# Age-Related Differences in Prefrontal Cortex Activity during Retrieval Monitoring: Testing the Compensation and Dysfunction Accounts

Ian M. McDonough[1], Jessica T. Wong[2] and David A. Gallo[2,3]

[1]Center for Vital Longevity, School of Behavioral and Brain Sciences, University of Texas at Dallas, Dallas, TX 75235, USA, [2]Department of Psychology and [3]Center for Cognitive and Social Neuroscience, University of Chicago, Chicago, IL 60637, USA

Address correspondence to David A. Gallo, University of Chicago, 5848 South University Avenue, Chicago, IL 60637, USA. Email: dgallo@uchicago.edu.

**Current theories of cognitive aging emphasize that the prefrontal cortex might not only be a major source of dysfunction but also a source of compensation. We evaluated neural activity associated with retrieval monitoring—or the selection and evaluation of recollected information during memory retrieval—for evidence of dysfunction or compensation. Younger and older adults studied pictures and words and were subsequently given criterial recollection tests during event-related functional magnetic resonance imaging. Although memory accuracy was greater on the picture test than the word test in both groups, activity in right dorsolateral prefrontal cortex (DLPFC) was associated with greater retrieval monitoring demands (word test > picture test) only in younger adults. Similarly, DLPFC activity was consistently associated with greater item difficulty (studied > nonstudied) only in younger adults. Older adults instead exhibited high levels of DLPFC activity for all of these conditions, and activity was greater than younger adults even when test performance was naturally matched across the groups (picture test). Correlations also differed between DLPFC activity and test performance across the groups. Collectively, these findings are more consistent with accounts of DLPFC dysfunction than compensation, suggesting that aging disrupts the otherwise beneficial coupling between DLPFC recruitment and retrieval monitoring demands.**

**Keywords:** aging, dorsolateral prefrontal cortex, false recognition, fMRI, source memory

## Introduction

One of the hallmark changes associated with cognitive aging is episodic memory decline (Zacks et al. 2000). Older adults are less likely than younger adults to recollect specific details from the past (i.e., forgetting), and they are more likely to recollect distorted details (i.e., false memories). While some of these recollection impairments may be due to incorrect binding of individual details or features into a single event at encoding (Johnson et al. 1993; Naveh-Benjamin 2000; Glisky et al. 2001), other evidence suggests additional impairments in monitoring processes at retrieval, such as the inappropriate selection and evaluation of recollected details or source information (Schacter et al. 1997; Mitchell and Johnson 2009). Some of the strongest evidence for age-related impairments in retrieval monitoring has come from false recognition tests or situations where familiar but nonstudied items must be rejected. A recent meta-analysis of this literature revealed that older adults claimed to falsely recollect nonstudied items twice as often as younger adults (McCabe et al. 2009). Because nonstudied items are not presented during the encoding phase, these memory errors implicate an age-related impairment in retrieval monitoring processes.

It has been argued that age-related retrieval monitoring impairments are primarily due to dysfunction of the prefrontal cortex (PFC), consistent with disproportionate structural and neurochemical alterations in PFC as a function of aging (e.g., Middleton and Strick 2000; Li et al. 2001; Head et al. 2004; Salat et al. 2004; Raz et al. 2005). Within the aging literature, evidence linking prefrontal function to retrieval monitoring primarily has come from the use of neuropsychological tests. These studies revealed that the rate of false recognition errors in older adults was negatively correlated with performance on tests that heavily rely on frontal lobe functioning, independent from performance on tests that heavily rely on medial temporal lobe functioning (e.g., Butler et al. 2004; Dornburg and McDaniel 2006; Roediger and McDaniel 2007; McCabe et al. 2009). Outside of the aging literature, evidence for the role of PFC in retrieval monitoring comes from patients with frontal lobe damage, who have shown increases in false recognition (Parkin et al. 1996; Schacter et al. 1996; Curran et al. 1997; but see Verfaellie et al. 2004; Hwang et al. 2007), and also from cognitively normal younger adults, who have shown that increases in retrieval monitoring demands are associated with greater functional magnetic resonance imaging (fMRI) activity in PFC regions, particularly in dorsolateral prefrontal cortex (DLPFC; Gallo et al. 2010; see also Henson et al. 1999; Cansino et al. 2002).

Recent neuroimaging work suggests that aging alters PFC activity during episodic memory retrieval, but the exact role of retrieval monitoring processes in these effects has not been directly investigated. For example, some research has found greater PFC activity in older adults relative to younger adults when contrasting memory retrieval to a nontask (or implicit) baseline (see Velanova et al. 2007; Davis et al. 2008). These findings potentially implicate PFC dysfunction, although poorly understood age-related physiological differences also might complicate the interpretation of baseline activity compared with task-related activity across age groups (Buckner et al. 2000; Rajah and D'Esposito 2005). Other neuroimaging studies have reported age-related differences in PFC activity when contrasting different retrieval conditions, which potentially avoids these baseline issues, but a clear pattern with respect to retrieval monitoring demands has yet to emerge. For example, Duarte et al. (2010) found that PFC activity in younger adults was greater for familiarity-based true recognition than for false recognition, and this difference was reduced in older adults (see also Dennis et al. 2008). This pattern might indicate that older adults were less sensitive to the retrieval monitoring demands of the different test items, but studies using source memory tasks have observed the opposite pattern, with greater PFC activity for correct relative to incorrect source judgments in older adults than in younger adults (Morcom et al. 2007; Duarte et al. 2008; Duverne et al. 2008). Similarly, some studies

have found greater differences in PFC activity between easy and hard memory tasks in older relative to younger adults (e.g., Morcom et al. 2007), yet other studies have found the opposite result (e.g., Gutchess et al. 2007; Rajah et al. 2010). It is unclear why these different results were obtained, but because these studies were not specifically designed to isolate retrieval monitoring processes, some of the age-related effects might have been due to activity associated with retrieval itself (i.e., recollection or familiarity) as opposed to retrieval monitoring processes.

When considering the relevance of neuroimaging studies to age-related effects on retrieval monitoring processes, an important theoretical question to consider is the meaning of age-related differences in fMRI activity more generally. Much debate has ensued over which aspects of age-related differences in PFC activity during memory retrieval are due to structural and neurochemical deficits with aging (i.e., dysfunction of brain regions), and conversely, which aspects might have been due to the attempted recruitment of cognitive processes to overcome these or other deficits (i.e., functional compensation; for reviews, see Cabeza 2002; Park and Reuter-Lorenz 2009). Age-related differences in fMRI activity, whether relative to a nontask baseline or another task, might be due to differences in processing strategies in the absence of functional impairments of the active region(s), and/or they might be due to age-related dysfunction of these same regions.

According to a compensation account, older adults over-recruit PFC regions relative to younger adults to compensate for age-related deficiencies, such as frontal volume shrinkage (Buckner 2004; Greenwood 2007), hippocampal dysfunction (Park and Gutchess 2005; Persson et al. 2006), or lack of specificity in occipitotemporal functioning (Park et al. 2004; Gazzaley et al. 2005; Davis et al. 2008). These kinds of theories assume that prefrontal regions are associated with cognitively controlled processes or strategies that can help support task performance. Support for these ideas comes from studies showing increases in PFC activity in more demanding relative to less demanding retrieval tasks, especially in high performing compared with low performing older adults (e.g., Cabeza et al. 2002; Gutchess et al. 2007). However, it is important to note that compensation-related activity and behavioral performance may not always be positively related because there is no guarantee that increased attempts to compensate in older adults will actually succeed in enhancing performance (cf. Rajah and D'Esposito 2005).

According to a dysfunction account, the over-recruitment of PFC regions in older adults might be due to inefficient processing in PFC regions, as opposed to strategic compensation. This dysfunction might cause over-recruitment across a wide variety of task conditions, resulting in less selectivity in the neural response of older adults relative to younger adults. This pattern of activity, sometimes described as dedifferentiation, may reflect a breakdown in functional specialization (Lindenberger and Baltes 1994; Baltes and Lindenberger 1997; Li and Lindenberger 1999). At the neural level, these decreases in functional specialization may be manifested as declines in signal and increases in noise in the blood oxygen level–dependent (BOLD) response of PFC regions (Rypma and D'Esposito 2000). Neuroimaging support for these ideas comes from semantic and phonetic fluency tasks that showed recruitment of different PFC regions in younger adults but recruitment of the same regions in both tasks in older adults

(Meinzer et al. 2009). In other fMRI studies, older adults showed less modulation of neural activity within single PFC regions relative to younger adults during perception (e.g., Grady 2002; Carp et al. 2011) and memory retrieval (e.g., Grady 2002; Gutchess et al. 2007), also potentially implicating dysfunction.

## Current Study

In the current study, we tested between these compensation and dysfunction accounts using a task that was specifically designed to isolate the neural correlates of retrieval monitoring processes. A body of behavioral research indicates that false recognition can be reduced when retrieval is oriented toward more distinctive recollections, such as pictures relative to words, in both younger and older adults (e.g., Schacter et al. 1999; Dodson and Schacter 2002; Gallo et al. 2007). These false recognition effects have been attributed to differences in retrieval monitoring, with more accurate and less effortful monitoring processes required when retrieval is oriented toward higher quality or more distinctive picture recollections.

To investigate these retrieval monitoring processes, Gallo et al. (2004) developed the criterial recollection task, which requires subjects to recollect different kinds of information across test blocks. They showed that testing memory for picture recollections increased memory accuracy and decreased retrieval monitoring effort relative to word recollections in both younger and older adults (Gallo et al. 2004, 2007; Scimeca et al. 2011). Furthermore, 2 fMRI studies in younger adults found increased DLPFC activity when tested for word recollections relative to pictures recollections (Gallo et al. 2006, 2010), even after controlling for retrieval success effects, consistent with the idea that increasing retrieval monitoring demands are associated with greater DLPFC recruitment (cf. Cabeza et al. 2002). These DLPFC effects were attributed to the use of a diagnostic monitoring process, whereby participants rejected familiar lures because they failed to elicit recollections that matched their retrieval expectations. When the target source was more distinctive (i.e., pictures relative to words), diagnostic monitoring was less effortful and hence less likely to recruit DLPFC (i.e., a distinctiveness heuristic; see Schacter and Wiseman 2006; Gallo 2010).

Using the criterial recollection task in the present study allowed us to manipulate retrieval monitoring demands in both younger and older adults while controlling for retrieval success effects (see Fig. 1). At study, subjects viewed a list of words in black font, some of which were associated with a picture, some with a semantic judgment, and some with both (nonconsecutively). At test, recollection of the different kinds of encoded information was assessed in an event-related fMRI design. Words were used as retrieval cues, and the retrieval instructions were varied so that subjects had to recollect if the test word had been presented in the picture study condition (picture test blocks) or in the word study condition (word test blocks). Based on prior behavioral research using a similar version of this task, we assumed that retrieval monitoring would be more demanding or effortful on the word test than on the picture test in both age groups. Moreover, Gallo et al. (2007) found that criterial recollection performance was impaired on the word test in older relative to younger adults, but criterial recollection performance was more closely matched on the picture test. These test differences are
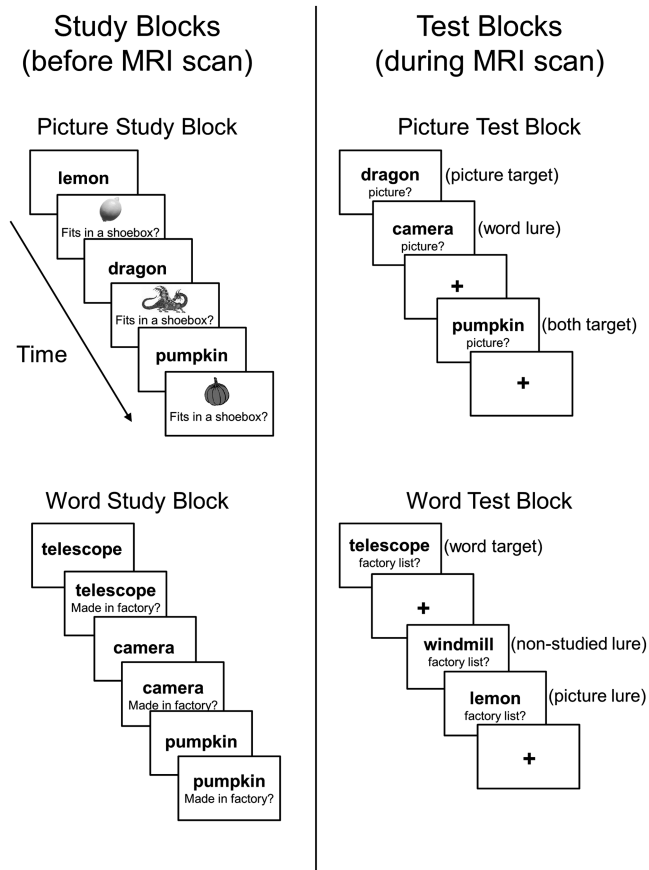
## Study Blocks
### (before MRI scan)

#### Picture Study Block

lemon
Fits in a shoebox?

dragon
Fits in a shoebox?

pumpkin
Fits in a shoebox?

Time

#### Word Study Block

telescope

telescope
Made in factory?

camera

camera
Made in factory?

pumpkin

pumpkin
Made in factory?

## Test Blocks
### (during MRI scan)

#### Picture Test Block

dragon (picture target)
picture?

camera (word lure)
picture?

+

pumpkin (both target)
picture?

+

#### Word Test Block

telescope (word target)
factory list?

+

windmill (non-studied lure)
factory list?

lemon (picture lure)
factory list?

+

**Figure 1.** Schematic of the criterial recollection task. During the study phase (left panel), words were associated with a semantic judgment and a picture (picture study blocks), or they were associated with a different semantic judgment and no picture (word study blocks). During the test phase (right panel), words from each study condition were intermixed on each test, and subjects responded whether they recollected previously seeing the item in the picture study condition (picture test blocks) or the word study condition (word test blocks).

theoretically important for the current study because they allowed us to investigate neural activity under conditions where age differences in retrieval success were naturally maximized (the word test) or minimized (the picture test), in addition to statistically controlling for such effects in the fMRI analysis. As described next, existing hypotheses make different age-related predictions under these 2 testing conditions.

Based on prior work, we expected that PFC activity would be greater on the word test than on the picture test in younger adults, owing to the increase in retrieval monitoring demands. The novel question was with respect to PFC activity in older adults, for which the compensation and dysfunction hypotheses make different predictions. According to the compensation hypothesis, age-related differences in PFC activity should be greatest in the most difficult testing conditions because these conditions are the most likely to require strategic changes or additional effort for older adults to optimize performance. Based on the behavioral work mentioned above, we expected larger age-related accuracy impairments on the word test than on the picture test, suggesting that older adults would need to differentially recruit more retrieval monitoring effort on the word test than on the picture test compared with younger adults. As a result, the word test should require more age-related compensatory processing than the picture test, leading to larger

differences in PFC across the 2 tests in older adults relative to younger adults.

In contrast to these compensation-related predictions, the dysfunction hypothesis predicts reduced differences in PFC regions between the word test and picture test in older adults relative to younger adults. In cognitively normal younger adults, PFC activity tracks retrieval monitoring demands (word test > picture test), suggesting that abnormal or dysfunctional PFC activity might not track retrieval monitoring demands as effectively, resulting in reduced differences in activity. For example, older adults might have difficulty recruiting effortful retrieval monitoring processes when they are required as well as difficulties inhibiting effortful strategies when they become suboptimal (Zacks et al. 2000; Logan et al. 2002). This dysfunction hypothesis does not deny the possibility that older adults may engage additional retrieval monitoring effort overall relative to younger adults, which could be considered a form of attempted compensation. Rather, this hypothesis focuses on the potential inability of older adults to selectively recruit PFC during more demanding retrieval monitoring situations, when such compensatory processing would be needed the most relative to less demanding situations.

## Materials and Methods

### Subjects
Twenty-one younger adults aged 18–30 years ($M$ = 21.2; 13 females) and 27 older adults aged 62-88 years ($M$ = 77.05; 18 females) participated for pay. Data from 1 younger adult and 4 older adults were discarded (2 to scanner/computer malfunctions, 1 to the subject aborting the session, 1 reported vision problems after the session, and 1 reported misunderstanding the instructions after the session). The older adults were recruited from the Chicago metropolitan area, lived independently, and had a high level of functioning as measured by the Mini-Mental State Examination ($M$ = 28.04; Folstein et al. 1975). All subjects were right-handed, and none reported neuropsychological conditions associated with cognitive decline (e.g., Alzheimer's disease, Parkinson's disease, etc.), taking excessive alcohol or narcotics, having a history of psychiatric diagnoses, or having recent head trauma. All subjects gave informed consent using methods approved by the appropriate human subjects committees at the University of Chicago. Vision was normal or corrected to normal using MR-compatible glasses or contact lenses.

### Materials and Design
The experiment consisted of a study phase outside the MRI scanner and a test phase inside the scanner. Study materials were pictures of 288 common objects (e.g., toaster, pig, clock) and corresponding verbal labels, with an average length of 6.56 letters and a Kucera-Francis written frequency of 35.86 per million. Pictures were presented on a white background with the object cropped from any surrounding context, and corresponding verbal labels were presented in black font. The study phase was divided into a picture block and a word block, with the order counterbalanced across subjects. Each block consisted of 144 stimuli in which 72 were presented in both study blocks ("both" items) and 72 in one format only (word-only or picture-only). An additional 72 stimuli were not studied and served as new items on the 2 subsequent memory tests. Stimuli were counterbalanced across the studied conditions, and presentation was randomized for each subject.

Two tests were given inside the scanner (picture test and word test) distributed across 3 functional runs lasting 10.1 min per run. Each run was subdivided into 2 test blocks separated by a minimum fixation of 10 s. The order of the tests was held constant across runs and was counterbalanced across subjects. Within each run, each test consisted of the 4 item types (12 picture-only, 12 word-only, 12 both, and 12 nonstudied) intermixed with jittered fixations. The order of items and fixations was determined using optseq program (part of the FS-FAST

analysis tools written by D. Greve, Charlestown, MA) to maximize the MR signal (e.g., Dale 1999). In total, there were 36 items of each type (pictures, words, both, and new) on each of the tests. On the picture test, picture-only items and both items served as targets, while word-only items and nonstudied items served as lures. On the word test, word-only items and both items served as targets, while picture-only items and nonstudied items served as lures. Both items were included so that recollection of the noncriterial source (e.g., picture lures on the word test) could not be used to reject studied lures.

### Procedure

Subjects were first given a practice study-test cycle using 24 stimuli not used in the actual experiment (10 min). Following the practice test, subjects studied the blocks of pictures and words. During the picture block, a verbal label in black font was presented in the center of the computer screen for 750 ms, followed 50 ms later by a corresponding picture for 2000 ms. After 2000 ms, the picture remained on the screen while a prompt appeared asking subjects whether the object represented in the picture typically fits inside a shoebox by pressing the "yes" or "no" key. The trial ended after the subject made the shoebox judgment or after 3000 ms elapsed. During the word block, a black word was presented in the center of the computer screen for 2750 ms. After 2750 ms, the word remained on the screen while a prompt appeared asking subjects whether the object was typically made in a factory by pressing the yes or no key. The trial ended after the subject made a factory judgment or after 3000 ms elapsed. A 100-ms interstimulus interval separated each trial on both tests.

Approximately 20 min after the end of the study phase, memory was tested inside the scanner. Verbal labels were presented for 4 s in black font in the center of the screen separated by a central fixation cross of jittered duration (2-12 s, mean stimulus onset asynchrony = 3.68 s). The onset of stimulus presentations was time-locked to the recording of the hemodynamic response. On the picture test, each word was paired with the prompt "picture?". Subjects were instructed to decide whether each item was shown during the picture study block and that remembering the item occurred during the word block was irrelevant on this test because some items were presented in both formats. On the word test, each word was paired with the prompt "factory judgment?". Subjects were instructed to decide whether each item was shown during the word study block in which they had made factory judgments. Subjects were also told that remembering that the item occurred during the picture block was irrelevant on this test because some items were presented in both formats. Test responses were made with the right hand using an MR-compatible button box such that the button for the index finger corresponded to yes responses and the button for the middle finger corresponded to no responses.

### fMRI Data Acquisition

Images were acquired using a 3-T Philips scanner at the University of Chicago Brain Research Imaging Center. Event-related functional scans were acquired using a $T_2^*$-weighted echo planar imaging sequence (repetition time [TR] = 2 s, echo time [TE] = 25 ms, field of view [FOV] = 224 mm; flip angle = 80°, matrix size = 64 × 63 mm, in-plane resolution = 3.5 × 3.5 mm$^2$). For whole-brain coverage, 32 interleaved slices (4 mm thickness, 0.5 mm skip between slices) were acquired. Prior to acquiring the 3 functional runs, we acquired 4 additional slices across orbital frontal region and applied a Z-shimming compensation gradient to regain signal loss due to nasal cavity artifact (Glover 1999). Structural scans were acquired last using high resolution $T_1$-weighted structural Turbo Field Echo (TR = 7.4 ms, TE = 3.4 ms, flip angle = 8°, FOV = 250 mm, matrix = 240 × 240 mm$^2$, in-plane resolution = 1.04 × 1.04 mm$^2$).

### fMRI Data Analysis

Preprocessing and data analysis were conducted using SPM8 (Wellcome Trust Centre for Neuroimaging, London). Standard preprocessing was performed on the functional data, including estimation of the realignment parameters using rigid body motion correction, anatomical coregistration, segmentation of the anatomical scan into gray matter, white matter, and cerebrospinal fluid, normalization to the Montreal

Neurological Institute template (resampling at 2 mm cubic voxels), and spatial smoothing (using a 10-mm full-width half-maximum isotropic Gaussian kernel).

For each subject, the BOLD response for each event type of interest was estimated using an unbiased whole-brain approach. This approach applied a voxel-by-voxel general linear model (GLM) that included temporal derivatives (Rombouts et al. 2005) and a high-pass filter of 128 s. The events were modeled using a mini-epoch that varied for each stimulus, starting at the stimulus onset and ending when subject responded. This method was selected because the retrieval processes of interest are thought to occur throughout these mini-epochs (cf. Grinband et al. 2008) (GLM analyses also were conducted using a stick function [duration = 0]. These results revealed similar patterns of activity including bilateral DLPFC [–42, 16, 46 and 48, 16, 44], but at lower $t$ values [$t$ < 3.81], suggesting less sensitivity to retrieval monitoring processes). Responses occurring after the offset of the stimulus or with multiple button presses were given a duration of the average response time value across all trials for that subject. Event types reflected a combination of the test condition (picture test and word test), item type (both, word, picture, and new), and the subjects' response (yes and no). Three runs were modeled as separate sessions, using rigid body motion parameters, outliers due to movement/signal spikes (Mazaika et al. 2005), and session effects as regressors. Group analyses were conducted by entering the first level contrasts of the correct responses into one-sample $t$-tests with subjects entered as random effects. To further characterize the BOLD response, we conducted region of interest (ROI) analyses using MarsBaR (Brett et al. 2002). ROI analyses extracted peak activity in the BOLD response relative to the mean signal from the ROI across the entire session and was done for each subject and each condition of interest within a 22 s window from stimulus onset, thus allowing assessment of the percent signal change associated with each condition.

## Behavioral Results

### Criterial Recollection Test Performance

We first analyzed memory performance for each age group separately and then directly compared age groups. Behavioral results can be found in Table 1. On the word test, younger adults responded yes more often to word targets (0.73) compared with picture lures (0.43), $t_{19}$ = 7.09, standard error of the mean (SEM) = 0.043, $P$ < 0.001, whereas on the picture test, younger adults responded yes more often to picture targets (0.71) compared with word lures (0.27), $t_{19}$ = 8.19, SEM = 0.053, $P$ < 0.001. These results indicate that younger adults appropriately adjusted their retrieval orientation on each test. Replicating prior work, younger adults committed fewer false recognition errors on the picture test (0.27) compared with the word test (0.43), $t_{19}$ = 3.73, SEM = 0.041, $P$ = 0.001, implicating reduced monitoring demands when retrieval was oriented toward higher quality picture recollections. Older adults also committed fewer false recognition errors on the picture test (0.34) compared with the word test (0.53), $t_{22}$ = 5.50, SEM = 0.035, $P$ < 0.001. On the word test, older adults numerically responded yes more often to word targets (0.57) compared with picture lures (0.53), but this difference was not significant, $t_{22}$ = 0.99, SEM = 0.045, $P$ = 0.33, whereas on the picture test, older adults responded yes significantly more often to picture targets (0.65) compared with word lures (0.34), $t_{22}$ = 6.20, SEM = 0.051, $P$ < 0.001. Consistent with prior work, the word test was very difficult for older adults relative to the picture test.

To directly compare the age groups, we calculated a source discrimination score (hits to items studied in the criterial source minus false alarms to items studied in the noncriterial source) and entered these data into a 2 (age: younger, older) × 2

**Table 1**

Mean proportion of items recognized, response latencies, and number of trials on the criterial recollection tests in younger and older adults

| | P, "yes" | | Latency correct responses (ms) | | Number of trials for correct responses | |
|---|---|---|---|---|---|---|
| | Younger | Older | Younger | Older | Younger | Older |
| Word test | | | | | | |
| Both targets | 0.79 (0.02) | 0.64 (0.04) | 1740 (63) | 1778 (80) | 27.3 (0.9) | 22.0 (1.4) |
| Word targets | 0.73 (0.02) | 0.58 (0.04) | 1654 (50) | 1796 (73) | 25.1 (0.8) | 19.8 (1.5) |
| Picture lures | 0.43 (0.04) | 0.53 (0.04) | 2012 (101) | 1963 (80) | 19.9 (1.3) | 16.2 (1.3) |
| New lures | 0.19 (0.04) | 0.33 (0.05) | 1711 (69) | 1773 (67) | 28.1 (1.6) | 23.3 (1.8) |
| Picture test | | | | | | |
| Both targets | 0.78 (0.03) | 0.72 (0.04) | 1692 (55) | 1782 (62) | 26.5 (1.1) | 25.0 (1.4) |
| Picture targets | 0.71 (0.03) | 0.65 (0.04) | 1673 (58) | 1857 (55) | 24.3 (1.1) | 22.7 (1.2) |
| Word lures | 0.27 (0.04) | 0.34 (0.04) | 1823 (84) | 2032 (74) | 24.9 (1.4) | 23.0 (1.5) |
| New lures | 0.10 (0.04) | 0.18 (0.03) | 1693 (62) | 1875 (63) | 30.5 (1.3) | 28.4 (1.2) |

Note: Standard errors of each mean are in parenthesis.

(test: word, picture) analysis of variance (ANOVA). A main effect of test indicated that discrimination was better on the picture test (0.38) than the word test (0.17), $F_{1,41} = 43.86$, Mean Square Error (MSE) = 0.020, $P < 0.001$. A main effect of age indicated that discrimination was better for younger adults (0.37) than older adults (0.18), $F_{1,41} = 9.67$, MSE = 0.080, $P < 0.001$. Critically, an age × test interaction, $F_{1,41} = 5.13$, MSE = 0.020, $P = 0.029$, indicated that discrimination was greater for younger compared with older adults on the word test, $t_{41} = 4.16$, SEM = 0.062, $P < 0.001$, but discrimination was more closely matched on the picture test, $t_{41} = 1.63$, SEM = 0.074, $P = 0.11$ (Similar behavioral effects were found in each of 3 functional runs, and an ANOVA including run as a factor revealed no main effect of run or interactions between run and age or test). These behavioral results replicate Gallo et al. (2007), showing an age-related decline in the ability to use fine-grained details to inform recollection decisions on the word test but minimal age differences in recollection accuracy on the picture test.

### Response Latencies

We next analyzed response latencies to correct responses (see Table 1), under the assumption that responses requiring more retrieval monitoring effort should have taken longer. Consistent with this assumption, we found that participants took longer to correctly reject studied lures relative to the other trials, likely because the familiarity of studied lures made them difficult to reject. Specifically, younger adults were slower to reject studied lures than to accept studied targets (2.01 vs. 1.65 s on the word test and 1.82 vs. 1.67 s on the picture test, both $Ps < 0.05$) and slower to reject studied lures than to reject nonstudied lures (2.01 vs. 1.71 s on the word test and 1.82 vs. 1.69 s on the picture test; $P = 0.003$ and $P = 0.055$ on the word and picture test, respectively). Likewise, older adults were slower to reject studied lures than to accept studied targets on each test (1.96 vs. 1.80 on the word test and 2.03 vs. 1.85 on the picture test, both $Ps < 0.05$) and were slower to reject studied lures than to reject nonstudied lures (1.96 vs. 1.77 s on the word test and 2.03 vs. 1.87 s on the picture test, both $Ps < 0.01$). These response patterns suggest that both age groups were engaging in retrieval monitoring to a greater extent when familiar, but noncriterial information was retrieved, and critically, these patterns were observed on each of the 2 tests. These results argue against a global strategy shift in older adults across the tests (i.e., using familiarity instead of recollection) and instead suggest that both age groups attempted to

recollect the criterial information when making their decisions on each of the 2 tests.

Across the 2 tests, we found that younger adults responded faster when correctly rejecting studied lures on the picture test (1.82 s) compared with the word test (2.01 s), $t_{19} = 2.77$, SEM = 0.068, $P = 0.012$, consistent with the idea that the picture test required less demanding retrieval monitoring than the word test. In contrast, response latencies for rejecting studied lures did not differ across the 2 tests in older adults (2.03 s on the picture test vs. 1.96 s on the word test; $P = 0.20$). Even though false recognition was lower on the picture test relative to the word test—suggesting that the word test placed more demands on retrieval monitoring than the picture test—older adults took an equal amount of time to reject studied lures across the 2 tests, potentially because they exerted a high degree of retrieval monitoring effort on both tests. We will return to this possibility after our presentation of the fMRI results.

To summarize these behavioral results, the 2 age groups were matched on their ability to recollect pictures on the picture test, but older adults were impaired in their ability to recollect the words on the word test. These results replicated the well-established age × distinctiveness interaction on memory accuracy (see Gallo et al. 2007). Moreover, even though older adults easily discriminated between studied and nonstudied items on each test (i.e., old/new recognition), the results from the word test replicated the typical age-related deficits in the recollection of more fine-grained information (e.g., Ferguson et al. 1992; Johnson et al. 1995). Importantly, although older adults were impaired in their ability to recollect the studied words, the response latencies indicated that both groups engaged in qualitatively similar retrieval monitoring processes. Participants in both age groups took longer to reject studied lures than nonstudied lures on each of the 2 tests, exactly as one would expect if they were basing their decisions for studied items on an effortful retrieval search (as opposed to familiarity-based responding). Both groups were following the test instructions and were attempting to recollect and monitor the criterial information.

### Neuroimaging Results

We report 3 sets of fMRI analyses. First, we directly contrasted correct responses on the 2 recollection tests to identify regions that were sensitive to retrieval monitoring demands (word test > picture test). These contrasts used an unbiased whole-brain approach and a threshold that maximized our ability to detect age-related differences in activity ($P < 0.005$, uncorrected, 10

contiguous voxel extent; cf. Lieberman and Cunningham 2009). These contrasts were accompanied by a conjunction analysis to control for retrieval success effects. Second, independent from the results of these statistical contrasts, we conducted an anatomically defined ROI analysis on the middle frontal gyrus (MFG), given that this area has consistently been associated with retrieval monitoring in this task as well as others (Gallo et al. 2006, 2010; for earlier review, see Rugg 2004). For this ROI analysis, we estimated activity associated with correct rejections of studied lures relative to nonstudied lures on each of the 2 tests. In addition to the expected difference in retrieval monitoring demands across the tests (word test > picture test), rejecting studied lures should have been more demanding than rejecting nonstudied lures within each test because the studied lures were more likely to elicit noncriterial recollection and familiarity. These analyses therefore allowed us to assess potential retrieval monitoring effects when criterial recollection performance differed between age groups (word test) and when it was matched (picture test). Last, to further explore the activity in the right DLPFC region that was associated with retrieval monitoring in our younger adult contrasts, we correlated criterial recollection performance with BOLD activity across the individuals in each age group.

### Cross-test Monitoring Effects

We first aimed to isolate regions associated with demanding retrieval monitoring by contrasting correct responses to studied items on the word test to these same items on the picture test (collapsing across word and picture study status), yielding an average of 46 observations per test for each subject (range = 32–61 observations on the word test and 31–64 observations on the picture test in younger adults; range = 27–52 observations on the word test and 32–61 observations on the picture test in older adults). This contrast holds item type constant across the tests while varying the retrieval monitoring demands (word test > picture test). As can be seen in Table 2 and Figure 2, this contrast yielded voxels with significantly more activity on the word test than on the picture test across many brain regions in younger adults. These regions collectively resembled the "executive network" or "frontoparietal control system" found in resting-state connectivity thought to underlie cognitively controlled processing (e.g., Seeley et al. 2007; Dosenbach et al. 2008). Critically, we found significant activity in DLPFC, including bilateral MFG near Brodmann area (BA) 8 and a more anterior area in left MFG (BA 9/46). These results replicate prior work in younger adults showing activity in these same DLPFC regions during demanding retrieval monitoring (Gallo et al. 2006, 2010). We also found activity in other regions consistent with previous

studies of episodic memory retrieval and cognitive control (e.g., Henson et al. 1999; Cansino et al. 2002; Dobbins et al. 2002; Donaldson et al. 2010) including bilateral ventrolateral PFC, left anterior PFC (BA 10), bilateral dorsomedial PFC (BA 6/8), bilateral anterior cingulate cortex (BA 32), lateral parietal, and medial parietal regions. The reverse contrast (picture test > word test) did not reveal activity near any of these regions in younger adults, consistent with the idea that activity in these regions was greater on the test that required more demanding retrieval monitoring processes.

In stark contrast to the findings in younger adults, the same contrast in older adults revealed no voxels that were more active on the word test relative to the picture test (Table 3), even though the older adults (like younger adults) were less accurate on the word test than on the picture test. These age-related differences between activity on the word test and picture test suggest that aging might alter DLPFC activity associated with more demanding retrieval monitoring. On the reverse contrast (picture test > word test), we found voxels with significant activity in DLPFC including right MFG (BA 8), a more anterior region of right MFG (near BA 9/10), right superior frontal gyrus (BA 6), and left superior frontal gyrus (BA 9) and a few other prefrontal regions. Additionally, several clusters in posterior regions showed significant activity

**Table 2**

Peak coordinates of activity for comparisons of studied items between the recollection tests in younger adults

| MNI coordinates (x, y, z) | Cluster size (voxels) | T score | Region | BA |
|---|---|---|---|---|
| Word test > picture test | | | | |
| −26, 62, 14 | 313 | 4.27 | L superior frontal gyrus | 10 |
| −26, 56, 36 | 148 | 4.34 | L superior frontal gyrus | 9 |
| −46, 38, 28 | 30 | 3.28 | L MFG | 9/46 |
| −2, 34, 44 | 168 | 4.13 | L medial frontal gyrus | 8 |
| −2, 30, −10 | 102 | 4.55 | L anterior cingulate | 32 |
| 26, 26, 40 | 253 | 3.99 | R MFG | 8 |
| −44, 18, 48 | 95 | 4.42 | L MFG | 8 |
| 36, 18, −12 | 121 | 4.62 | R inferior frontal gyrus | 47 |
| −38, 16, −10 | 728 | 4.82 | L inferior frontal gyrus | 47 |
| −12, 12, 66 | 53 | 3.97 | L superior frontal gyrus | 6 |
| 46, 12, 46 | 223 | 4.45 | R MFG | 8 |
| −50, 10, 8 | 32 | 3.33 | L precentral gyrus | 44 |
| 6, −40, 22 | 129 | 4.23 | R posterior cingulate | 23 |
| 60, −50, 8 | 21 | 3.11 | R middle temporal gyrus | 21 |
| 48, −52, 44 | 22 | 3.35 | R inferior parietal lobule | 40 |
| −44, −56, 34 | 79 | 3.84 | L inferior parietal lobule | 40 |
| 6, −72, 40 | 283 | 4.19 | R precuneus | 7 |
| 48, −80, −6 | 14 | 3.02 | R inferior occipital gyrus | 18 |
| Picture test > word test | | | | |
| −28, −10, −28 | 16 | 3.27 | L parahippocampal gyrus | |

Note: Coordinates are the peak activation within a cluster, arranged anterior to posterior and laterally (R = right, L = left). BA = approximate Brodmann's areas.
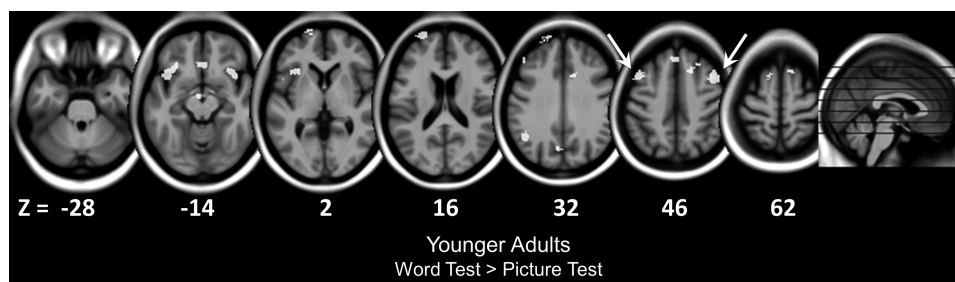


**Figure 2.** Axial slices illustrating activity observed on the word test > picture test contrast in younger adults. Younger adults recruited several PFC regions as a function of the retrieval monitoring demands (word test > picture test), but older adults did not reveal any significant activity in this contrast. Arrows highlight DLPFC activity.

**Table 3**

Peak coordinates of activity for comparisons of studied items between the recollection tests in older adults

| MNI coordinates (x, y, z) | Cluster size (voxels) | T score | Region | BA |
|---|---|---|---|---|
| Word test > picture test | | | | |
| No significant voxels | | | | |
| Picture test > word test | | | | |
| −20, 36, 34 | 30 | 3.20 | L superior frontal gyrus | 9 |
| 18, 36, 30 | 52 | 3.48 | R medial frontal gyrus | 9 |
| 32, 34, 18 | 17 | 3.21 | R MFG | 9/10 |
| 10, 24, 56 | 21 | 3.36 | R superior frontal gyrus | 6 |
| 30, 10, 42 | 20 | 3.14 | R MFG | 8 |
| 12, −8, 72 | 32 | 3.41 | R superior frontal gyrus | 6 |
| 28, −16, 72 | 156 | 4.68 | R precentral gyrus | 6 |
| 38, −18, 60 | 61 | 3.36 | R precentral gyrus | 4 |
| −26, −20, 74 | 22 | 3.25 | L precentral gyrus | 6 |
| −44, −26, 64 | 645 | 4.81 | L postcentral gyrus | 3 |
| 40, −32, 64 | 43 | 3.31 | R postcentral gyrus | 3 |
| −24, −56, 64 | 20 | 3.43 | L superior parietal lobule | 7 |
| 14, −70, 26 | 29 | 3.07 | R precuneus | 31 |

Note: Coordinates are the peak activation within a cluster, arranged anterior to posterior and laterally (R = right, L = left). BA = approximate Brodmann's areas.

**Table 4**

Center of mass coordinates of activity for monitoring conjunction in younger and older adults

| MNI coordinates (x, y, z) | Cluster size (voxels) | Region | BA |
|---|---|---|---|
| Monitoring conjunction for word test > picture test in younger adults | | | |
| −29, 60, 12 | 137 | L middle/superior frontal gyrus | 10 |
| −12, 51, 25 | 60 | L superior frontal gyrus | 9 |
| − 6, 36, 39 | 150 | L medial frontal gyrus | 6 |
| −39, 23, −8 | 289 | L inferior frontal gyrus | 47 |
| 35, 21, −9 | 85 | R inferior frontal gyrus | 47 |
| 47, 10, 45 | 14 | R MFG | 8 |
| 58, −20, −9 | 64 | R middle temporal gyrus | 21 |
| 59, −43, 2 | 145 | R middle temporal gyrus | 22 |
| −46, −57, 28 | 104 | L angular gyrus | 39 |
| −1, −66, 38 | 594 | L precuneus | 7 |
| Monitoring conjunction for word test > picture test in older adults | | | |
| No significant overlapping voxels | | | |

Note: Coordinates are the center of mass within a cluster, arranged anterior to posterior and laterally (R = right, L = left). CR = correct rejections. BA = approximate Brodmann's areas.

including left superior parietal lobule and right precuneus. Although this activity on the reverse contrast may reflect more effortful retrieval monitoring on the picture test relative to the word test in older adults, these contrasts also might have been sensitive to the relatively larger differences in retrieval success across the 2 recollection tests in older adults. Consistent with this retrieval success interpretation, follow-up analyses of the picture test showed that studied targets (pictures) relative to studied lures (words) were more likely than nonstudied lures to activate prefrontal regions, including right MFG. Because our main goal was to identify activity associated with retrieval monitoring processes, independent of retrieval success effects, we next turn to analyses that more directly targeted retrieval monitoring processes.

We conducted a conjunction analysis to isolate activity associated with retrieval monitoring while controlling for retrieval success. For this analysis, we focused on the correct rejections of studied lures on the word test because as discussed these items in particular should have required effortful retrieval monitoring to reject. In the conjunction, we identified regions that overlapped between 2 different contrasts. The first contrast compared the rejection of studied lures on the word test with the picture test. This contrast is associated with more demanding retrieval monitoring across the 2 tests (word test > picture test) but also potentially varies retrieval success (i.e., the lures on the word test had been studied with pictures). To control for the possibility that rejecting studied lures on the word test involved picture recollections, the second contrast identified voxels that were more active when rejecting studied lures on the word test compared with correctly accepting targets on the picture test (i.e., items that were clearly associated with picture recollections). Although the contrasts in these analyses were not independent, controlling for retrieval success in this way is theoretically important because DLPFC activity is often attributed to both retrieval monitoring and retrieval success. Because a conjunction analysis is more conservative than a simple contrast, we used a more liberal threshold for each contrast that contributed to the conjunction analysis to avoid Type II error ($P < 0.01$, uncorrected, 5 contiguous voxel extent).

As seen in Table 4, this conjunction analysis in younger adults revealed overlapping voxels in DLPFC regions including right MFG (BA 8) and left superior frontal gyrus (BA 9). Other

prefrontal regions included bilateral ventrolateral PFC (BA 47), left anterior PFC spanning middle and superior frontal gyrus (BA 10), and bilateral dorsomedial PFC (BA 6/8/9). Overlapping voxels in posterior regions included left angular gyrus (BA 39), left precuneus (BA 31), and right middle temporal gyrus (BA 21/22). No overlapping voxels were found in older adults, consistent with the lack of significant activity in the cross-test comparison above (word test > picture test).

To summarize these contrasts, younger adults selectively recruited several PFC regions (including right posterior DLPFC) as a function of retrieval monitoring demands (word test > picture test), replicating prior work. In contrast, older adults did not differentially activate these PFC regions as a function of test, despite the fact that the word test should have been even more demanding than the picture test in older adults relative to younger adults. These findings clearly demonstrate that PFC activity was more sensitive to retrieval monitoring demands in younger adults than in older adults, consistent with the dysfunction hypothesis. In contrast, none of these regions showed greater test differences (word test > picture test) in older adults relative to younger adults, providing no support for the predictions of the compensation hypothesis.

### Anatomical ROI Analyses

As an additional way to investigate PFC activity associated with retrieval monitoring, we created anatomically defined ROIs in the right MFG, which is inclusive of DLPFC. For this analysis, we split the right MFG from the AAL library (Tzourio-Mazoyer et al. 2002) into anterior and posterior regions and then extracted percent signal change for the different kinds of items on each test for younger and older adults. Given the results of our contrasts in the previous section, we expected that activity in the posterior MFG (which includes the DLPFC region) would show similar word test > picture test effects in younger adults but not in older adults. However, these ROI analyses further allowed us to characterize the magnitude of the BOLD signal across the different test and item conditions, which is more informative than simple contrasts alone and allows for additional comparisons.

To analyze these ROIs for test effects (word test > picture test), the BOLD signal for correct responses to studied items on each test was averaged and entered into a 2 (age: younger, older) × 2 (test: word, picture) ANOVA (see Fig. 3). For

posterior MFG, there was a main effect of age, $F_{1,41} = 17.52$, MSE = 0.010, $P < 0.001$, indicating that overall activity was greater in older than younger adults, and this effect was qualified by an age × test interaction, $F_{1,41} = 7.48$, MSE = 0.004, $P = 0.009$, indicating that while activity was greater on the word test than the picture test in younger adults, $t_{19} = 2.68$, SEM = 0.014, $P = 0.015$, activity between the 2 tests did not differ in older adults, $t_{22} = 1.59$, SEM = 0.021, $P = 0.13$. For anterior MFG, the ANOVA revealed only a marginal main effect of age, $F_{1,41} = 3.30$, MSE = 0.009, $P = 0.08$, as older adults had greater activity than younger adults, but no effect of test or significant interaction (all $Ps > 0.24$), suggesting that this region was less critical for retrieval monitoring than posterior MFG. Overall, these analyses illustrated the expected test effect in posterior MFG in younger adults but not in anterior MFG, and they also show that the lack of this effect in older adults was associated with overall high levels of activity on each of the tests.

We also analyzed these ROIs for potential retrieval monitoring effects within each test, comparing the BOLD signal for correct rejections to studied lures to correct rejections to nonstudied lures (Fig. 4). Unlike our conjunction analysis, these particular ROI comparisons could be affected by differences in retrieval success (studied lures > nonstudied lures) as well as associated differences in retrieval monitoring. Nevertheless, to the extent that these MFG regions are sensitive to retrieval monitoring demands, they should have been more active when rejecting studied lures relative to nonstudied lures because subjects had to monitor the additional familiarity of studied lures. We first describe activity on the word test, where accuracy was greater in younger than older adults and then describe activity on the picture test, where accuracy was more closely matched between the age groups.

For the word test, a 2 (age: younger and older) × 2 (lure: studied and nonstudied) ANOVA in posterior MFG revealed a main effect of lure, $F_{1,41} = 5.98$, MSE = 0.007, $P = 0.033$, indicating that activity was greater for correct rejections to studied lures compared with nonstudied lures, a main effect of age, $F_{1,41} = 9.21$, MSE = 0.014, $P = 0.004$, indicating that activity
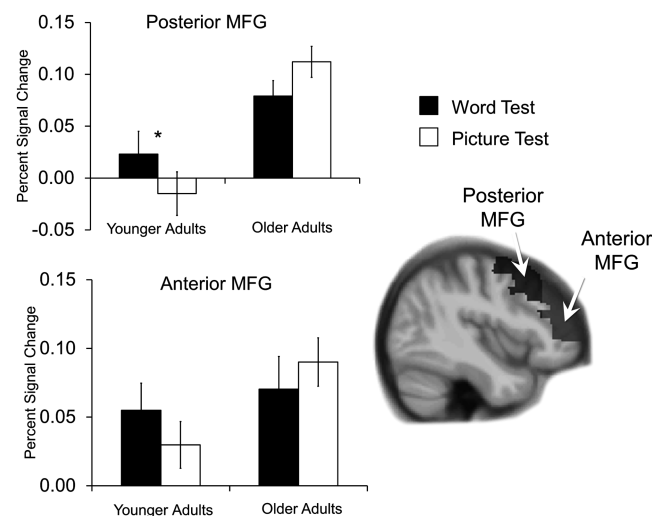
was greater for older than younger adults, and a critical age × lure interaction, $F_{1,41} = 4.85$, MSE = 0.007, $P = 0.033$, indicating that activity was only greater for correct rejections to studied compared with nonstudied lures in younger adults ($t_{19} = 2.77$, SEM = 0.031, $P = 0.012$ and $t_{22} < 1$, $P = 0.84$, for younger and older adults, respectively). Consistent with our cross-test analysis, this analyses revealed that activity in posterior MFG tracked the expected retrieval monitoring demands of the test items (studied lures > nonstudied lures) only in younger adults and that older adults instead showed high levels of activity for each item type. For anterior MFG, a 2 (age: younger, older) × 2 (lure: studied, nonstudied) ANOVA on the word test revealed a main effect of lure, $F_{1,41} = 4.80$, MSE = 0.011, $P = 0.034$, indicating that activity was greater for correct rejections to studied lures compared with nonstudied lures, but no main effect of age or interaction (all $Ps > 0.29$). However, follow-up $t$-tests indicated that the effect of lure was weak in each age group ($t_{19} = 1.35$, SEM = 0.038, $P = 0.19$ and $t_{22} = 1.83$, SEM = 1.83, $P = 0.081$, for younger and older adults, respectively), again suggesting that this region was not as critical for retrieval monitoring as posterior MFG.

For the picture test, a 2 (age: younger, older) × 2 (lure: studied, nonstudied) ANOVA in posterior MFG revealed a main effect of lure, $F_{1,41} = 9.89$, MSE = 0.004, $P = 0.003$, indicating that activity was greater for correct rejections to studied lures compared with nonstudied lures, a main effect of age, $F_{1,41} =$
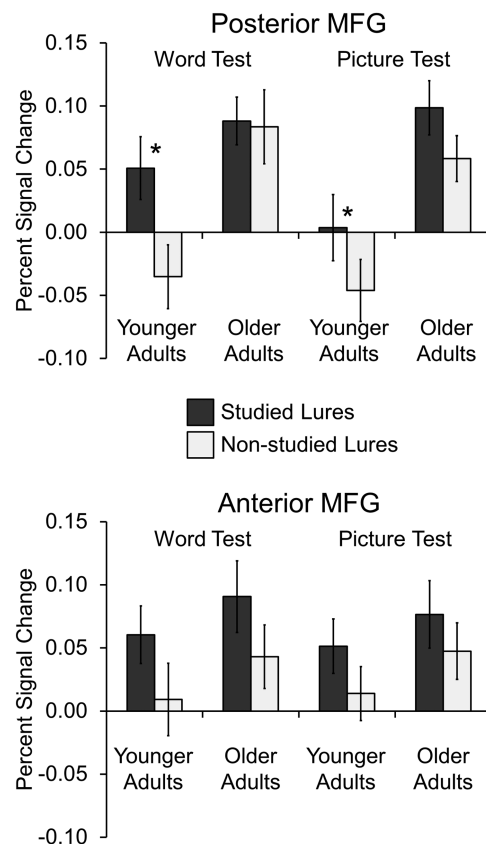


**Figure 3.** Percent signal change in right MFG observed during correct responses to studied items on the word test (dark bars) and picture test (light bars). Posterior MFG showed an age × test interaction, as monitoring demands modulated activity in the expected direction (word test > picture test) only in younger adults. In anterior MFG, only a marginal effect of age was found. Standard error of the mean is represented in the error bars. Asterisks indicate $t$-test significance ($P < 0.05$).



**Figure 4.** Percent signal change from MFG observed during correct rejections to studied lures (dark bars) and nonstudied lures (light bars) on each test. In posterior MFG, analyses on the word test revealed an age × lure interaction, indicating that lure difficulty modulated activity (studied lure > nonstudied lure) only in younger adults. The picture test showed a similar pattern, but the interaction was not significant. By contrast, anterior MFG showed lure effects but no age interactions. Standard error of the mean is represented in the error bars. Asterisks indicate $t$-test significance ($P < 0.05$).

12.23, MSE = 0.017, $P$ = 0.001, indicating that activity was greater for older than younger adults, but no age × lure interaction, $F_{1,41}$ < 1, $P$ = 0.74. Although the interaction was not significant, follow-up $t$-tests indicated that the effect of lure was significant in younger adults but only marginally significant in older adults ($t_{19}$ = 2.67, SEM = 0.019, $P$ = 0.015 and $t_{22}$ = 1.90, SEM = 0.021, $P$ = 0.071, for younger and older adults, respectively). For anterior MFG, a 2 (age: younger, older) × 2 (lure: studied, nonstudied) ANOVA on the picture test revealed a marginal main effect of lure, $F_{1,41}$ = 3.75, MSE = 0.006, $P$ = 0.060, indicating that activity was greater for correct rejections to studied lures compared with nonstudied lures, but no main effect of age or interaction (all $P$s > 0.30). Follow-up $t$-tests again indicated that the effect of lure was relatively weak in each age group in anterior MFG ($t_{19}$ = 1.92, SEM = 0.020, $P$ = 0.07 and $t_{22}$ = 1.07, SEM = 0.027, $P$ = 0.30, for younger and older adults, respectively), again suggesting that this region was less critical for retrieval monitoring.

To summarize these ROI analyses, younger adults but not older adults exhibited the expected word test > picture test pattern in posterior MFG. The same pattern was not found in anterior MFG. Younger adults also exhibited the predicted studied lure > nonstudied lure pattern in posterior MFG, on both the word and the picture tests, but older adults did not show this effect as reliably. Older adults instead showed elevated levels of activity relative to younger adults in all conditions, demonstrating relative insensitivity of the MFG to retrieval monitoring demands. These ROI analyses are consistent with the results of the whole-brain contrasts and further implicate age-related differences in DLPFC activity as a function of retrieval monitoring demands (We also reran the posterior MFG analyses after statistically equating performance on the word test across the 2 age groups, by taking out an equal number of subjects from each group [remaining $N$ = 31]. Even with this reduced data set, the critical findings were maintained. Specifically, the test × age interaction persisted, $F_{1,30}$ = 4.19, MSE = 0.004, $P$ = .05, such that activity was greater for the word test than the picture test for studied items in younger adults, $t_{14}$ = 2.20, SEM = 0.018, $P$ = 0.045, but not in older adults, $t_{16}$ < 1, $P$ > 0.33. With respect to the within-test effects, there again was an age × lure interaction on the word test, $F_{1,30}$ = 4.43, MSE = 0.009, $P$ = 0.044, such that activity was greater for correct rejections to studied lures than nonstudied lures in young adults, $t_{14}$ = 2.76, SEM = 0.039, $P$ = 0.015, but not in older adults, $t_{16}$ < 1, $P$ > 0.72. These findings further indicate that the age-related differences in DLPFC were not an artifact of different levels of performance on the word test but instead were due to age differences in the recruitment of DLPFC as a function of retrieval monitoring demands).

### Individual Differences Correlations

To further characterize the relationship between brain activity and behavior in right DLPFC, we calculated correlations between DLPFC activity and recollection test activity within each age group (Fig. 5). For this analysis, we used a 5-mm sphere ROI based on the peak coordinates of the right DLPFC region observed in the retrieval monitoring conjunction in younger adults, under the assumption that this region would be most likely to show a relationship to accuracy in younger adults. Specifically, criterial recollection performance (hits to criterial targets minus false alarms to studied lures) was correlated with the average BOLD activity associated with

correct responses to these same items (hits to criterial targets and correct rejections to studied lures) across subjects.

For younger adults, it could be assumed that increases in retrieval monitoring effort—as potentially measured by BOLD activity in DLPFC—would lead to improved criterial recollection performance, especially on the word test when retrieval monitoring demands were greatest (cf. Cruse and Wilding 2009). Consistent with this assumption, criterial recollection performance was positively correlated with BOLD activity on both the word test and the picture test in younger adults, although this correlation was only significant on the word test ($r_{20}$ = +0.57, $P$ = 0.009 and $r_{20}$ = +0.31, $P$ = 0.18). These results are consistent with the idea that DLPFC recruitment can enhance retrieval monitoring accuracy, especially when the test places a high demand on retrieval monitoring (i.e., the word test). If older adults had recruited this region in a similar manner as younger adults, then a positive correlation also should have been observed on the word test in older adults. However, unlike younger adults, older adults showed no relationship between criterial recollection performance and DLPFC activity on the word test ($r_{23}$ = +0.02, $P$ = 0.94), whereas on the picture test, older adults showed a "negative" correlation between criterial recollection performance and DLPFC activity ($r_{23}$ = −0.42, $P$ = 0.047). These correlations provide further evidence of altered age-related PFC activity during retrieval monitoring, consistent with the idea that these age-related differences reflect dysfunctional recruitment of these regions as opposed to compensation (Given that prior work has implicated homologous contralateral regions as compensatory, we also calculated the same criterial recollection performance correlations with activity in left DLPFC found on the word test > picture test contrast in younger adults [−46, 16, 48]. Contrary to a compensatory role, this region showed the same correlation patterns as in right DLPFC but to a lesser extent [$r_{20}$ = 0.45, $P$ = 0.047 and $r_{20}$ = 0.34, $P$ = 0.15 in younger adults on the word and picture test, respectively; $r_{23}$ = 0.11, $P$ = 0.60 and $r_{23}$ = −0.24, $P$ = 0.26 in older adults on the word and picture test, respectively]).
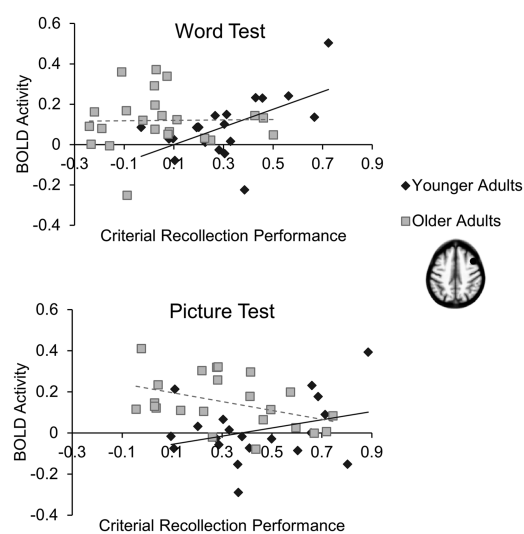


**Figure 5.** Correlations between BOLD activity in right DLPFC ($x$-axis), averaged across correct responses for studied items, and criterial recollection performance ($y$-axis), measured as criterial hits minus studied false alarms. Younger adults are shown in the diamonds (trend in solid lines), and older adults are shown in squares (trend in dotted lines). A significant positive correlation was found only on the word test in younger adults. A significant negative correlation was found only on the picture test in older adults.

## General Discussion

We investigated age-related differences in neural activity associated with retrieval monitoring demands. While both younger and older adults showed reduced false recognition when searching memory for pictures relative to words, only younger adults showed activity in prefrontal regions that was modulated by the retrieval monitoring demands of the tests (word test > picture test). Of particular interest was the activity found in right posterior DLPFC that has been associated with retrieval monitoring in several prior studies (e.g., Gallo et al. 2010; see Rugg 2004). The same analyses in older adults failed to reveal any prefrontal regions with activity that was modulated by the retrieval monitoring demands of the tests (word test > picture test). Older adults instead demonstrated elevated activity in right DLPFC across the conditions, and this pattern was observed not only when criterial recollection performance differed between the 2 age groups (the word test) but also when criterial recollection performance was matched across the 2 groups (the picture test). We also found that activity in right posterior DLPFC more consistently tracked the retrieval monitoring demands of the different test items (correct rejections to studied lures > nonstudied lures) in younger adults compared with older adults. Finally, while DLPFC activity was positively correlated with criterial recollection performance on the more demanding word test in younger adults, DLPFC activity was not correlated with criterial recollection performance on this test in older adults.

When considered in light of the hypotheses proposed in the Introduction, these patterns of activity are more consistent with an age-related dysfunction of PFC regions during retrieval monitoring, as opposed to a hypothesis based entirely on compensation. According to the compensation hypothesis of PFC activity, older adults should have been more likely than younger adults to recruit compensatory processes on the word test, where accuracy was the lowest and hence retrieval monitoring was most demanding, relative to the picture test. These patterns of activity were not observed in the PFC, and unlike younger adults, older adults showed more activity on the picture test than on the word test in some regions. For the same reason, a compensation hypothesis would have predicted greater PFC activity for correct rejections of studied lures than for nonstudied lures in older adults relative to younger adults because studied lures should have been harder for older adults to reject. This pattern was not found. Instead, right PFC activity was less likely to differentiate between the more demanding items in older adults than in younger adults. These findings suggest that older adults recruited DLPFC to the same degree regardless of the retrieval demands of the different conditions, owing to neural dysfunction. These results are consistent with the results of some prior studies of memory retrieval (e.g., Gutchess et al. 2007; Morcom et al. 2007; Duarte et al. 2008, 2010; Duverne et al. 2008; Rajah et al. 2010), which also found age-related differences in DLPFC activity, and they suggest that these prior patterns may have reflected dysfunction of the retrieval monitoring process.

The finding that older adults did not recruit DLPFC regions more heavily during demanding retrieval monitoring situations might represent the neural underpinnings of limited processing resources in aging (cf. Craik and Byrd 1982; Craik 1983). For example, according to the CRUNCH model (compensation-related utilization of neural circuits, see Reuter-Lorenz and Cappell 2008), age-related processing inefficiencies cause older adults to exert more effort than younger adults to achieve similar levels of performance on easy tasks, but as task difficulty increases, a resource ceiling is eventually reached and older adults are unable to benefit further. In the current study, if older adults expended all of their processing resources in the least demanding conditions, then they could not expend additional processing resources in conditions associated with greater monitoring demands. To the extent that DLPFC activity is associated with processing resources, this resource ceiling could explain why older adults failed to recruit DLPFC more heavily in those conditions that should have been more demanding. Furthermore, a resource ceiling in older adults may have limited our ability to find a significant association between BOLD activity and criterial recollection performance on the word test. Similar arguments have been made elsewhere when older adults failed to show modulation of activity across working memory and episodic memory tasks varying in processing demands (e.g., Mattay et al. 2006; Morcom et al. 2007; Reuter-Lorenz and Cappell 2008; Rajah et al. 2010). It should be noted, though, that the currents study was not specifically designed to test this resource ceiling account. Future studies that parametrically modulate the familiarity of studied lures, which would be more analogous to the working memory load manipulations used in studies that have tested the CRUNCH model (Nagel et al. 2009; Cappell et al. 2010; Schneider-Garces et al. 2010) would provide a more direct test of this resource ceiling interpretation in the domain of episodic memory retrieval.

Given the evidence for age-related dysfunction that we observed in association with retrieval monitoring, how were older adults able to achieve such a high level of recollection performance on the picture test? One possibility is that the high quality of picture recollections mitigated the need for effortful retrieval monitoring on the picture test in either age group, consistent with the false recognition literature (Schacter et al. 1999; Gallo et al. 2007). In this case, the increased activity in DLPFC on the picture test observed in older adults, relative to younger adults, may have been due to increased noise or interference owing to neural dysfunction. While speculative, this interpretation also may explain the negative correlation between picture test accuracy and DLPFC activity observed in older adults, to the extent that older adults with the worst performance also were the most likely to show dysfunctional recruitment of DLPFC on this test. This interpretation stands in contrast to a purely compensatory account of DLPFC activity, which would have predicted a positive correlation between DLPFC activity and memory accuracy in older adults, with all other factors being equal (see Reuter-Lorenz et al. 2000).

It is important to note that this dysfunctional interpretation of the observed DLPFC activity in older adults does not rule out the possibility that older adults had attempted to engage in compensatory processes altogether, as both dysfunction and attempted compensation are likely to be associated with aging. Indeed, to the extent that compensation during episodic retrieval is conceptualized as the recruitment of additional retrieval monitoring effort (e.g., Cabeza et al. 2002), the overall elevated activity in older adults relative to younger adults that we observed in some PFC regions might be partly due to this form of attempted compensation. Nevertheless, while these age-related increases in overall activity could be attributed to either compensation or to dysfunction, the failure of older adults to show the same task-based modulations of DLPFC activity as did

younger adults more clearly points to dysfunction of these regions.

In conclusion, we found that the quality of the to-be-recollected information affected the accuracy of retrieval monitoring in both age groups but also led to a different pattern of underlying neural activity. Younger adults recruited prefrontal regions most heavily when retrieval monitoring demands were greatest, whereas older adults showed elevated levels of activity compared with younger adults that did not differ with retrieval monitoring demands. Moreover, older adults demonstrated elevated activity when criterial recollection performance was impaired relative to younger adults (the word test) and also when criterial recollection performance was closely matched with younger adults (the picture test). These and other findings reported here are inconsistent with accounts of age-related differences in fMRI activity that are based entirely on compensation and instead provide strong evidence for age-related dysfunction of regions in the PFC during retrieval monitoring.

## Funding

## Notes

*Conflict of Interest*: None declared.

## References

Baltes PB, Lindenberger U. 1997. Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? Psychol Aging. 12:12-21.

Brett M, Anton JL, Valabregue R, Poline JB. 2002. Region of interest analysis using an SPM toolbox. Neuroimage. 16:497.

Buckner RL. 2004. Memory and executive function in aging and AD: multiple factors that cause decline and reserve factors that compensate. Neuron. 4:195-208.

Buckner RL, Snyder AZ, Sanders AL, Raichle ME, Morris JC. 2000. Functional brain imaging of young, nondemented, and demented older adults. J Cogn Neurosci. 12:24-34.

Butler K, McDaniel MA, Dornburg CC, Price AL, Roediger HL III. 2004. Age differences in veridical and false recall are not inevitable: the role of frontal lobe function. Psychon Bull Rev. 11:921-925.

Cabeza R. 2002. Hemispheric asymmetry reduction in old adults: the HAROLD model. Psychol Aging. 17:85-100.

Cabeza R, Anderson ND, Locantore JK, McIntosh AR. 2002. Aging gracefully: compensatory brain activity in high-performing older adults. Neuroimage. 17:1394-1402.

Cansino S, Maquet P, Dolan RJ, Rugg MD. 2002. Brain activity underlying encoding and retrieval of source memory. Cereb Cortex. 12:1048-1056.

Cappell KA, Gmeindl L, Reuter-Lorenz PA. 2010. Age differences in prefrontal recruitment during verbal working memory maintenance depend on memory load. Cortex. 46:462-473.

Carp J, Park J, Polk TA, Park DC. 2011. Age differences in neural distinctiveness revealed by multi-voxel pattern analysis. Neuroimage. 56:736-743.

Craik FIM. 1983. On the transfer of information from temporary to permanent memory. Philos Trans R Soc Lond. B302:341-359.

Craik FIM, Byrd M. 1982. Aging and cognitive deficits: the role of attentional resources. In: Craik FIM, Trehub S, editors. Aging and cognitive processes. New York: Plenum. p. 191-211.

Cruse D, Wilding EL. 2009. Prefrontal contributions to episodic retrieval monitoring and evaluation. Neuropsychologia. 47:2779-2789.

Curran T, Schacter DL, Norman KA, Galluccio L. 1997. False recognition after a right frontal lobe infarction: memory for general and specific information. Neuropsychologia. 35:1035-1049.

Dale AM. 1999. Optimal experimental design for event-related fMRI. Hum Brain Mapp. 8:109-114.

Davis SW, Dennis NA, Daselaar SM, Fleck MS, Cabeza R. 2008. Qué PASA? The posterior-anterior shift in aging. Cereb Cortex. 18:1201-1209.

Dennis NA, Kim H, Cabeza R. 2008. Age-related differences in brain activity during true and false memory retrieval. J Cogn Neurosci. 20:1390-1402.

Dobbins IG, Foley H, Schacter DL, Wagner AD. 2002. Executive control during episodic retrieval: multiple prefrontal processes subserve source memory. Neuron. 35:989-996.

Dodson CS, Schacter DL. 2002. Aging and strategic retrieval processes: reducing false memories with a distinctiveness heuristic. Psychol Aging. 17:405-415.

Donaldson DI, Wheeler ME, Petersen SE. 2010. Remembering the source: dissociating frontal and parietal contributions to episodic memory. J Cogn Neurosci. 22:371-391.

Dornburg CC, McDaniel MA. 2006. The cognitive interview enhances long-term free recall of older adults. Psychol Aging. 21:196-200.

Dosenbach NUF, Fair DA, Cohen AL, Schlagger BL, Petersen SE. 2008. A dual-networks architecture of top-down control. Trends Cogn Sci. 12:99-105.

Duarte A, Graham KS, Henson RN. 2010. Age-related differences in neural activity associated with familiarity, recollection and false recognition. Neurobiol Aging. 31:1814-1830.

Duarte A, Henson RN, Graham KS. 2008. The effect of aging on the neural correlates of subjective and objective recollection. Cereb Cortex. 18:2169-2180.

Duverne S, Habibi A, Rugg MD. 2008. Regional specificity of age effects on the neural correlates of episodic retrieval. Neurobiol Aging. 29:1902-1916.

Ferguson SA, Hashtroudi S, Johnson MK. 1992. Age-differences in using source-relevant cues. Psychol Aging. 7:443-452.

Folstein MF, Folstein SE, McHugh PR. 1975. "Mini-mental state": a practical method for grading the mental state of patients for the clinician. J Psychiatr Res. 12:189-198.

Gallo DA. 2010. False memories and fantastic beliefs: 15 years of the DRM illusion. Mem Cognit. 38:833-848.

Gallo DA, Cotel SC, Moore CD, Schacter DL. 2007. Aging can spare recollection-based retrieval monitoring: the importance of event distinctiveness. Psychol Aging. 22:209-213.

Gallo DA, Kensinger EA, Schacter DL. 2006. Prefrontal activity and diagnostic monitoring of memory retrieval: fMRI of the criteria recollection task. J Cogn Neurosci. 18:135-148.

Gallo DA, McDonough IM, Scimeca J. 2010. Dissociating source memory decisions in the prefrontal cortex: fMRI of diagnostic and disqualifying monitoring. J Cogn Neurosci. 22:955-969.

Gallo DA, Weiss JA, Schacter DL. 2004. Reducing false recognition with criterial recollection tests: distinctiveness heuristic versus criterion shifts. J Mem Lang. 51:473-493.

Gazzaley A, Cooney JW, Rissman J, D'Esposito M. 2005. Top-down suppression deficit underlies working memory impairment in normal aging. Nat Neurosci. 8:1298-1300.

Glisky EL, Rubin SR, Davidson PSR. 2001. Source memory in older adults: an encoding or retrieval problem? J Exp Psychol Learn Mem Cogn. 27:1131-1146.

Glover GH. 1999. 3D z-shim method for reduction of susceptibility effects in BOLD fMRI. Magn Reson Med. 42:290-299.

Grady CL. 2002. Age-related differences in face processing: a meta-analysis of three functional neuroimaging experiments. Can J Exp Psychol. 56:208-220.

Greenwood PM. 2007. Functional plasticity in cognitive aging: review and hypothesis. Neuropsychology. 21:657-673.

Grinband J, Wager TD, Lindquist M, Ferrera VP, Hirsch J. 2008. Detection of time-varying signals in event-related fMRI designs. Neuroimage. 43:509-520.

Gutchess AH, Hebrank A, Sutton B, Leshikar E, Chee MWL, Tan JC, Goh JOS, Park DC. 2007. Contextual interference in recognition memory with age. Neuroimage. 35:1338-1347.

Head D, Buckner RL, Shimony JS, Williams LE, Akbudak E, Conturo TE, McAvoy M, Morris JC, Snyder AZ. 2004. Differential vulnerability of anterior white matter in nondemented aging with minimal acceleration in dementia of the Alzheimer type: evidence from diffusion tensor imaging. Cereb Cortex. 14:410-423.

Henson RNA, Shallice T, Dolan RJ. 1999. Right prefrontal cortex and episodic memory of retrieval: a functional MRI test of the monitoring hypothesis. Brain. 122:1367-1381.

Hwang DY, Gallo DA, Ally BA, Black PM, Schacter DL, Budson AE. 2007. Diagnostic retrieval monitoring in patients with frontal lobe lesions: further exploration of the distinctiveness heuristic. Neuropsychologia. 45:2543-2552.

Johnson MK, DeLeonardis DM, Hashtroudi S. 1995. Aging and single versus multiple cues in source monitoring. Psychol Aging. 10:507-517.

Johnson MK, Hashtroudi S, Lindsay DS. 1993. Source monitoring. Psychol Bull. 114:3-28.

Li SC, Lindenberger U. 1999. Cross-level unification: a computational exploration of the link between deterioration of neurotransmitter systems and dedifferentiation of cognitive abilities in old age. In: Nilsson LG, Markowitsch HJ, editors. Cognitive neuroscience of memory. Ashland (OH): Hogrefe & Huber Publishers. p. 103-146.

Li SC, Lindenberger U, Sikstrom S. 2001. Aging cognition: from neuromodulation to representation. Trends Cogn Sci. 5:479-486.

Lieberman MD, Cunningham WA. 2009. Type I and Type II error concerns in fMRI research: re-balancing the scale. Soc Cogn Affect Neurosci. 4:423-428.

Lindenberger U, Baltes PB. 1994. Sensory functioning and intelligence in old age: a strong connection. Psychol Aging. 9:339-355.

Logan JM, Sanders AL, Snyder AZ, Morris JC, Buckner RL. 2002. Under-recruitment and nonselective recruitment: dissociable neural mechanisms associated with aging. Neuron. 33:827-840.

Mattay VS, Fera F, Tessitore A, Hariri AR, Berman KF, Das S, Weinberger DR. 2006. Neurophysiological correlates of age-related differences in working memory capacity. Neurosci Lett. 392:32-37.

Mazaika P, Whitfield S, Cooper JC. 2005. Detection and repair of transient artifacts in fMRI data. Neuroimage. 26(Supp 1):S36.

McCabe DP, Roediger HL III, McDaniel MA, Balota DA. 2009. Aging reduces veridical remembering but increases false remembering: neuropsychological test correlates of remember-know judgments. Neuropsychologia. 47:2164-2173.

Meinzer M, Wilser L, Flaisch T, Eulitz C, Rockstroh B, Conway T, Rothi LJG, Crosson B. 2009. Neural signatures of semantic and phonemic fluency in young and old adults. J Cogn Neurosci. 21:2007-2018.

Middleton FA, Strick PL. 2000. Basal ganglia and cerebellar loops: motor and cognitive circuits. Brain Res Rev. 31:236-250.

Mitchell KJ, Johnson MK. 2009. Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? Psychol Bull. 135:638-677.

Morcom AM, Li J, Rugg MD. 2007. Age effects on the neural correlates of episodic retrieval: increased cortical recruitment with matched performance. Cereb Cortex. 17:2491-2506.

Nagel IE, Preuschhof C, Li S, Nyberg L, Bäckman L, Lindenberger U, Heekeren HR. 2009. Performance level modulates adult age differences in brain activation during spatial working memory. Proc Natl Acad Sci U S A. 106:22552-22557.

Naveh-Benjamin M. 2000. Adult age differences in memory performance: tests of an associative deficit hypothesis. J Exp Psychol Learn Mem Cogn. 26:1170-1187.

Park DC, Gutchess AH. 2005. Long-term memory and aging: a cognitive neuroscience perspective. In: Cabeza R, Nyberg L, editors. Cognitive neuroscience of aging: linking cognitive and cerebral aging. New York: Oxford University Press. p. 218-245.

Park DC, Polk TA, Park R, Minear M, Savage A, Smith MR. 2004. Aging reduces neural specialization in ventral visual cortex. Proc Natl Acad Sci U S A. 101:13091-13095.

Park DC, Reuter-Lorenz P. 2009. The adaptive brain: aging and neurocognitive scaffolding. Annu Rev Psychol. 60:173-196.

Parkin AJ, Bindschaedler C, Harsent L, Metzler C. 1996. Pathological false alarm rates following damage to the left frontal cortex. Brain Cogn. 32:14-27.

Persson J, Nyberg L, Lind J, Larsson A, Nilsson LG, Ingvar M, Buckner RL. 2006. Structure-function correlates of cognitive decline in aging. Cereb Cortex. 16:907-915.

Rajah MN, D'Esposito M. 2005. Region-specific changes in prefrontal function with age: a review of PET and fMRI studies on working and episodic memory. Brain. 128:1964-1983.

Rajah MN, Languay R, Valiquette L. 2010. Age-related differences in prefrontal cortex activity are associated with behavioral deficits in both temporal and spatial context memory retrieval in older adults. Cortex. 46:535-549.

Raz N, Lindenberger U, Rodrigue KM, Kennedy KM, Head D, Williamson A, Dahle C, Gerstorf D, Acker JD. 2005. Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. Cereb Cortex. 15:1676-1689.

Reuter-Lorenz PA, Cappell KA. 2008. Neurocognitive aging and the compensation hypothesis. Curr Dir Psychol Sci. 17:177-182.

Reuter-Lorenz PA, Jonides J, Smith EE, Hartley A, Miller A, Marshuetz C, Koeppe RA. 2000. Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. J Cogn Neurosci. 12:174-187.

Roediger HL III, McDaniel MA. 2007. Illusory recollection in older adults: testing mark twain's conjecture. In: Garry M, Hayne H, editors. Do justice and let the sky fall: Elizabeth Loftus and her contributions to science, law, and academic freedom. Mahwah (NJ): Lawrence Erlbaum Associates. p. 105-136.

Rombouts SA, Goekoop R, Stam CJ, Barkhof F, Scheltens P. 2005. Delayed rather than decreased BOLD response as a marker for early Alzheimer's disease. Neuroimage. 26:1078-1085.

Rugg MD. 2004. Retrieval processing in human memory: electrophysiological and fMRI evidence. In: Gazzaniga MS, editor. The cognitive neurosciences. 3rd ed. Cambridge (MT): MIT Press. p. 727-738.

Rypma B, D'Esposito M. 2000. Isolating the neural mechanisms of age-related differences in human working memory. Nat Neurosci. 3:509-515.

Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RS, Busa E, Morris JC, Dale AM, Fischl B. 2004. Thinning of the cerebral cortex in aging. Cereb Cortex. 14:721-730.

Schacter DL, Curran T, Galluccio L, Milberg WP, Bates JF. 1996. False recognition and the right frontal lobe: a case study. Neuropsychologia. 34:793-808.

Schacter DL, Israel L, Racine C. 1999. Suppressing false recognition in younger and older adults: the distinctiveness heuristic. J Mem Lang. 40:1-24.

Schacter DL, Koutstaal W, Norman KA. 1997. False memories and aging. Trend Cogn Sci. 1:229-236.

Schacter DL, Wiseman AL. 2006. Reducing memory errors: the distinctiveness heuristic. In: Hunt RR, Worthen JB, editors. Distinctiveness and memory. New York: Oxford University Press. p. 89-107.

Schneider-Garces NJ, Gordon BA, Brumback-Peltz CR, Shin E, Lee Y, Sutton BP, Fabiani M. 2010. Span, CRUNCH and beyond: working memory capacity and the aging brain. J Cogn Neurosci. 15:655-669.

Scimeca JM, McDonough IM, Gallo DA. 2011. Quality trumps quantity at reducing memory errors: implications for retrieval monitoring and mirror effects. J Mem Lang. 65:363-377.

Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci. 27:2349-2356.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15:273-289.

Velanova K, Lustig C, Jacoby LL, Buckner RL. 2007. Evidence for frontally mediated controlled processing differences in older adults. Cereb Cortex. 17:1033-1046.

Verfaellie M, Rapcsak SZ, Keane MM, Alexander MP. 2004. Elevated false recognition in patients with frontal lobe damage is neither a general nor a unitary phenomenon. Neuropsychology. 18:94-103.

Zacks RT, Hasher L, Li KZH. 2000. Human memory. In: Craik FIM, Salthouse TA, editors. The Handbook of aging and cognition. 2nd ed. Mahwah (NJ): Lawrence Erlbaum Associates. p. 293-357.