Psychology and Aging
2007, Vol. 22, No. 1, 209–213

0882-7974/07/$12.00 DOI: 10.1037/0882-7974.22.1.209

Aging Can Spare Recollection-Based Retrieval Monitoring: The Importance of Event Distinctiveness

David A. Gallo
University of Chicago

Sivan C. Cotel
Wesleyan University

Christopher D. Moore
Princeton University

Daniel L. Schacter
Harvard University

The authors investigated two retrieval-monitoring processes. Subjects studied red words and pictures and then decided whether test words had been studied in red font (red word test) or as pictures (picture test). Memory confusions were lower on the picture test than on the red word test, implicating a distinctiveness heuristic. Memory confusions also were lower when study formats were mutually exclusive (the same item was never studied as both a red word and a picture), compared with a nonexclusive condition, implicating a recall-to-reject process. When the to-be-recollected events were pictures, older adults used each monitoring strategy as effectively as did younger adults.

Keywords: aging, false recognition, retrieval monitoring, criterial recollection, distinctiveness

Older adults can be more susceptible to memory errors than are younger adults, especially when accuracy depends on specific recollections (see Pierce, Simons, & Schacter, 2004, for review). For example, Jacoby (1999) had subjects study two lists of words (one auditory and one visual). At test, they were instructed to respond positively to words heard in the inclusion list (targets), but to reject words seen in the exclusion list (familiar lures). Aging was associated with increased false recognition of lures from the exclusion list, indicating that older adults had difficulties recollecting modality-specific information to reject these lures. On these sorts of exclusion tasks, recollecting a word from one source allows one to reject it from having occurred in the other source (a *recall-to-reject* process). Such mutual-exclusivity rules have been used in a variety of memory tasks (e.g., associative recognition, word conjunctions, false fame, repetition lag, and the Deese-Roediger-McDermott or DRM task), and aging has been found to increase false recognition in each (e.g., Castel & Craik, 2003; Jones & Jacoby, 2005; Multhaup, 1995; Jennings & Jacoby, 1997; Gallo, Bell, Beier, & Schacter, 2006, respectively).

In contrast to these findings, other researchers have found that older adults are not impaired in their ability to use recollection to avoid false recognition. For instance, Schacter, Israel, and Racine (1999) found that both younger and older adults reduced familiarity-based errors after studying pictures, compared with words. They argued that subjects expected more distinctive recollections for pictures, allowing the rejection of nonstudied lures (which did not elicit picture recollections) via a process dubbed the *distinctiveness heuristic* (for review, see Schacter & Wiseman, 2006). These findings indicate that aging does not necessarily impair the use of recollection to monitor recognition accuracy.

One potentially important difference between these two paradigms is in the distinctiveness of the task-relevant recollections. The distinctiveness heuristic work has involved visually complex and unique pictures, but in all of the aforementioned exclusion tasks, the to-be-recollected information has been relatively nondistinctive (e.g., the occurrence of a word in a particular list). Given that age-related deficits in source memory are largest when similar sources need to be differentiated (e.g., Henkel, Johnson, & De Leonardis, 1998), the latter types of information may be more difficult for older adults to monitor. If so, the differential effects of aging on the recall-to-reject process and the distinctiveness heuristic may have been due to differences in the study materials.

Another potentially important difference between these two paradigms is the way that to-be-recollected information was used in the memory decisions. Recollection-based monitoring processes can involve two qualitatively different types of decisions (see Gallo, Bell, et al., 2006). Many recall-to-reject processes involve the *successful* recollection of information that disqualifies the event as having occurred. The various mutual-exclusion tasks discussed above are all examples of this sort of decision process. In contrast, the distinctiveness heuristic involves the *failure* to recollect expected information, which is diagnostic that an event did not occur. This decision process is analogous to the memorability heuristics that have been documented in the source memory literature (e.g., Hicks & Starns, 2006; Johnson, Hashtroudi, &

David A. Gallo, Department of Psychology, University of Chicago; Sivan C. Cotel, Department of Psychology, Wesleyan University; Christopher D. Moore, Department of Psychology, Princeton University; Daniel L. Schacter, Department of Psychology, Harvard University.

Preliminary analyses of these data were presented at the 46th Annual Meeting of the Psychonomic Society (November 2005, Toronto, Ontario, Canada). This work was supported by National Institute on Aging Grants AG21369 (to David A. Gallo) and AG08441 (to Daniel L. Schacter). We thank Jonathan Weiss for assistance in collecting pilot data for these experiments.

Correspondence concerning this article should be addressed to David A. Gallo, University of Chicago, Department of Psychology, 5848 South University Avenue, Chicago, IL 60637. E-mail: dgallo@uchicago.edu

Lindsay, 1993). By this view, the differential effects of aging in prior tasks were attributable to different decision process, as opposed (or in addition) to differences in the materials.

A major theoretical hurdle is that recall-to-reject processes and the distinctiveness heuristic have been investigated in separate studies, using different types of materials and tasks. In one recent exception, Gallo, Bell, et al. (2006) reported evidence that younger adults could use both of these strategies in the same task. However, there was little evidence that older adults had used either strategy, and this finding is at odds with previous experiments on the distinctiveness heuristic. As discussed by Gallo, Bell, et al., instructing older adults to focus on a difficult list-differentiation strategy (as in Jacoby, 1999) may have precluded their use of the distinctiveness heuristic in that study.

To overcome these limitations, in the present study we investigated the effects of aging on these two monitoring strategies using the criterial recollection task (Gallo, Kensinger, & Schacter, 2006; Gallo, Weiss, & Schacter, 2004). Subjects studied object names (in black font) paired with the same word in red font or with a colored picture of the object. They then took several recognition memory tests (using white object names as retrieval cues). On the red word test, subjects were instructed to respond positively if they recollected that the item had been presented as a red word, and on the picture test, they responded positively if they recollected a picture. It is important to note that some items were studied as both red words and pictures. Because these sources were nonexclusive, subjects could not use an exclusion-based recall-to-reject strategy, and instead they had to selectively search memory for criterial recollections (e.g., pictures on the picture test, red words on the red word test). Gallo et al. (2004) found that younger adults were less likely to make recognition errors when they monitored memory for picture recollections (the picture test) compared with word recollections (the red word test). This result was obtained even when the red words were made more familiar than pictures (via study repetitions), providing strong evidence for a recollection-based distinctiveness heuristic, as opposed to familiarity-based responding. By expecting more distinctive recollections on the picture test, subjects were able to avoid false recognition (e.g., "I don't remember studying a picture of this object").

One goal of the present study was to use this task to provide a more direct measure of the distinctiveness heuristic in older adults. By repeating red words at study, familiarity differences between red words and pictures can be minimized. As a result, differences in false recognition across the red word test and picture test can be attributed to recollection-based expectations (i.e., the distinctiveness heuristic), as opposed to familiarity. Ruling out familiarity-based explanations is important, because aging is thought to differentially affect recollection and familiarity (e.g., Jacoby, 1999). In addition, this task provides more external support for the use of a distinctiveness heuristic than has been provided in prior tasks (e.g., Gallo, Bell, et al., 2006), because subjects are explicitly instructed to base their recognition decisions on picture recollections. This is also an important consideration for aging studies, as older adults sometimes require special guidance to demonstrate that they can use certain types of recollection-based strategies (Multhaup, 1995).

The other goal was to investigate the degree to which younger and older adults could use a recall-to-reject strategy to reduce false recognition. To this aim, we created a separate "exclusive" condition, in which an item was studied either as a red word or as a picture and never as both. This manipulation added a mutual-exclusivity rule that subjects were instructed to use while making their memory decisions (e.g., "I remember studying a picture of this object, so it wasn't studied as a red word"). The advantage of this recall-to-reject rule is that it is based on the same type information (e.g., picture recollections) as the distinctiveness heuristic. By equating the type of to-be-recollected information for each monitoring process and by designing the task instructions so that both strategies would be clear to subjects, we allowed a more direct comparison of these two monitoring processes in the current study than did researchers in prior aging studies.

Method

Subjects

Forty-eight younger adults (mean age = 20.9 years, SD = 2.3; range = 18–27 years) and 48 older adults (mean age = 70.8 years, SD = 4.5; range = 63–78 years) participated for course credit or money and were assigned to one of the two experimental conditions (equated on age). Younger adults were recruited at Harvard University, and older adults were volunteers drawn from the same pool used in prior studies of the distinctiveness heuristic (but with no experience in the current task). Older adults were generally high functioning, and they were prescreened for neurological disease or insults or psychoactive medications (self-report) and for depression (Geriatric Depression Scale; Yesavage et al., 1983). Data from 2 younger adults were replaced because of failure to follow instructions.

Materials and Design

The task was presented on an Apple Macintosh G4 computer, with PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993), and the nonexclusive condition was based on the design of Gallo, Kensinger, and Schacter (2006). The stimuli were 360 pictures of common objects (e.g., pumpkin, tree) from the Internet (see Koutstaal et al., 2001, for more details). For each studied item, a label was presented for 250 ms in black font, immediately followed by the same word in red letters (1 s) or a colored picture of the object (1.5 s). Pictures were studied for a longer duration to ensure encoding of their features. During the test phase, white words were used as memory cues, along with the appropriate test prompt (i.e., "studied?" for the standard test, "red word?" for the red word test, and "picture?" for the picture test). In the nonexclusive condition, the stimuli were counterbalanced across the studied conditions (red word, picture, both, or neither) and test conditions (standard test, red word test, or picture test). The exclusive condition was similar, except that no items were studied as both red words and pictures. Instead, half of the items studied in both formats in the nonexclusive condition were studied only as red words in the exclusive condition, and the other half were studied only as pictures. These items were not analyzed, so all comparisons across conditions were based on the same items.

Procedure

All subjects first completed a 10-min practice version of the experiment to ensure that they understood the procedures and instructions. Subjects then proceeded directly to the study phase of the main experiment. During study, subjects were instructed to remember 270 words and pictures. In the nonexclusive condition, 90 stimuli were presented in each of the three study formats (red word only, picture only, and both red word and picture [*both* items]). To equate recognition memory for red words and pictures, we repeated each red word three times. Each picture was presented once. All items were randomly ordered during the study phase, with the exception that 30 items of each type occurred in each third of the study list (including the repetitions of a given item). The exclusive condition was similar, except that no items were presented as both red words and pictures. The test phase consisted of three runs, each divided into three test blocks (standard, red word, and picture test). Each block was separated by a prompt indicating the type of test. The order of the test blocks was varied across the three runs (within subject) and counterbalanced. During each test block, subjects were tested on 40 items (randomly arranged). Each test word was presented for 3 s, with brief fixation periods interspersed throughout. There were 30 critical items of each item type (red word, picture, new) in each of the three testing conditions, with 30 additional items designated as *both* items (exclusive condition) or as fillers (nonexclusive condition).

On the standard test, subjects were instructed to press "yes" for any test word that corresponded to a studied stimulus, regardless of the study format. On the red word test, they were to press "yes" if they remembered studying a corresponding red word; on the picture test, they were to press "yes" if they remembered studying a picture. In the nonexclusive condition, subjects were reminded that some items were studied in both formats, so the recollection of one format (e.g., a picture) did not preclude presentation in the other format (e.g., a red word). Thus, they had to selectively search memory for the to-be-remembered format. In the exclusive condition, subjects were reminded that items were only studied in one format or the other. Thus, if they could remember a picture, the item could not have been studied as a red word, and if they recalled a red word, it could not have been studied as a picture.

Results

We first analyzed results from the nonexclusive condition (see Table 1). On the standard test, a 4 (item) × 2 (age) analysis of variance (ANOVA) revealed no effect of age and no interaction (both $ps > .05$). By design, the hit rates for red words and pictures were equated (both $ts < 1$), and all studied items were successfully discriminated from nonstudied lures (all $ps < .001$). On the red word test, there was an interaction between age and item, $F(3, 138) = 6.23$, $MSE = .016$. In younger adults, true recognition of red words (.60) was greater than false recognition of pictures (.40), $t(23) = 5.25$, $SEM = .038$, but older adults were unable to make this source discrimination, $Ms = .55$ and .56, respectively, $t(23) < 1$. Picture false alarms were greater in older adults (.56) relative to younger adults (.40), $t(36) = 3.11$, $SEM = .051$, with no other age differences (all $ps > .05$). On the picture test, there was no effect of age or interaction (both $ps > .05$). Searching memory for more distinctive recollections reduced the age differences found on the red word test, and both age groups were able to discriminate pictures from red words (collapsed across age groups, $Ms = .51$ and .14 for older and younger groups, respectively), $t(47) = 15.39$, $SEM = .024$.

To test the distinctiveness heuristic hypothesis, we compared false recognition across the criterial recollection tests (see Figure 1). A 2 (lure) × 2 (test) × 2 (age) ANOVA on these false recognition data revealed a significant three-way interaction, $F(1, 46) = 5.09$, $MSE = .039$. To explore this interaction, we conducted a 2 (test) × 2 (age) ANOVA on studied lures, which revealed an effect of age, $F(1, 46) = 9.74$, $MSE = .035$, confirming that older adults made more familiarity-based false alarms. There also was an effect of test, $F(1, 46) = 228.76$, $MSE = .012$, demonstrating the distinctiveness effect on false recognition (red word test > picture test). These variables did not interact ($p > .05$), and effect sizes were similar across age groups ($d = 2.14$ for younger adults, and $d = 2.11$ for older adults). Although older adults were more likely to make false alarms to studied lures, the

Table 1
Mean Recognition of Each Age Group in the Nonexclusive and Exclusive Conditions

Test	Nonexclusive				Exclusive			
	Younger		Older		Younger		Older	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Standard test								
Both hits	.71	.04	.75	.04				
Red word hits	.57	.04	.54	.05	.63	.04	.64	.03
Picture hits	.54	.04	.56	.04	.63	.04	.65	.03
New FAs	.17	.02	.18	.03	.21	.03	.22	.03
Red word test								
Both hits	.70	.04	.71	.04				
Red word hits	.60	.04	.55	.04	.66	.04	.67	.03
Picture FAs	.40	.03	.56	.04	.28	.03	.42	.04
New FAs	.22	.04	.21	.04	.29	.05	.32	.04
Picture test								
Both hits	.50	.04	.56	.05				
Picture hits	.48	.04	.55	.04	.55	.04	.60	.03
Red word FAs	.10	.02	.18	.03	.12	.03	.12	.02
New FAs	.08	.02	.11	.03	.07	.02	.09	.02

Note. FA = false alarm.

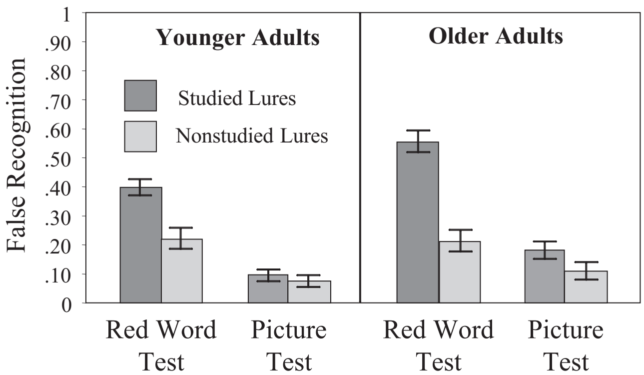

Figure 1. False recognition means (and standard errors) from the nonexclusive condition. In both age groups, false recognition of both studied and nonstudied lures was reduced on the picture test (implicating the distinctiveness heuristic).

two groups were equally likely to suppress false alarms on the picture test relative to the red word test. For nonstudied lures, there was a main effect of test, $F(1, 46) = 34.16$, $SEM = .01$, and no effect of age or interaction, both $F\text{s} < 1$, $d = .83$ and $d = .55$, respectively. These results further suggest that the groups were equally likely to use the distinctiveness heuristic.

We next analyzed results from the exclusive condition (see Table 1). On the standard test, there again was no effect of age and no interaction (both $p\text{s} > .05$). On the red word test, there was an interaction between age and item, $F(2, 92) = 3.76$, $MSE = .018$, and again this interaction was driven by age-related increase in picture false alarms (.28 vs. .42), $t(46) = 3.09$, $SEM = .047$, with no other age differences (both $t\text{s} < 1$). Younger adults were able to discriminate red words (.66) from pictures (.28), $t(23) = 9.08$, $SEM = .043$, but unlike in the nonexclusive condition, older adults also were able to make this discrimination (.67 vs. .42), $t(23) = 5.79$, $SEM = .043$. As expected, allowing a mutual-exclusion rule facilitated source discrimination, a point that we discuss more below. On the picture test, there was no effect of age or interaction (both $p\text{s} > .05$), and both age groups were able to discriminate pictures from red words (collapsed across age groups, $M\text{s} = .57$ and .12), $t(47) = 14.61$, $SEM = .031$.

To test the recall-to-reject hypothesis, we compared false recognition on the red word test in the nonexclusive and exclusive conditions (see Figure 2). This comparison is most analogous to that made for the distinctiveness heuristic, because the recall-to-reject strategy required subjects to monitor memory for picture recollections. A 2 (lure) × 2 (condition) × 2 (age) ANOVA revealed an interaction between lure and condition, $F(1, 92) = 49.64$, $MSE = .012$, and lure and age, $F(1, 92) = 21.25$, $MSE = .012$. To explore these interactions, we conducted a 2 (condition) × 2 (age) ANOVA on studied lures, which revealed an effect of age, $F(1, 92) = 19.18$, $MSE = .029$, confirming that older adults were more susceptible to familiarity-based false recognition than were younger adults. There also was an effect of condition, $F(1, 92) = 13.86$, $MSE = .029$, suggesting that both groups had used a recall-to-reject process to reduce false recognition of studied lures in the exclusive condition. There was no interaction, $F < 1$, and effect sizes were similar across age groups ($d = .78$ for younger adults, and $d = .77$ for older adults), suggesting that younger and older adults were equally likely to use a recall-to-reject process. On nonstudied lures, there was an effect of condition, $F(1, 92) = 4.71$, $MSE = .04$, but no effect of age and no interaction, both $F\text{s} < 1$. There were fewer false alarms to nonstudied lures in the nonexclusive condition, potentially because the addition of *both* items suppressed positive responses (i.e., a list-strength effect). This effect was opposite to that on studied lures, consistent with the idea that a recall-to-reject strategy should only reduce false alarms to studied lures (i.e., items eliciting picture recollections).

Discussion

We found evidence for two qualitatively different types of retrieval monitoring. False recognition of all lures (studied and nonstudied) was reduced on the picture test relative to the red word test, consistent with the use of a distinctiveness heuristic. In contrast, introducing a mutual-exclusivity rule selectively reduced false recognition of studied lures, consistent with the use of a recall-to-reject strategy. Both age groups showed similar patterns of false recognition suppression, and there were minimal age differences in the absolute size of these effects. These findings suggest that aging can spare these two types of retrieval monitoring, at least when searching memory for distinctive pictures. Unlike Gallo, Bell, et al. (2006), we placed equal emphasis on the two strategies in the current task, and providing unbiased support potentially allowed older adults to demonstrate each monitoring process.

Although older adults were able to suppress false recognition, aging did increase overall levels of familiarity-based false recognition (i.e., source confusions for studied lures), especially on the red word test. This aspect of our results is consistent with those studies cited in the introduction that have demonstrated age-related monitoring impairments for nondistinctive memories (e.g., Jacoby, 1999). Recent functional magnetic resonance imaging results in younger adults also are relevant to these findings. Gallo, Kensinger, and Schacter (2006) found that regions in the dorsolateral prefrontal cortex were more active on the red word test than on the picture test. Other neuroimaging studies have implicated these same regions in effortful postretrieval monitoring (see Rugg, 2004, for review) and have shown that aging can impair (or otherwise

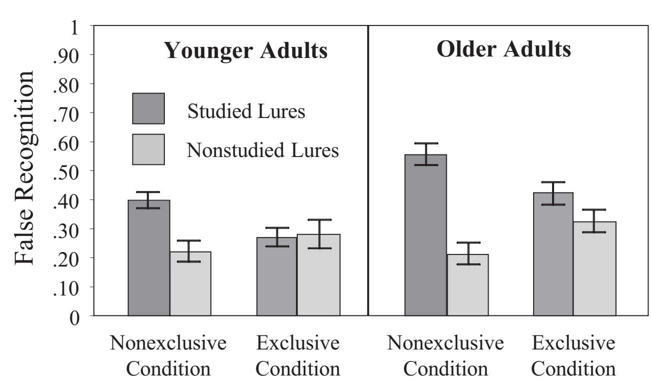

Figure 2. False recognition means (and standard errors) from the red word tests. In both age groups, only false recognition of studied lures was reduced in the exclusive condition (implicating recall-to-reject process).

alter) these frontally based retrieval processes (e.g., Grady, McIntosh, & Craik, 2005; Schacter, Savage, Alpert, Rauch, & Albert, 1996). These findings are only indirectly linked to the current results, but they support the idea that older adults were more impaired on the red word test due to frontally mediated retrieval deficits. More generally, age-related deficits in recollection-based monitoring may depend on the degree to which the monitoring process taps frontal functions, which in turn may depend on the distinctiveness of the memories.

In conclusion, our results indicate that the quality of the to-be-recollected events needs to be considered more extensively in studies of aging and retrieval monitoring. We equated overall levels of recognition memory for the two types of studied events so that red words and pictures mostly differed in terms of recollective distinctiveness (e.g., perceptual richness). Relative to younger adults, older adults were less accurate when the to-be-recollected information was nondistinctive and were more susceptible to familiarity-based false recognition. However, with more distinctive information, older adults were just as likely as younger adults to monitor memory for the presence or absence of detailed recollections. To the degree that the important events of people's lives are relatively distinctive, laboratory tasks that require the monitoring of nondistinctive events might overestimate the consequences of aging on recollection-based retrieval strategies.

References

Castel, A. D., & Craik, F. I. M. (2003). The effects of aging and divided attention on memory for item and associative information. *Psychology and Aging, 18,* 873–885.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25,* 257–271.

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in young and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory, 14,* 730–741.

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.

Gallo, D. A., Weiss, J. A., & Schacter, D. L. (2004). Reducing false recognition with criterial recollection tests: Distinctiveness heuristic versus criterion shifts. *Journal of Memory and Language, 51,* 473–493.

Grady, C. L., McIntosh, A. R., & Craik, F. I. M. (2005). Task-related activity in prefrontal cortex and its relation to recognition memory performance in young and old adults. *Neuropsychologia, 43,* 1466–1481.

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127,* 251–268.

Hicks, J. L., & Starns, J. J. (2006). The roles of associative strength and source memorability in the contextualization of false memory. *Journal of Memory and Language, 54,* 39–53.

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 3–22.

Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology and Aging, 12,* 352–361.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Jones, T. C., & Jacoby, L. L. (2005). Conjunction errors in recognition memory: Modality-free errors for older adults but not for younger adults. *Acta Psychologia, 120,* 55–73.

Koutstaal, W., Wagner, A. D., Rotte, M., Maril, A., Buckner, R. L., & Schacter, D. L. (2001). Perceptual specificity in visual object priming: Functional magnetic resonance imaging evidence for a laterality difference in fusiform cortex. *Neuropsychologia, 39,* 184–199.

Multhaup, K. S. (1995). Aging, source, and decision criteria: When false fame errors do and do not occur. *Psychology and Aging, 10,* 492–497.

Pierce, B. H., Simons, J. S., & Schacter, D. L. (2004). Aging and the seven sins of memory. *Advances in Cell Aging and Gerontology, 15,* 1–40.

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.

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., Savage, C. R., Alpert, N., Rauch, S., & Albert, M. S. (1996). The role of hippocampus and frontal cortex in age-related memory loss: A PET study. *NeuroReport, 7,* 1165–1169.

Schacter, D. L., & Wiseman, A. L. (2006). Reducing memory errors: The distinctiveness heuristic. In R. R. Hunt & J. Worthen (Eds.), *Distinctiveness and memory* (pp. 89–107). New York: Oxford University Press.

Yesavage, J. A., Brink T. L., Rose T. L., Lum, O., Huang, V., Adey, M. B., & Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research, 17,* 37–49.

Received March 23, 2006
Revision received July 18, 2006
Accepted September 25, 2006 ■