical recollections (i.e., a recall-to-reject strategy). Consistent with the idea that left DLPFC supports systematic processes, we found that left DLPFC was more active on the exclusion test than on the red word test in several analyses, although a direct test of the laterality of this effect was not significant.² A similar effect in right DLPFC was not observed in any of our analyses, providing little support for the idea that right DLPFC is critical for postretrieval monitoring (Rugg et al., 2003; Henson et al., 1999).

It is less clear what the source monitoring framework would predict for the retrieval monitoring process that occurred on our nonexclusion test. If one considers the distinctiveness heuristic to be a special type of metacognitive monitoring process, one that is engaged only when distinctive recollections are expected (cf. Schacter & Wiseman, 2006; Dodson et al., 2000), then the picture test involved a heuristic process that did not occur on the red word test. This conceptualization also would be keeping with the idea that heuristic processes are relatively fast acting, as response latencies were fastest on the picture test. In this case, the systematic/heuristic distinction of the source monitoring framework would incorrectly predict greater right DLPFC for the picture test compared with the red word test because the right DLPFC has been associated with heuristic-based responding.

In contrast, the current approach conceptualizes the distinctiveness heuristic as just one instance of a more general diagnostic monitoring process, one that occurs whenever subjects must base a memory decision on specific source recollections. On each of our critical recollection tests, subjects set a response criterion along the relevant recollection dimension to make a memory decision. Such diagnostic monitoring was more effortful on the red word test than the picture test, making it more likely to recruit right DLPFC. From this perspective, previous neuroimaging findings that were attributed to heuristic processes (e.g., fast and effortless decisions) may instead have been caused by diagnostic monitoring (e.g., setting response criterion along a single dimension). Either way, the current study indicates the importance of studying specific types of decision processes to understand

retrieval monitoring, as opposed to more general notions of retrieval speed or difficulty.

Outstanding Questions

One remaining question is the role of the more posterior activity observed in our conjunction analyses, including left parietal and right occipital regions (e.g., fusiform and precuneus). Activity in these regions has been attributed to retrieval success in many memory studies (for a review, see Skinner & Fernandes, 2007), with parietal activity potentially reflecting attention to information that is thought to be “old” and occipital activity potentially reflecting the perceptual reactivation of picture recollections (e.g., Woodruff, Johnson, Uncapher, & Rugg, 2005; Wheeler & Buckner, 2004). Although pictures also were involved in each of our conjunctions, we controlled for retrieval success effects by including picture hits on the picture test. This activity observed here instead may reflect the top-down modulation of these regions during the retrieval monitoring process (see Hornberger, Rugg, & Henson, 2006). Although speculative, our subjects may have allocated more attention to picture representations when attempting to use a recall-to-reject strategy.

A related question regards the activity in right DLPFC and posterior regions that we observed for new items on the exclusion test. Activity is not always obtained for new items across conditions that differ in retrieval monitoring demands (Gallo et al., 2006; Rugg et al., 2003), but Woodruff et al. (2006) also observed activity for new items in DLPFC regions and similar posterior regions during an exclusion task. They attributed this activity to retrieval monitoring processes, such as orienting toward a specific type of information during a memory search. The current data are consistent with this account. As discussed, response latencies were slower for new items on the exclusion test than on the red word test, potentially because subjects had attempted a recall-to-reject strategy. However, new items could not benefit from this strategy and instead needed to be rejected via diagnostic monitoring processes. The observed activity might reflect the use of these additional monitoring processes.

One final question is the degree that the current findings will generalize to the monitoring of other types of events in memory. As discussed by Mitchell et al. (2008) and others, retrieval-monitoring processes operate across different types of materials, but they also can be influenced by the quality of retrieved information. We attempted to overcome any material-specific effects in the current study with our specific contrasts and conjunction analyses. It remains to be seen, though, whether the current pattern of effects can be found with other types of materials. As discussed, evidence for a recall-to-reject strategy and the distinctiveness heuristic has been found across a variety of cognitive tasks and materials. To the degree that the same types of decision processes underlie these various effects, studies with these other materials should find similar patterns of

DLPFC activity as were found here. Such findings would bolster the idea that an understanding of the types of decision processes involved is critical for understanding prefrontal contributions to retrieval monitoring.

Acknowledgments

This project was funded by grants from the University of Chicago’s Office of the Vice President, Social Sciences Division, and Brain Research Imaging Center. Thanks to Jean Decety, Robert Lyons, and Steve Small for technical assistance.

Reprint requests should be sent to David A. Gallo, Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, IL 60637, or via e-mail: dgallo@uchicago.edu.

Notes

1. It is important to acknowledge that this is a biased selection of ROIs because it is based on observed activation patterns from the whole-brain analysis. Nevertheless, this analysis allows for a direct test of whether the right-lateralized effects observed in the unbiased whole-brain conjunction were significantly different from any subthreshold effects that also may have been present in the left hemisphere.
2. We do not believe that this laterality effect is due to the materials that we used. First, we used verbal labels on all of our memory tests, so that the critical difference between these two tests was in the distinctiveness of the to-be-recollected stimulus (a red word or a picture). Second, the conjunction analysis attempted to control for any material-specific retrieval success effects. Finally, we found that searching memory for the verbal information (red words) was more likely to activate right DLPFC compared with picture information, a finding that is opposite to what one might expect based on material-specific lateralization of DLPFC activity (e.g., Kelley et al., 1998).

REFERENCES

- Achim, A. M., & Lepage, M. (2005). Dorsolateral prefrontal cortex involvement in memory post-retrieval monitoring revealed in both item and associative recognition tests. *Neuroimage, 24*, 1113–1121.
- Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer’s disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language, 46*, 199–226.
- Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review, 110*, 762–784.
- Brett, M., Anton, J. L., Valabregue, R., & Poline, J. B. (2002). *Region of interest analysis using an SPM toolbox*. Abstract presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan. *Neuroimage, 16*, Abstract No. 2.
- Budson, A. E., Droller, D. B. J., Dodson, C. S., Schacter, D. L., Rugg, M. D., Holcomb, P. J., et al. (2005). Electrophysiological dissociations of picture versus word encoding: The distinctiveness heuristic as a retrieval orientation. *Journal of Cognitive Neuroscience, 17*, 1181–1193.
- Cabeza, R., Rao, S. M., Wagner, A. D., Mayer, A. R., & Schacter, D. L. (2001). Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study

- of veridical and illusory recognition memory. *Proceedings of the National Academy of Sciences, U.S.A.*, *98*, 4805–4810.
- 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.
- Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, *8*, 109–114.
- Dobbins, I. G., & Han, S. (2006). Isolating rule- versus evidence-based prefrontal activity during episodic and lexical discrimination: A functional magnetic resonance imaging investigation of detection theory distinctions. *Cerebral Cortex*, *16*, 1614–1622.
- Dobbins, I. G., Simons, J. S., & Schacter, D. L. (2004). fMRI evidence for separable and lateralized prefrontal memory monitoring processes. *Journal of Cognitive Neuroscience*, *16*, 908–920.
- Dodson, C. S., Koutstaal, W., & Schacter, D. L. (2000). Escape from illusion: Reducing false memories. *Trends in Cognitive Sciences*, *4*, 391–397.
- Dudukovic, N. M., & Wagner, A. D. (2007). Goal-dependent modulation of declarative memory: Neural correlates of temporal recency decisions and novelty detection. *Neuropsychologia*, *45*, 2608–2620.
- Fleck, M. S., Daselaar, S. M., Dobbins, I. G., & Cabeza, R. (2005). Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. *Cerebral Cortex*, *16*, 1623–1630.
- Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory—Insights from functional neuroimaging. *Brain*, *124*, 849–881.
- Fraser, C. S., Brisdon, N. C., & Wilding, E. L. (2007). Controlled retrieval processing in recognition memory exclusion tasks. *Brain Research*, *1150*, 131–142.
- Gallo, D. A. (2004). Using recall to reduce false recognition: Diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 120–128.
- Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York: Psychology Press.
- Gallo, D. A., Cotel, S. C., Moore, C. D., & Schacter, D. L. (2007). Aging can spare recollection-based retrieval monitoring: The importance of event distinctiveness. *Psychology and Aging*, *22*, 209–213.
- Gallo, D. A., Kensinger, E. A., & Schacter, D. L. (2006). Prefrontal activity and diagnostic monitoring of memory retrieval: fMRI of the criterial recollection task. *Journal of Cognitive Neuroscience*, *18*, 135–148.
- Ghetti, S. (2003). Memory for nonoccurrences: The role of metacognition. *Journal of Memory and Language*, *48*, 722–739.
- Hayama, H. R., Johnson, J. D., & Rugg, M. D. (2008). The relationship between the right frontal old/new ERP effect and post-retrieval monitoring: Specific or non-specific? *Neuropsychologia*, *46*, 1211–1223.
- Henson, R. N. A., Shallice, T., & Dolan, R. J. (1999). Right prefrontal cortex and episodic memory retrieval: A functional MRI test of the monitoring hypothesis. *Brain*, *122*, 1367–1381.
- Herron, J. E., & Rugg, M. D. (2003). Strategic influences on recollection in the exclusion task: Electrophysiological evidence. *Psychonomic Bulletin & Review*, *10*, 703–710.
- Hintzman, D., & Curran, D. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language*, *33*, 1–18.
- Hornberger, M., Rugg, M. D., & Henson, R. N. A. (2006). fMRI correlates of retrieval orientation. *Neuropsychologia*, *44*, 1425–1436.
- Hwang, D. Y., Gallo, D. A., Ally, B. A., Black, P. M., Schacter, D. L., & Budson, A. E. (2007). Diagnostic retrieval monitoring in patients with frontal lobe lesions: Further exploration of the distinctiveness heuristic. *Neuropsychologia*, *45*, 2543–2552.
- Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of know judgments and exclusion errors. *Psychonomic Bulletin & Review*, *5*, 705–709.
- Johnson, M. K. (2006). Memory and reality. *American Psychologist*, *61*, 760–771.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, *88*, 67–85.
- Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology*, *94*, 37–64.
- Kelley, W. M., Miezin, F. M., McDermott, K. B., Buckner, R. L., Raichle, M. E., Cohen, N. J., et al. (1998). Hemispheric specialization in human dorsal frontal cortex and medial temporal lobe for verbal and nonverbal memory encoding. *Neuron*, *20*, 927–936.
- Kensinger, E. A., Clarke, R. J., & Corkin, S. (2003). What neural correlates underlie successful encoding and retrieval? A functional magnetic resonance imaging study using a divided attention paradigm. *Journal of Neuroscience*, *23*, 2407–2415.
- Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 332–342.
- Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping*, *10*, 120–131.
- Lepage, M. (2004). Differential contribution of left and right prefrontal cortex to associative cued-recall memory: A parametric PET study. *Neuroscience Research*, *48*, 297–304.
- Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1137–1151.
- McDermott, K. B., Jones, T. C., Petersen, S. E., Lageman, S. K., & Roediger, H. L., III (2000). Retrieval success is accompanied by enhanced activation in anterior prefrontal cortex during recognition memory: An event-related fMRI study. *Journal of Cognitive Neuroscience*, *12*, 965–975.
- McDonough, I. M., & Gallo, D. A. (2008). Autobiographical elaboration reduces false recognition: Cognitive operations and the distinctiveness heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1430–1445.
- Mitchell, K. J., Johnson, M. K., Raye, C. L., & Greene, E. J. (2004). Prefrontal cortex activity associated with source monitoring in a working memory task. *Journal of Cognitive Neuroscience*, *16*, 921–934.
- Mitchell, K. J., Raye, C. L., McGuire, J. T., Frankel, H., Greene, E. J., & Johnson, M. K. (2008). Neuroimaging evidence for agenda-dependent monitoring of different features during short-term source memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 780–790.
- Nolde, S. F., Johnson, M. K., & D'Esposito, M. (1998). Left prefrontal activation during episodic remembering: An event-related fMRI study. *NeuroReport*, *15*, 3509–3514.

- Nolde, S. F., Johnson, M. K., & Raye, C. L. (1989). The role of prefrontal cortex during tests of episodic memory. *Trends in Cognitive Sciences*, 2, 399–406.
- Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review*, 15, 105–130.
- Rajah, M. N., Ames, B., & D'Esposito, M. (2008). Prefrontal contributions to domain-general executive control processes during temporal context retrieval. *Neuropsychologia*, 46, 1088–1103.
- Ranganath, C. (2004). The 3-D prefrontal cortex: Hemispheric asymmetries in prefrontal activity and their relation to memory retrieval processes. *Journal of Cognitive Neuroscience*, 16, 903–907.
- Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory and Cognition*, 28, 907–922.
- Rugg, M. D. (2004). Retrieval processing in human memory: Electrophysiological and fMRI evidence. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (3rd ed., pp. 727–738). Cambridge, MA: MIT Press.
- Rugg, M. D., Fletcher, P. C., Chua, P. M.-L., & Dolan, R. J. (1999). The role of prefrontal cortex in recognition memory and memory for source: An fMRI study. *Neuroimage*, 10, 520–529.
- Rugg, M. D., Henson, R. N. A., & Robb, W. G. K. (2003). Neural correlates of retrieval processing in the prefrontal cortex during recognition and exclusion tasks. *Neuropsychologia*, 41, 40–52.
- Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, 40, 1–24.
- Schacter, D. L., & Wiseman, A. L. (2006). Reducing memory errors: The distinctiveness heuristic. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 89–107). Oxford: Oxford University Press.
- Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia*, 45, 2163–2179.
- Talairach, J., & Tournoux, P. (1998). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- Velanova, K., Jacoby, L. L., Wheeler, M. E., McAvoy, M. P., Petersen, S. E., & Buckner, L. (2003). Functional-anatomic correlates of sustained and transient processing components engaged during controlled retrieval. *Journal of Neuroscience*, 23, 8460–8470.
- 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.
- 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.
- Woodruff, C. C., Uncapher, M. R., & Rugg, M. D. (2006). Neural correlates of differential retrieval orientation: Sustained and item-related components. *Neuropsychologia*, 44, 3000–3010.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.