

Touron, D. R., & Hertzog, C. (2004). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging*, 19(3): 452-466. (Sep 2004) Published by the American Psychological Association (ISSN: 1939-1498). This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

## **Distinguishing Age Differences in Knowledge, Strategy Use, and Confidence During Strategic Skill Acquisition**

Dayna R. Touron and Christopher Hertzog

### **ABSTRACT**

The authors examined how age differences in strategy selection are related to associative learning deficits and metacognitive variables, including memory ability confidence. In Experiment 1, increases in memory reliance for performance of the noun-pair lookup task were compared with increases in noun-pair memory ability. In Experiment 2, memory reliance was assessed for noun pairs memorized prior to the task. In each experiment, older adults manifested a substantial delay in transition to a retrieval-based strategy despite comparable noun-pair knowledge. In Experiment 3, young and older adults reported comparable confidence ratings for the accuracy of each memory probe response. However, older adults reported lower confidence in their general ability to use the memory retrieval strategy, which correlated with avoidance of the retrieval strategy.

Learning to associate new elements is a component of many skill-acquisition tasks (Anderson, 1996; Ericsson & Charness, 1994). Older adults are able to acquire new associations but require a greater amount of study to do so (e.g., Naveh-Benjamin, 2000). It is therefore not surprising that skill-acquisition functions, mapping performance as a function of task practice, differ for young and older adults (Charness & Campbell, 1988; Fisk & Rogers, 1991; Hertzog, Cooper, & Fisk, 1996; Strayer & Kramer, 1994; Touron, Hoyer, & Cerella, 2001).

Skill acquisition can be thought of as a progression from effortful processing toward more fluent processing (Ackerman, 1987; Shiffrin & Schneider, 1977). This progression might involve quantitative improvements in efficiency as well as qualitative transitions in task approach. In simulated skill-acquisition tasks that measure response time (RT), such as memory search and mental arithmetic, skilled performance involves a transition from relatively slow controlled processing (e.g., computation by algorithm) to a (typically) faster memory retrieval process (Logan, 1988; Reder & Ritter, 1992; Rickard, 1997; Shiffrin & Schneider, 1977). This transition is reflected in substantial RT improvements, as well as in the elimination of effects that influence controlled processes but not fluent memory retrieval (e.g., memory search set size; Fisk & Rogers, 1991).

The current study evaluates skill acquisition in the noun-pair lookup task. Participants verify whether a centrally presented target noun pair matches one of a set of pairs contained in a lookup table at the top of the screen (Ackerman & Woltz, 1994). Consistent mapping (CM; Shiffrin & Schneider, 1977)—maintaining the same word pairings across practice—affords learning of the associative pairings. With repetition of CM noun pairs, individuals are able to respond via memory retrieval. Correct discrimination of matched from unmatched pairs can be achieved by either visual search of the lookup table (which we refer to as scanning) or by memory retrieval. The memory retrieval strategy yields much faster RTs.

The noun-pair task permits the separation of scanning from retrieval, observing the contribution of strategy shift to performance improvements. Scanning behavior can be observed by using a variably mapped (VM) version of the task, in which the pairings change randomly from trial to trial. In VM noun-pair trials, participants must scan the table to ensure accurate performance. Retrieval behavior can be observed by using recognition memory probes (Ackerman & Woltz, 1994). These probes present a target pair as in standard trials but without displaying the lookup table. Above-chance performance requires memory retrieval. Standard CM noun-pair task RTs typically fall between these two extremes, aligning with VM RTs early in practice and approaching retrieval probe RTs late in practice. A shift from scanning to retrieval has been inferred when CM noun-pair RTs become reliably faster than VM noun-pair RTs (Ackerman & Woltz, 1994; Rogers & Gilbert, 1997; Rogers, Hertzog, & Fisk, 2000).

The shift from scanning to retrieving during noun-pair task practice can be considered a strategic choice, one made rapidly on stimulus presentation that determines subsequent item processing. In the present context, the term strategy is used to describe a conscious or unconscious approach to selecting a mode of processing (see Schunn, Reder, Nhounyvanisvong, Richards, & Stroffolino, 1997). The description of this transition as a strategy

shift underscores the fact that following a rule-to-retrieval course of skill development is under an individual's (either direct or indirect) control.

Selection of the memory retrieval strategy can be measured by participants' self-reports of the strategy they used on each trial. Strategy reports in such skill-acquisition tasks appear to be valid reflections of strategic choice (e.g., retrieval reports are associated with reliably faster RT), and responding to strategy probes does not produce reactive effects on young or older adults' rates of skill acquisition (Hoyer, Cerella, & Onyper, 2004; Touron & Hertzog, 2004; Touron, Hoyer, & Cerella, in press).

Although it is clear that cognitive strategies do not completely account for age differences in cognitive task performance (Dunlosky & Hertzog, 1998; Salthouse, 1991), they can influence them (Kausler, 1994; Siegler & Lemaire, 1997). Rogers et al. (2000) showed that older adults were less likely to shift to a memory retrieval strategy, which exaggerated estimated age differences in rates of skill acquisition. Higher associative memory ability (assessed in independent psychometric tests) predicted faster noun-pair RT improvements and also discriminated older scanners from older retrievers.

One possible explanation for this outcome is that older adults' ability to rely on memory retrieval is delayed because of age-related impairments in associative learning. By this account, older adults use the retrieval strategy as soon as their level of learning affords it. We refer to associations that have been learned (i.e., are sufficiently well-learned to be accessible in memory) as items that are known. Some treatments of skill acquisition use a strength metaphor to capture the concept that associations are graded in their degree of learning and, hence, fluency of access (e.g., Schneider & Detweiler, 1987). We treat the concept of item knowledge as similarly graded rather than absolute. In the noun-pair task, item knowledge is presumed to be a direct but latent outcome of incidental learning of the new associations. However, individuals can achieve a correct discrimination for any given trial by scanning the lookup table, even when item knowledge is sufficient to achieve accurate and faster retrieval discrimination. Use of the memory retrieval strategy is under participant control and, as such, is subject to a variety of factors other than item knowledge.

Indeed, age differences in the shift to a retrieval strategy can be experimentally manipulated in ways that are not necessarily consistent with an associative learning deficit. Touron and Hertzog (2004) showed that expanding the size of the lookup table, thus making scanning more time consuming, caused older adults to shift to the retrieval strategy more rapidly. Faster rates of noun-pair learning resulted, despite the fact that the number of associates and exposures to those associates was held constant. Touron and Hertzog's findings suggest that strategy selection may involve a malleable choice that (perhaps implicitly) weighs the costs and benefits of different processing options. Although models of strategy choice typically do not invoke the concept of a criterion for strategy selection (see Rickard, 1997), one can consider a decision to scan when the individual could effectively rely on item knowledge to be a manifestation of a conservative strategy choice. Touron and Hertzog found that some older adults use a scanning strategy more often than would be expected from their posttask memory for the noun pairs (see also Rogers et al., 2000). Moreover, older adults' confidence in their ability to use the memory

strategy correlated with reported retrieval use. These findings are consistent with what we term the retrieval aversion hypothesis: Older adults avoid the retrieval strategy even when they have sufficient item knowledge to use it effectively. Older adults may be conservative in their criterion for choosing memory retrieval for a number of reasons, including greater perceived risk of memory failure. Moreover, this phenomenon may be a more general manifestation of conservatism in cognitive task performance. Older adults have been shown to have conservative response criteria in other RT tasks, preferring to trade speed for accuracy (e.g., Hertzog, Vernon, & Rypma, 1993; Salthouse & Somberg, 1982; see also Ratcliff, Thapar, & McKoon, 2001).

The relationship of rated confidence and retrieval strategy use is consistent with literature on aging and beliefs about memory. Older adults rate their memory ability as being lower than young adults and view their own memory as having declined (see Hertzog & Hulstsch, 2000). Despite the correspondence of these beliefs with what is known about aging and memory (Zacks, Hasher, & Li, 2000), older individuals' subjective memory is not necessarily based on a veridical assessment of their objective memory ability (e.g., Rabbitt & Abson, 1990). Indeed, age differences in memory beliefs may instead reflect an implicit theory that memory declines with age (Lachman, Bandura, Weaver, & Elliot, 1995; Lineweaver & Hertzog, 1998; McDonald-Miszczak, Hertzog, & Hulstsch, 1995).

The specific goals of the current project were to isolate age differences in noun-pair task performance, to compare the acquisition of noun-pair knowledge with reliance on a retrieval strategy, and to examine the relationship between memory confidence and retrieval strategy use. In Experiment 1, we concurrently measured improvements in noun-pair knowledge and strategy use in the noun-pair task. In Experiment 2, participants were first trained to criterion on either a subset of the noun pairs to be trained or on the entire set of noun pairs, and we examined age differences in strategic behavior given substantial item knowledge. In Experiment 3, we evaluated improvements in confidence judgments about noun-pair recognition memory in relation to changes in strategy reports. Although we expected both young and older adults to learn new associations and to use this information to guide performance, we expected that older adults would shift more slowly to the retrieval strategy, despite availability of item knowledge that affords its use. We also expected older adults' aversion to the memory retrieval strategy to be reflected by lower confidence in their ability to effectively use the memory retrieval strategy.

## **EXPERIMENT 1**

Experiment 1 examined how the level of item knowledge relates to strategy choice and performance in the noun-pair lookup task. To measure improvements in noun-pair knowledge, we embedded recognition memory probes within blocks of standard CM noun-pair task trials. Memory testing can alter the probability of using the memory retrieval strategy (Ackerman & Woltz, 1994). Hence, participants were randomly assigned to a standard noun-pair task (i.e., without memory probes) or to a task with memory probes. In theory, memory probes could increase the rate of learning new associations because of retrieval practice (Allen, Mahler, & Estes, 1969; Bahrick, 2000). Therefore, we randomly assigned half of the items to receive

memory probes; the remaining half of the item set was not probed. We hypothesized that recognition memory probes increase global confidence in the effectiveness of the memory retrieval strategy, leading to its increased use by older adults. Thus, we hypothesized equivalent effects for probed and unprobed items on the use of memory retrieval strategies, relative to the condition with no memory probes.

To examine the delay between item learning and strategy shift, we computed (for each trial) the conditional probability of retrieval given that the individual had correctly recognized that pairing at its most recent prior memory-probe trial. This conditional probability can be taken as an indirect measure of retrieval aversion. We hypothesized that older adults would manifest a lower probability of using memory retrieval even after demonstrating correct recognition memory performance for that item.

## **Method**

### **Design**

The between-subject independent variables were age and memory-probe condition. Equal numbers of young and older adults were randomly assigned to either a no-memory-probes condition (20 noun pairs without memory probes) or a memory-probes condition (10 noun pairs with memory probes and 10 items without memory probes). The within-subject independent variables were item type (probed or unprobed, for the memory-probe condition only) and practice (noun-pair repetition).

### **Participants**

The study included 40 young adults between the ages of 18–23 years and 40 older adults between the ages of 60–75 years. Young adults were undergraduate volunteers participating for extra credit. Older adults were recruited from the community, receiving an honorarium for their participation. Older adults had to be sufficiently healthy and intact to visit the laboratory on the Georgia Institute of Technology campus and were screened for some postsecondary education.

Mean scores and standard errors for these measures and the demographic characteristics for the participants in this experiment are reported in Table 1. Older adults scored higher on the Shipley Vocabulary Test (Zachary, 1986) than young adults. Young adults scored higher on the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981) Digit Symbol Substitution subtest (a speeded task of paired associates matching) and the Educational Testing Service First and Last Names Test (a measure of associative memory; French, Price, & Thurstone, 1962). The obtained age differences were comparable with those typically reported in the literature.

**Table 1**  
*Means (and Standard Errors) of Participant Characteristics for Experiment 1*

Measure	Young		Older	
	Control	Probes	Control	Probes
Age (years) <sup>a</sup>	20.7 (0.35)	20.8 (0.30)	68.7 (0.77)	67.4 (1.00)
Education (years)	15.2 (0.30)	15.1 (0.33)	15.7 (0.71)	16.5 (0.65)
Vocabulary <sup>a</sup>	32.2 (0.53)	30.6 (0.74)	35.8 (0.65)	35.1 (0.86)
Digit Symbol <sup>a</sup>	70.0 (2.82)	73.3 (3.16)	50.6 (1.70)	53.1 (2.62)
First–Last Names <sup>a</sup>	18.9 (1.57)	14.4 (1.34)	8.2 (1.10)	8.0 (0.72)
No. medications <sup>a,b</sup>	0.4 (0.15)	0.4 (0.21)	2.2 (0.32)	3.9 (0.50)

*Note.* Vocabulary refers to the number correct out of 40. Digit Symbol = Wechsler Adult Intelligence Scale—Revised Digit Symbol subtest; First–Last Names = Educational Testing Service First and Last Names subtest; No. medications = self-reported number of daily medications.

<sup>a</sup> Age comparison significant at  $p < .05$ . <sup>b</sup> Older adult condition comparison significant at  $p < .05$ .

*Means (and Standard Errors) of Participant Characteristics for Experiment 1*

**Procedure**

A Visual Basic 6.0 program controlled stimulus presentations and response recordings. Timing accuracy to the nearest millisecond was achieved by a routine that directly accessed the computer's internal clock. Stimuli were presented in 15-point Arial font at a resolution of 800 × 600 pixels. Participants were seated at a height and distance that optimized their own screen viewing and comfort.

A trial started with a centrally presented fixation point, followed by a centrally presented target item (e.g., cat —market). These two words were paired horizontally and separated by spaces and a dash. A table of 20 noun pairs was presented at the top of the screen, with the set of pairs arranged in five rows (each row containing four noun pairs) across the screen. The pairings did not change, with the location of each pair in the table randomly selected on each trial. Half of the trials presented a target pair matching a pair in the lookup table. The remaining nonmatched trials paired a left-hand word from one pair with a randomly selected right-hand word from a different pair.

After a short set of questionnaires and cognitive ability tests, participants completed self-paced computer instruction on the noun-pair task and then began training. Across training, all participants received 100 repetitions per item. Two sessions of 25 blocks of trials were completed on consecutive days. For participants in the no-memory-probe condition, a training block contained 40 randomly ordered trials, with two trial target repetitions each for the 20 word-pair stimuli. For participants in the memory-probe condition, a training block contained 50 randomly ordered trials, with two trial target repetitions each for the 20 noun-pair stimuli and one memory probe for each of the 10 probed noun-pair stimuli. For persons assigned to the memory-probe condition, the 10 noun pairs to be probed throughout training were randomly selected. The recognition memory probes had the same form as training trials, except that the lookup table was not provided.

We instructed participants to press a key labeled “Y” if the target pair had a match in the lookup table or a key labeled “N” if it did not. Strategy reports were obtained following each trial. Participants were instructed to indicate which of the following strategies they used on the previous trial by pressing a key labeled with the first letter of the response strategy: (1) scan (“S”), (2) memory (“M”), (3) both (“B”), or (4) other (“O”). Trial-level error feedback followed each strategy probe.

Participants were offered a rest break following each block, during which they received feedback on their mean RT and accuracy for noun-pair trials in the preceding block, as well as their mean accuracy for memory-probe trials. We instructed participants to keep their noun-pair accuracy at 95%. If the participant was more than 95% accurate, they were asked to speed their responding; if the participant was less than 95% accurate, they were asked to slow their responding. To prevent fatigue, every fifth break was constrained to at least 1 min in length. Young adults typically required about 1.5 hr to complete each session; older adults typically required about 2 hr to complete each session. Afterward, participants completed a posttest survey and debriefing.

## **Results and Discussion**

Both age groups performed close to the instructed 95% accuracy level in the noun-pair trials for all experiments, although older adults consistently showed slightly higher levels of accuracy (see Touron & Hertzog, 2004, for similar outcomes). Therefore, in the interest of brevity, detailed accuracy analyses are not presented.

### **Noun-Pair Trials**

#### **RT improvements**

RTs are presented in the top panel of Figure 1. 1 The temporary elevation in RT in the middle of the curve reflects the session break. Except where noted, practice effect analyses aggregated the repetitions variable over two repetitions into 50 blocks. Young adults responded consistently faster than older adults,  $F(1, 73) = 112.74$ ,  $MSE = 3,489,747,418$ ,  $p < .01$ ,  $r = .53$ . 2 The main effect of memory-probe condition was significant,  $F(1, 73) = 5.30$ ,  $MSE = 164,102,863$ ,  $p = .02$ , but was qualified by a significant Age  $\times$  Condition effect,  $F(1, 73) = 7.30$ ,  $MSE = 225,834,942$ ,  $p < .01$ . Older adults who received memory probes responded more quickly than older adults who did not receive memory probes ( $r = .36$ ), but the effect was not obtained for young adults ( $r = -.05$ ; perhaps because of a ceiling effect in the data from early in practice).

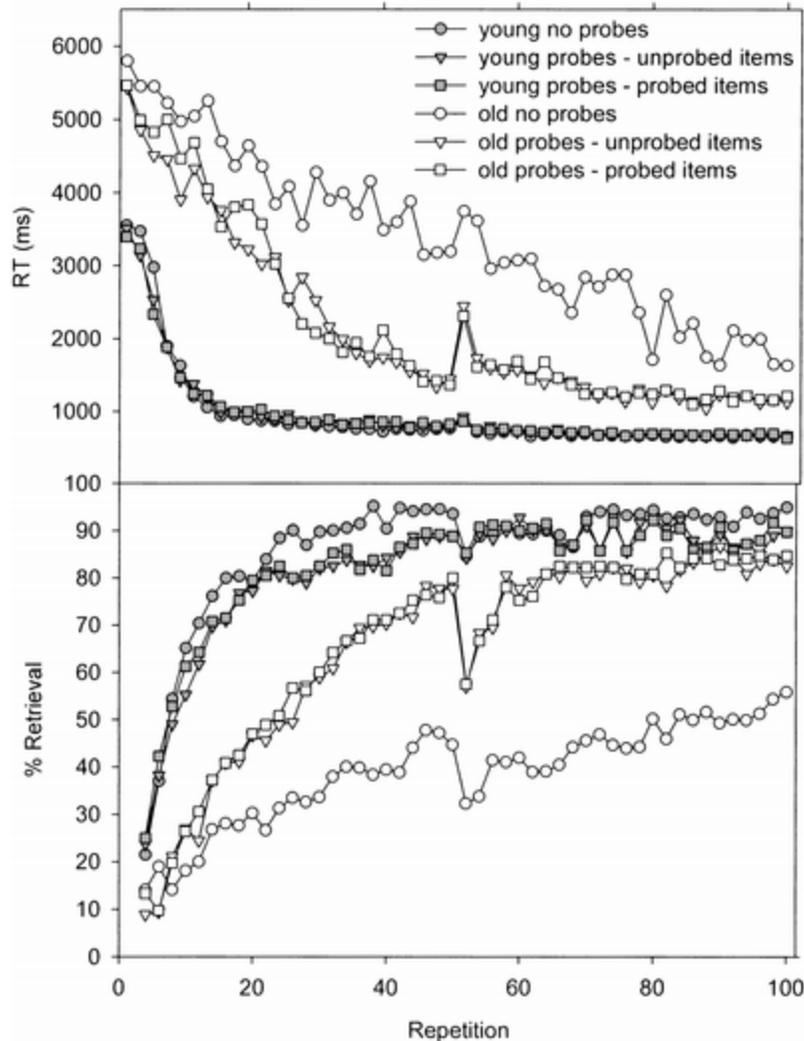


Figure 1. Experiment 1 mean median reaction times (RTs; in milliseconds; top panel) and mean percentage retrieval strategy use (bottom panel) by age, memory-probe condition, pair probing type, and repetition

RT improved with repetitions,  $F(49, 3577) = 131.31$ ,  $MSE = 55,308,267$ ,  $p < .01$ , and young adults improved more rapidly than older adults,  $F(49, 3577) = 19.48$ ,  $MSE = 8,204,836$ ,  $p < .01$ . Although the interaction of Memory-Probe Condition  $\times$  Repetitions was not significant ( $p = .27$ ), the interaction of Age  $\times$  Condition  $\times$  Repetitions was significant,  $F(49, 3577) = 1.78$ ,  $MSE = 749,193$ ,  $p < .01$ , and can be interpreted as follows. Older adults in the memory-probe condition improved more rapidly than those in the no-memory-probe condition,  $F(49, 3577) = 2.55$ ,  $MSE = 1,072,465$ ,  $p < .01$ , but the same interaction was not obtained for young adults ( $p > .99$ ). For participants in the memory-probe condition, RTs for probed and unprobed noun pairs did not differ and there were no reliable interactions of probing with age and repetitions ( $ps > .89$ ).

Whereas probing memory improved older adults' performance, memory probes did not influence young adults' performance. The benefit of memory probing on response latency was global, with equivalent improvements to probed and unprobed noun pairs for older participants in the memory-probe condition.

### Strategy separation

Reports of the "other" and "both" strategies were minimal, occurring for approximately 0.6% and 7.3%, respectively, of the responses. Because of the relative infrequency of "other" and "both" strategy reports, RTs were compared only for the "scan" and "memory" strategies (see Figure 2).

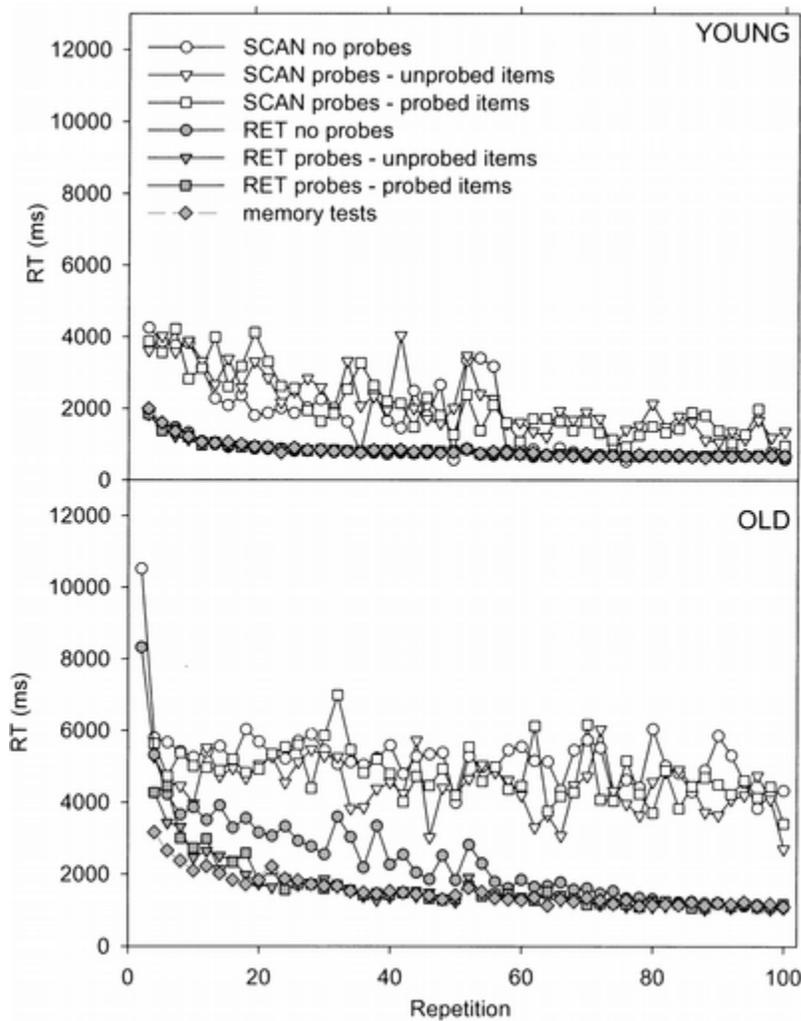


Figure 2. Young (top panel) and older (bottom panel) adults' mean median reaction times (RTs; in milliseconds) for Experiment 1 by strategy, memory-probe condition, pair probing type, and

*repetition. Memory probe RTs are also shown, but data were available only for the probed pairs in the memory-probe condition. SCAN = scanning; RET = retrieval*

In the strategy-separated data, no effect of (or interaction with) memory-probe condition was detected ( $p > .10$ ). RTs were considerably faster for trials with reported memory retrieval than with reported scanning,  $F(1, 76) = 32.71$ ,  $MSE = 70,655,620$ ,  $p < .01$ ,  $r = .66$ . Strategy report did not interact with age group or memory-probe condition ( $p > .06$ ), indicating that group differences in overall RT primarily reflect strategy differences. The pattern of RTs by strategy type supported the validity of the strategy reports.

A notable difference in retrieval RTs early in practice can be seen in the bottom panel of Figure 2, with slower RTs for older adults in the no-memory-probe condition. To assess the reliability of this difference, we compared older adults' retrieval RTs across memory-probe conditions, restricting the analysis to the first 50 repetitions. Reported memory retrieval RTs were indeed faster in the memory-probe condition,  $F(1, 34) = 9.85$ ,  $MSE = 17,023,638$ ,  $p < .01$ . We interpret this difference as resulting from lower memory ability confidence for older participants in the no-memory-probe condition. Higher confidence in response accuracy is typically associated with faster RT (Van Zandt, 2000; Vickers, 1979).

### **Retrieval strategy use**

Young adults consistently reported using retrieval more often than older adults,  $F(1, 73) = 30.42$ ,  $MSE = 816,852$ ,  $p < .01$ ,  $r = .48$  (see Figure 1, bottom panel). The main effect of memory-probe condition was not significant ( $p = .13$ ) but was qualified by a significant Age  $\times$  Condition interaction,  $F(1, 73) = 5.08$ ,  $MSE = 136,480$ ,  $p = .03$ . Older adults who received memory probes retrieved more often than older adults who did not receive memory probes ( $r = .31$ ). This effect was not obtained for young adults ( $r = -.07$ ).

Retrieval use increased with repetitions,  $F(49, 3431) = 58.04$ ,  $MSE = 17,141$ ,  $p < .01$ , and this increase was more rapid for young adults than for older adults,  $F(49, 3431) = 4.45$ ,  $MSE = 1,315$ ,  $p < .01$ . The interaction of Age  $\times$  Condition  $\times$  Repetitions was significant,  $F(49, 3431) = 1.38$ ,  $MSE = 406$ ,  $p < .05$ , and can be interpreted as follows: Older adults' frequency of retrieval in the memory-probe condition increased more rapidly than those of older adults in the no-memory-probe condition,  $F(49, 3431) = 2.17$ ,  $MSE = 640$ ,  $p < .01$ , but this effect was not obtained for young adults ( $p = .99$ ).

For participants in the memory-probe condition, retrieval strategy use for probed and unprobed noun pairs did not differ and there were no interactions of probing with age and repetitions ( $p > .85$ ). The finding of no difference between probed and unprobed items refutes the hypothesis that interspersed memory probes affect the level of specific item learning. Instead, the data indicate a global benefit from completion of memory probe.

Older adults in the memory-probe condition were more likely to report retrieval than older adults in the condition without memory probes, consistent with RT results reported by Rogers and Gilbert (1997). We did not find a reliable memory-probe benefit for young adults, which is inconsistent with results from Ackerman and Woltz (1994). The outcome difference might be accounted for by discrepant analytic methods (categorization of RT distributions vs. analysis of strategy reports) or differences in young adult sample characteristics.

This pattern of results is consistent with the hypothesis that adding recognition memory probes, and the experience with successful retrieval it generates, globally increases older adults' confidence in their ability to use the memory retrieval strategy.

## **Recognition Memory Performance**

### **Accuracy data**

Probe accuracy did not differ by age group ( $p = .11$ ; see Figure 3, top panel). Improvements in probe accuracy across repetitions were significant,  $F(48, 1824) = 15.87$ ,  $MSE = 1,732$ ,  $p < .01$ , and the speed of improvement differed between the two age groups,  $F(48, 1824) = 2.88$ ,  $MSE = 314$ ,  $p < .01$ , with young adults showing faster improvement than older adults. Figure 3 indicates that this difference in improvement occurs early in training. Indeed, when confining analysis to Session 2, no group or repetition differences in accuracy were found ( $ps > .05$ ). In summary, older adults' noun-pair knowledge increased more slowly than young adults' knowledge in early practice but reached young adults' levels of performance after a moderate amount of training.

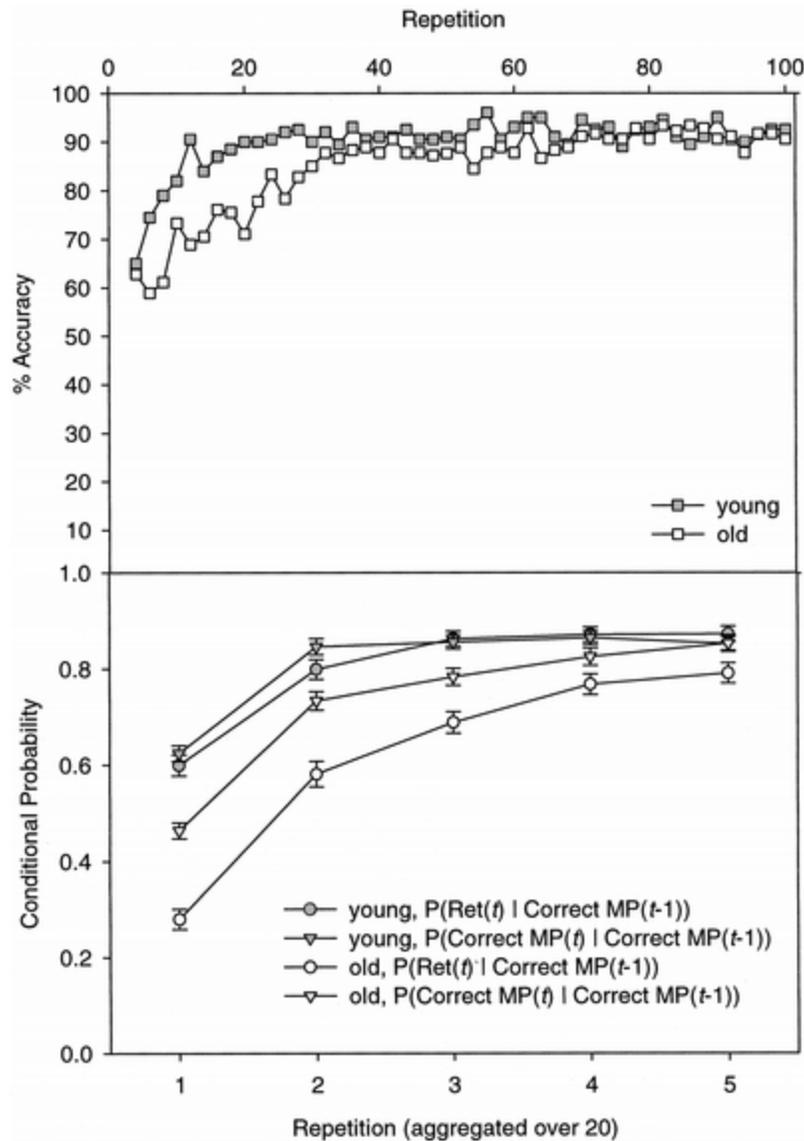


Figure 3. Top: Experiment 1 mean memory probe accuracy, by age and repetition. Note that accuracy data were available only for the 10 probed pairs in the memory-probe condition. Bottom: Experiment 1 mean probability of retrieval [ $P(\text{Ret}(t))$ ] and mean probability of a correct memory-probe response [ $P(\text{Correct MP}(t))$ ], both conditional on success for an item's last memory probe [ $P(\text{Correct MP}(t - 1))$ ], by age and block (to increase precision, data were aggregated over 20 repetitions). Standard error bars are also shown

### RT improvements

Young adults' RTs were consistently faster than older adults' RTs,  $F(1, 38) = 59.10$ ,  $MSE = 288,746,557$ ,  $p < .01$ ,  $r = .54$  (see Figure 2, diamond symbols). RTs improved with repetitions,  $F(48, 1824) = 51.36$ ,  $MSE = 5,688,771$ ,  $p < .01$ , and the young adults improved more quickly than did the older adults,  $F(48, 1824) = 5.78$ ,  $MSE = 640,441$ ,  $p < .01$ . Hence, there were age

differences in the speed of memory retrieval. Except early in practice, probe trial RT overlaid RT for reported memory retrieval trials. The equivalence of RT for required memory probes and reported retrieval-based RT again supports the validity of the strategy reports.

### **Conditional Probability of Retrieval**

The critical question was whether older adults' lower use of the memory retrieval strategy late in practice was governed solely by slower learning or also by a reluctance to use the retrieval strategy. We examined retrieval aversion by comparing age differences in the probability of retrieval given that the individual had correctly answered the previous memory probe for a specific noun pair (see Figure 3, bottom panel). Retrieval use following correct recognition was reliably lower for older adults than for young adults,  $F(1, 38) = 6.01$ ,  $MSE = 1.607$ ,  $p = .02$ ,  $r = .26$ . Increases with practice in the conditional probability of retrieval were significant,  $F(4, 152) = 49.65$ ,  $MSE = 1.049$ ,  $p < .01$ , and occurred more rapidly for young adults,  $F(4, 152) = 4.40$ ,  $MSE = 0.093$ ,  $p < .01$ .

The difference in conditional probabilities of retrieval reports might be attributed to less consistency in recognition memory across trials for older adults (see Salthouse & Dunlosky, 1995; Widner, Otani, & Smith, 2000, for analogous effects in recall tasks). We examined this possibility by computing age differences in the probability of correct recognition for a given noun pair, conditional on correct recognition on the previous memory probe for that same noun pair (see Figure 3, bottom panel). Although there was no main effect for age, significant improvements in the consistency of memory-probe accuracy occurred over blocks,  $F(4, 152) = 79.43$ ,  $MSE = 0.662$ ,  $p < .01$ , more rapidly for young adults than for older adults,  $F(4, 152) = 4.70$ ,  $MSE = 0.039$ ,  $p < .01$ . Nevertheless, age differences in the conditional probability of retrieval persisted when controlling on the covariate of consistency of recognition memory,  $F(1, 39) = 26.92$ ,  $p < .01$ .

In summary, older adults were more reluctant to rely on memory retrieval than young adults, and this reluctance cannot be accounted for by inconsistent noun-pair knowledge. We take these results as evidence that, in addition to an associative learning deficit, older adults are more likely to choose scanning over retrieval, as stipulated by the retrieval aversion hypothesis.

## **EXPERIMENT 2**

Experiment 2 was designed to more definitively test the hypothesis that age differences in noun-pair performance are merely a function of an associative learning deficit. Participants first memorized a set of noun pairs and were then given the noun-pair lookup task that included these memorized items. At issue was the degree of transfer of acquired item knowledge to noun-pair lookup task performance. In a first condition, young and older adults memorized to criterion a subset (half) of the noun pairs to be presented during noun-pair practice (we refer to this treatment as the 50% prelearning condition). Performance on prelearned noun pairs was then compared with performance on noun pairs that had not been prelearned. If age differences

in item learning are completely responsible for age differences in retrieval use, then both young and older adults should show elevated retrieval strategy use early in noun-pair practice for the prelearned items but not for the new items. In contrast, the retrieval aversion hypothesis states that older adults will still show a delayed strategy shift, even for prelearned items.

A complication for interpretation is that this design deliberately establishes major differences between subsets in degree of item knowledge. If older adults fail to demonstrate rapid shifts to the retrieval strategy for the subset of prelearned items, then it could be explained by them being less flexible in adapting strategy selection to level of item knowledge, perhaps because of task-switching deficits (e.g., Mayr & Kliegl, 2003). For example, older adults might not adopt a retrieval strategy until all items have been learned. Therefore, a second condition was tested in which all items were prelearned (100% prelearning condition). Across both conditions, we expected young adults to show relatively little delay in using item knowledge to shift to retrieval strategy use. In contrast, we expected older adults to show a substantial delay before shifting to the retrieval strategy, as predicted by the retrieval aversion hypothesis.

## **Method**

### **Design**

For the 50% prelearning condition, the within-subject independent variables were prelearning (10 noun pairs were prelearned and 10 items were not prelearned) and practice. For the 100% prelearning condition, the within-subject independent variable was practice. The 100% condition was added after data for the 50% condition were collected; hence, condition comparisons are not based on a fully randomized experiment.

### **Participants**

Each prelearning condition included 20 young adults between the ages of 18–23 years and 20 older adults between the ages of 60–75 years. Recruitment and compensation were the same as in Experiment 1. As shown in Table 2, there were no condition differences in any background variables, despite the lack of random assignment to prelearning conditions. Hence, the groups appeared to be comparable in relevant cognitive abilities that correlate with noun-pair performance (Rogers et al., 2000).

**Table 2**  
*Means (and Standard Errors) of Participant Characteristics and Metacognitive Variables for Experiment 2*

Measure	50% prelearning		100% prelearning	
	Young	Older	Young	Older
Age (years) <sup>a</sup>	19.6 (0.42)	68.9 (0.93)	19.6 (0.31)	66.9 (0.97)
Education (years)	13.6 (0.31)	14.8 (0.56)	13.7 <sup>a</sup> (0.33)	15.6 <sup>a</sup> (0.45)
Vocabulary <sup>a</sup>	31.2 (0.69)	35.2 (0.89)	32.9 (0.68)	36.4 (0.58)
Digit Symbol <sup>a</sup>	74.2 (2.25)	48.8 (2.70)	70.8 (2.37)	49.4 (2.54)
First–Last Names <sup>a</sup>	18.3 (1.64)	9.3 (1.32)	19.2 (1.39)	8.7 (1.12)
No. medications <sup>a</sup>	0.8 (0.26)	3.4 (0.50)	0.8 (0.19)	3.9 (0.44)
Criterion 1 <sup>a</sup>	2.2 (0.15)	8.3 (2.04)	3.0 (0.32)	13.5 (3.37)
Criterion 2	2.0 (0.00)	2.4 (1.17)	2.0 (0.00)	3.7 (0.92)
Effort <sup>a</sup>	2.7 (0.21)	3.6 (0.20)	2.9 (0.22)	3.7 (0.20)
Improvement	1.5 (0.21)	1.9 (0.27)	1.1 (0.14)	1.6 (0.17)
Global confidence			1.5 <sup>a</sup> (0.18)	2.2 <sup>a</sup> (0.25)

*Note.* Vocabulary refers to the number correct out of 40. No main or interaction effects of prelearning condition were significant ( $ps > .05$ ). Digit Symbol = Wechsler Adult Intelligence Scale—Revised Digit Symbol subtest; First–Last Names = Educational Testing Service Last Names subtest; No. medications = self-reported number of daily medications; Criterion 1 = number of blocks to criterion in Session 1; Criterion 2 = number of blocks to criterion in Session 2; Effort = self-rated effort required for memorization; Improvement = self-rated perceived improvement offered by the memory strategy for task performance; Global confidence = self-rated confidence in using the memory strategy.

<sup>a</sup> Age comparison significant at  $p < .05$ .

*Means (and Standard Errors) of Participant Characteristics and Metacognitive Variables for Experiment 2*

## Procedure

The apparatus and general procedure for the noun-pair task were identical to Experiment 1. In the 50% prelearning condition, the 10 noun pairs to be learned by each individual prior to the noun-pair task were randomly chosen by the computer program. A list of these items (and containing all 20 items for the 100% prelearning condition) was presented to the participant, and they were asked to study the list. After the subject-paced study time, participants completed blocks of prelearning trials until reaching a criterion of 9 out of 10 pairs correct (50% condition) or 18 out of 20 pairs correct (100% condition). Prelearning test trials had the same form as training trials, except that the noun-pair lookup table was not provided.

Across noun-pair lookup training, all participants responded to 100 repetitions per stimulus. The two sessions, run on consecutive days, consisted of 25 presentation blocks. A training block contained 40 trials, with 2 target repetitions each for the 20 word-pair stimuli. The second session began with a second exposure to the prelearning task prior to shifting participants back into the noun-pair task to ensure that participants had not forgotten the item pairings.

Following the computer testing, participants completed a posttest survey and debriefing. They answered questions eliciting verbal descriptions of their strategy use and completed self-ratings of effort (“How much effort did it require for you to memorize the word pairs?” on a scale from 1 = automatic to 5 = effortful) and perceived improvement (“How much does using memory

improve performance on this task?" on a scale from 1 = very much to 5 = not at all). Participants in the 100% condition also completed self-ratings of global memory confidence ("Were you confident to use your memory?" on a scale from 1 = yes to 5 = no).

## **Results and Discussion**

### **Prelearning Data**

Young adults reached criterion in fewer blocks than did older adults,  $F(1, 75) = 14.95$ ,  $MSE = 756$ ,  $p < .01$  (see Table 2). Criterion was reached in fewer blocks for Session 2 compared with Session 1,  $F(1, 75) = 22.50$ ,  $MSE = 622$ ,  $p < .01$ . A larger session difference occurred for older adults than for young adults, resulting in a significant Age  $\times$  Session interaction,  $F(1, 75) = 16.45$ ,  $MSE = 455$ ,  $p < .01$ . The age difference in blocks to criterion was significant in Session 1 but not in Session 2 (see Table 2). Condition differences in blocks to criterion were not reliable, and all interactions with condition were nonsignificant ( $ps > .10$ ).

### **RT Improvements**

Prelearning clearly affected RT early in practice (see Figure 4, top panel). This effect disappeared for young adults by the 20th repetition but persisted longer in practice for older adults. Prelearning did not eliminate age differences in rates of noun-pair learning, even in the 100% prelearning condition.

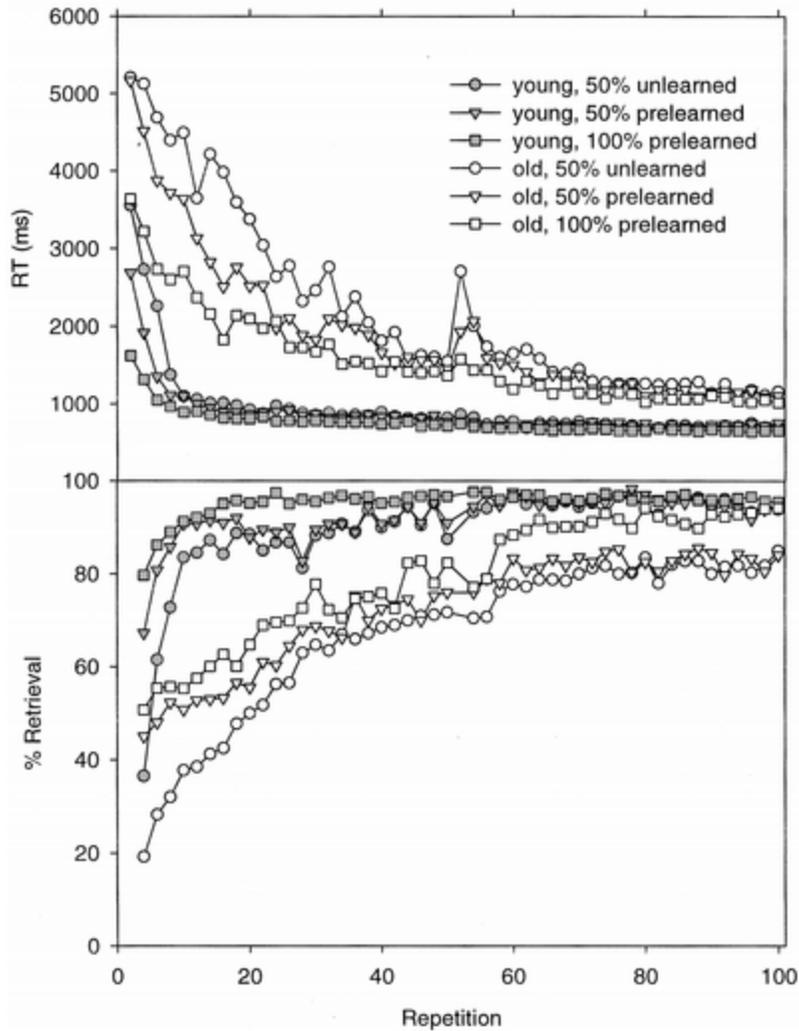


Figure 4. Experiment 2 mean median reaction times (RTs; in milliseconds; top panel) and mean percentage retrieval strategy use (bottom panel) by age, prelearning condition, prelearning stimulus type, and repetition

Young adults' RTs were consistently faster than older adults' RTs,  $F(1, 75) = 81.24$ ,  $MSE = 1,315,428,427$ ,  $p < .01$ ,  $r = .48$ . Participants in the 100% prelearning condition responded more quickly than those in the 50% prelearning condition,  $F(1, 75) = 8.13$ ,  $MSE = 131,612,390$ ,  $p < .01$ ,  $r = .16$ , but the condition effect did not differ by age group ( $p = .14$ ). RTs improved with repetitions,  $F(49, 3675) = 96.76$ ,  $MSE = 25,972,900$ ,  $p < .01$ . Improvement was more rapid for young adults than for older adults,  $F(49, 3675) = 25.06$ ,  $MSE = 6,726,625$ ,  $p < .01$ , and was more rapid for the 100% condition than for the 50% condition,  $F(49, 3675) = 5.84$ ,  $MSE = 1,568,710$ ,  $p < .01$ . Young and older adults had comparable Condition  $\times$  Repetitions effects, as shown by the nonsignificant three-way interaction ( $p = .72$ ).

In the 50% prelearning condition, prelearned noun pairs produced faster RTs than unlearned pairs,  $F(1, 37) = 11.39$ ,  $MSE = 29,642,034$ ,  $p < .01$ ,  $r = .09$ , and the interaction of prelearning type with age was nonsignificant ( $p = .12$ ). Improvements in RT were more rapid for prelearned pairs than for pairs that were not prelearned,  $F(49, 1813) = 5.64$ ,  $MSE = 776,148$ ,  $p < .01$ , but the interaction of Age  $\times$  Prelearning Type  $\times$  Repetitions was not significant ( $p = .43$ ).

### Retrieval Strategy Use

The critical data for the two hypotheses concerns the probability of retrieval strategy use. Young adults reported using retrieval more often than did older adults,  $F(1, 75) = 23.53$ ,  $MSE = 783,554$ ,  $p < .01$ ,  $r = .33$  (see Figure 4, bottom panel). Participants in the 100% prelearning condition used retrieval more often than those in the 50% condition,  $F(1, 75) = 146.66$ ,  $MSE = 4,883,534$ ,  $p < .01$ ,  $r = .13$ , and the condition effect was somewhat larger for older adults ( $r = .15$ ) than for young adults ( $r = .13$ ),  $F(1, 75) = 4.92$ ,  $MSE = 163,967$ ,  $p < .03$ . Retrieval use increased with repetitions,  $F(48, 3600) = 74.84$ ,  $MSE = 39,945$ ,  $p < .01$ . Improvement was more rapid for young adults than for older adults,  $F(48, 3600) = 9.43$ ,  $MSE = 5,031$ ,  $p < .01$ ,  $r = .33$ , and was more rapid for the 100% prelearning condition than for the 50% prelearning condition,  $F(49, 3675) = 10.51$ ,  $MSE = 5,609$ ,  $p < .01$ ,  $r = .13$ . Young and older adults had comparable Condition  $\times$  Repetitions effects, as shown by the nonsignificant three-way interaction ( $p = .15$ ).

In the 50% condition, retrieval was used more often for prelearned noun pairs than for unlearned pairs,  $F(1, 37) = 21.36$ ,  $MSE = 13,558$ ,  $p < .01$ ,  $r = .06$ , and the interaction of prelearning type with age was nonsignificant ( $p = .14$ ). Improvement in RTs were more rapid for prelearned pairs than for pairs that were not prelearned,  $F(48, 1775) = 11.76$ ,  $MSE = 587$ ,  $p < .01$ , and the interaction of Age  $\times$  Prelearning Type  $\times$  Repetitions was not significant ( $p = .84$ ).

In summary, prelearning the associations used in the noun-pair task clearly did not eliminate age differences in retrieval use. Despite having substantial knowledge of a partial subset or the complete set of noun pairings, older adults continued to manifest a slower strategy shift. Older adults' reported frequency of retrieval strategy use in the 100% condition did not approach the retrieval frequencies of young adults for items that were not prelearned. The slowness of older adults' strategy shift weighs against a simple learning deficit explanation and is more consistent with the retrieval aversion hypothesis.

The manipulations of prelearning did materially affect rates of noun-pair learning and strategy shift, with the curves showing separation early in Session 1 (as in the RT curves). The lower retrieval strategy reporting for prelearned items in the 50% prelearning condition, relative to 100% prelearning, suggests the effect of a strategic set. When prelearned items are mixed with items that are not prelearned, participants are less likely to report using the retrieval strategy. However, the strategy set effect cannot explain older adults' slow strategy shift in the 100% prelearning condition.

## **Posttask Ratings**

Despite prelearning of the noun pairs, older adults in both prelearning conditions still rated the memory retrieval strategy as more effortful (see Table 2). Indeed, there was no reliable difference between the 100% prelearning and 50% prelearning conditions in this regard. Prelearning all the noun pairs did not increase older adults' confidence in the use of memory retrieval strategy nor did it alter rated effort in using memory retrieval. There was no effect of condition on young adults' ratings either, but in their case, rated effort was low in both cases.

The ratings data favor the retrieval aversion hypothesis. Older adults apparently avoid the retrieval strategy because of low confidence in their ability to successfully use it—despite the extensive prelearning and noun-pair training. This argument is further supported by the fact that individual differences in retrieval use in the older group were substantially correlated with self-rated confidence in using the retrieval strategy (100% prelearning condition only, Pearson  $r = .58$ ,  $p < .01$ ). As suggested by previous studies, many but not all older adults are averse to using the retrieval strategy (Rogers et al., 2000; Touron & Hertzog, 2004).

## **EXPERIMENT 3**

Experiment 3 more closely examined the role of metacognitive variables in determining age differences in strategy shift. Metacognitive theory embraces two classes of variables that are relevant to the current study: beliefs about memory and online monitoring (Hertzog & Hulstsch, 2000).

### **Memory Beliefs**

The first two experiments suggest the possible importance of beliefs—specifically, low levels of confidence in the ability to use the retrieval strategy. An important question was whether older adults' lower confidence reflects a more general tendency for older adults to have a low memory self-concept. A belief in one's own memory ability can be either relatively general across task contexts or specific to particular contexts. Bandura (1997) argued that the context-specific self-efficacy—a belief in one's own ability to successfully execute a task—is a critical determinant of persistent goal-directed behavior, including strategy use (e.g., Berry, 1999). We were interested in determining whether the posttask (task-specific) measure of confidence in the retrieval strategy (which predicted retrieval use in the Experiment 2 100% prelearning condition and also predicted retrieval strategy use in Touron & Hertzog, 2004) or more global measures of memory self-concept and personal control over memory, as measured by a number of metamemory questionnaires (e.g., Lachman et al., 1995; Lineweaver & Hertzog, 1998; see Hertzog & Hulstsch, 2000), better predicted retrieval strategy use in the noun-pair task. We hypothesized that a lack of confidence in using the retrieval strategy would be associated with negative beliefs about one's own memory abilities and that both types of variable would predict retrieval strategy avoidance.

## Retrieval Monitoring Accuracy

Experiment 3 also evaluated the hypothesis that age deficits in monitoring memory retrieval outcomes account for age differences in strategy shift. We included two measures of retrieval monitoring accuracy in this experiment: the relative accuracy of confidence judgments to recognition memory probes—collected throughout noun-pair training—and the relative accuracy of delayed judgments of learning (JOLs) for the noun pairs (see Connor, Dunlosky, & Hertzog, 1997). If older adults are less able to monitor the accuracy of retrieval processes and associated recognition memory responses, then they may not be able to optimally regulate their use of the retrieval strategy in the standard noun-pair task. The available evidence suggests that older adults maintain a broad array of monitoring skills, even in the face of declining episodic memory (e.g., Connor et al., 1997; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002). Monitoring the accuracy of retrieved information from episodic memory may be an important exception to this generalization, however (Johnson & Raye, 2000; Souchay, Isingrini, & Espanet, 2000). Kelley and Sahakyan (2003) recently reported that older adults demonstrate lower accuracy of confidence judgments for forced cue-recall responses, being more likely to report high confidence that target options they are required to generate (typically, options that are normatively associated with the cue) were originally studied when they were actually not presented. Lower accuracy of confidence judgments was associated with poorer control of accuracy in free recall when individuals could withhold answers they were uncertain were correct. By analogy, it is possible that older adults' delayed strategy shift in the noun-pair task is caused by inaccurate monitoring of their recognition memory for the new associations. Experiencing instances of deciding to retrieve, only to then be incorrect, might aversively reinforce older adults' use of scanning over retrieval. To examine confidence in associative recognition memory, we included interspersed memory probes (as in Experiment 1), adding a recognition memory confidence rating after each probe. Dunlosky and Hertzog (2000) demonstrated highly accurate confidence judgments for older and younger adults in a paired-associate recall test, but this finding might not generalize to age equivalence in the accuracy of associative recognition confidence ratings for our noun pairs. Because we did not include misleading recognition lures (as in Kelley & Sahakyan, 2003), we hypothesized minimal age differences in the relative accuracy of these confidence judgments.

We also assessed retrieval monitoring with a cued-recall task for the noun pairs administered after noun-pair training. Individuals made JOLs by being cued with one element of the noun pair, predicting their confidence that they could recall its associated element a short time later. These JOLs after extensive noun-pair training are analogous to delayed JOLs (Connor et al., 1997) and should reflect monitoring retrieval of the associated nouns. We used the cued-recall approach to measure retrieval monitoring after skill acquisition because extensive noun-pair practice would result in near-ceiling associative recognition performance by the end of training. We hypothesized that participants in both age groups would show high (and equivalent) ability to monitor levels of item knowledge acquired in the noun-pair task.

## **Underconfidence in Recognition Memory**

Even if older adults are able to discriminate between items they had correctly recognized and those they had not (suggesting accurate monitoring), they could have lower mean confidence in their recognition responses. In metacognitive research, the former phenomenon is termed resolution (or relative accuracy) whereas the latter phenomenon is termed underconfidence (a manifestation of absolute accuracy; see Hertzog & Hultsch, 2000). Indeed, we believe that older adults' underconfidence in recognition memory was a more viable hypothesis to account for the delayed shift to the memory retrieval strategy than a deficit in retrieval monitoring, as measured by relative accuracy of confidence ratings. We hypothesized that underconfidence in recognition memory accuracy would be associated with older adults' retrieval strategy aversion.

In summary, Experiment 3 considered the relationship of the following metacognitive variables to strategy choice and performance in the noun-pair lookup task: (a) ratings of task-specific confidence in using the memory retrieval strategy in the noun-pair task, (b) ratings of memory self-concept and personal control over memory, (c) recognition memory confidence judgments, and (d) JOLs for cued-recall of the new associations made after noun-pair training.

## **Method**

### **Design**

To allow an investigation into whether making confidence judgments alters the course of noun-pair learning or the strategy shift, we randomly assigned equal numbers of young and older participants to conditions with or without confidence judgments. Thus, the between-subject independent variables were age and confidence judgment condition. The within-subject independent variable was practice.

### **Participants**

Forty young adults between the ages of 18–25 years and 40 older adults between the ages of 62–75 years participated in the study. Recruitment and compensation were the same as in Experiments 1 and 2. As in the previous experiments, age differences in background abilities (see Table 3) were comparable with those typically reported. We also administered a short version of the Personal Beliefs About Memory Inventory (PBMI; Lineweaver & Hertzog, 1998), which included scales measuring memory self-concept (or memory self-efficacy) and personal control over memory. Items were presented by a computer, which obtained item responses on a visual analog scale. The PBMI scales included in this study were as follows: Global Memory Ability (memory self-concept), Relative Standing (memory ability relative to one's own age peers), and Control Over Memory (see Lineweaver & Hertzog, 1998).

**Table 3**  
*Means (and Standard Errors) of Participant Characteristics for Experiment 3*

Measure	Young		Older	
	Control	Confidence	Control	Confidence
Age (years) <sup>a</sup>	20.6 (0.36)	20.3 (0.35)	68.0 (0.97)	67.6 (0.80)
Education (years)	15.0 (0.39)	14.3 (0.32)	15.6 (0.50)	15.8 (0.41)
Vocabulary <sup>a</sup>	31.4 (0.73)	31.8 (0.64)	34.4 (0.83)	35.1 (0.65)
Digit Symbol <sup>a</sup>	72.9 (3.57)	75.4 (2.25)	43.6 (3.28)	49.1 (3.07)
DS Memory <sup>a</sup>	0.8 (0.26)	3.4 (0.50)	0.8 (0.19)	3.9 (0.44)
No. medications <sup>a</sup>	0.8 (0.35)	0.7 (0.24)	2.5 (0.49)	2.7 (0.55)

*Note.* Vocabulary refers to the number correct out of 40. Digit Symbol = Wechsler Adult Intelligence Scale—Revised Digit Symbol subtest; DS Memory = symbol recall memory following the Digit Symbol subtest; No. medications = self-reported number of daily medications.

<sup>a</sup> Age comparison significant at  $p < .05$ .

### *Means (and Standard Errors) of Participant Characteristics for Experiment 3*

## **Procedure**

The apparatus and general procedure for the noun-pair task were identical to Experiment 1 except where noted. Across training, all participants responded to 60 repetitions for each 12 noun-pair stimuli. Accordingly, the table of noun pairs presented at the top of the screen contained only 12 noun pairs. Because of the smaller set size (relative to Experiments 1 and 2), we were able to conduct Experiment 3 in a single session. Training consisted of between 30 and 40 presentation blocks. <sup>4</sup> A training block contained 36 randomly ordered trials, with two trial target repetitions and one memory probe for each of the 12 noun-pair stimuli.

All participants responded to recognition memory probes of the same form as in Experiment 1. For participants in the condition with confidence judgments, memory probes were followed by the prompt, “How confident are you that the answer you just gave is correct? Using the marked keys, indicate your percent confidence from 0% to 100%.” Keys on the number pad were marked in decades (i.e., 0, 10, 20 ... 100).

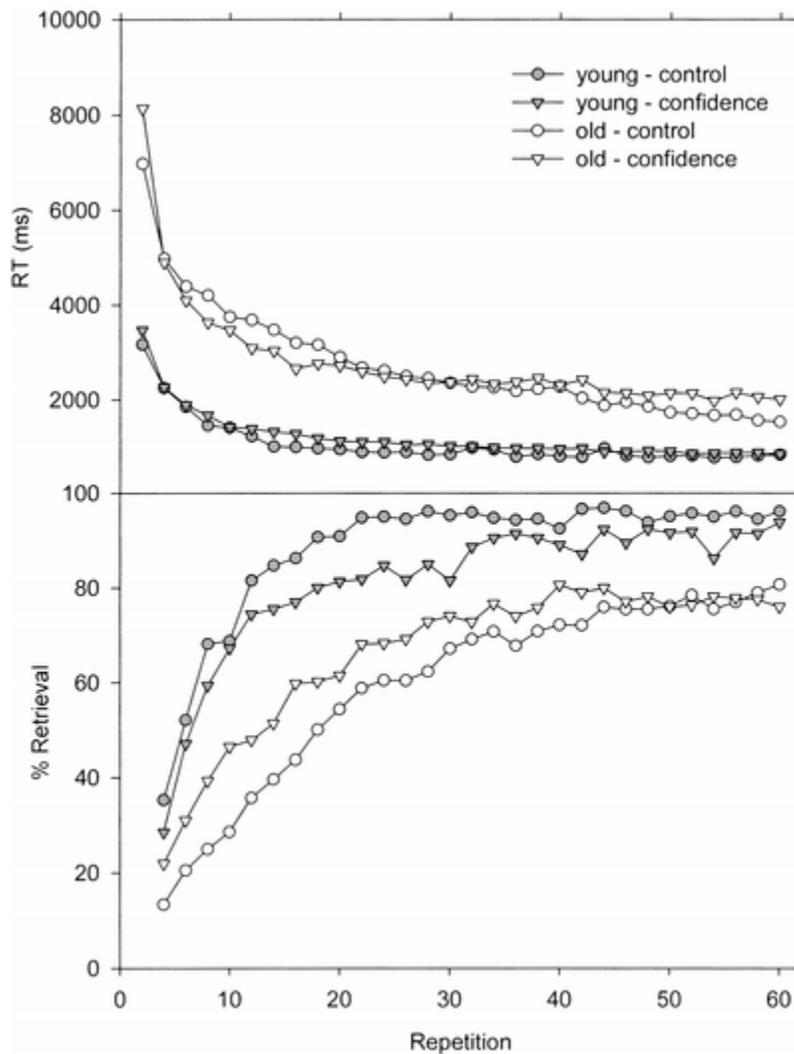
Following the computer testing, we collected JOLs—ratings of confidence in likelihood of later recall—for each noun-pair item. Participants were presented with the left-hand word of each pair and were asked, “How confident are you that you can recall the second word of the pair when prompted with the first word, 0% (won’t recall) to 100% (definitely will recall).” JOL confidence levels were scaled in decades, similar to the recognition memory confidence judgments. Participants then completed a cued-recall task, in which they were given the first word of each pair and asked to write in the second word. Finally, participants estimated the percentage of pairs memorized during training, a variable we do not report in this article. Additional posttest survey questions were identical to Experiment 2, 100% prelearning condition.

## Results and Discussion

### Noun-Pair Trials

#### *RT improvements*

Young adults responded consistently faster than older adults,  $F(1, 75) = 88.28$ ,  $MSE = 1,574,117,249$ ,  $p < .01$ ,  $r = .50$  (see Figure 5, top panel). Neither the main effect of confidence condition nor the interaction of Age  $\times$  Condition was significant ( $ps > .50$ ). RT improved with repetitions,  $F(29, 2175) = 132.14$ ,  $MSE = 56,837,247$ ,  $p < .01$ , and young adults improved more rapidly than older adults,  $F(29, 2175) = 20.90$ ,  $MSE = 8,990,931$ ,  $p < .01$ . Both the interaction of Condition  $\times$  Repetitions,  $F(29, 2175) = 1.58$ ,  $MSE = 679,790$ ,  $p < .01$ , and the interaction of Age  $\times$  Condition  $\times$  Repetitions were significant,  $F(29, 2175) = 1.99$ ,  $MSE = 856,903$ ,  $p < .01$ .



*Figure 5. Experiment 3 mean median reaction times (RTs; in milliseconds; top panel) and mean percentage retrieval strategy use (bottom panel) by age, confidence judgment condition, and repetition*

These outcomes can be interpreted as follows. Older adults who made confidence judgments improved their noun-pair RT more rapidly than those who did not, whereas the reverse pattern was found for young adults. Perhaps producing confidence judgments makes actual item knowledge more salient, causing typically underconfident older adults to be less conservative in strategy selection (opting more often for retrieval) whereas causing typically overconfident young adults to be more conservative in strategy selection.

### **Retrieval strategy use**

Young adults consistently used retrieval more often than older adults,  $F(1, 75) = 23.14$ ,  $MSE = 271,501$ ,  $p < .01$ ,  $r = .30$  (see Figure 5, bottom panel). Neither the main effect of confidence condition nor the interaction of Age  $\times$  Condition was significant ( $ps > .13$ ). However, the nonsignificant outcome for the Age  $\times$  Condition interaction is surprising, given the trends shown in the bottom panel of Figure 5, with differing effects of confidence judgments on retrieval probabilities in the two age groups (as for RTs above). Confidence judgments appeared to increase retrieval use in older adults and decrease retrieval use in young adults. We suspect limited statistical power to be the culprit. Indeed, 2 older participants who made confidence judgments reported extremely low levels of retrieval strategy use and generated negative correlations of confidence judgments with recognition memory performance. Eliminating them reduced error variance (from  $MSE = 11,734$  to  $MSE = 9,797$ ) and produced a reliable Age  $\times$  Condition interaction,  $F(1, 73) = 5.38$ ,  $p < .05$ .

Retrieval use increased with repetitions,  $F(28, 2100) = 76.33$ ,  $MSE = 18,884$ ,  $p < .01$ , and this increase was more rapid for young adults than for older adults,  $F(28, 2100) = 3.75$ ,  $MSE = 926$ ,  $p < .01$ . Surprisingly, neither the Condition  $\times$  Repetitions interaction nor the Age  $\times$  Condition  $\times$  Repetitions interaction were significant ( $ps > .15$ ).

### **Recognition Memory Probes**

Age differences in accuracy and latency of recognition-memory-probe responses were qualitatively similar to those presented in Experiment 1. Therefore, in the interest of brevity, we do not present the detailed outcomes for this experiment.

To our surprise, no age difference in noun-pair confidence was observed ( $p > .53$ ; see Figure 6, bottom panel). Confidence increased with repetitions,  $F(29, 1160) = 50.58$ ,  $MSE = 4,771$ ,  $p < .01$ , and increases were comparable for young and older adults ( $p > .96$ ). Considered together with the recognition memory accuracy data, older adults actually appear to be overconfident in the accuracy of their recognition memory responses. Indeed, we computed the difference between the percentage of correct recognition memory responses and the percentage

confidence ratings; this measure can be treated as an indication of average over- or underconfidence in recognition memory accuracy. Older adults were, on average, overconfident in their recognition memory, but this age difference interacted with repetitions,  $F(28, 2156) = 3.68$ ,  $MSE = 894$ ,  $p < .01$ . Both young and older adults showed discrepant confidence-accuracy relations early in practice, but older adults persisted in being overconfident whereas young adults did not. At the end of training, the mean difference was 23% for older adults but only 5% for young adults. These data refute the hypothesis that older adults' lower mean levels of confidence in the accuracy of associative recognition memory drives age differences in the timing of the strategy shift. Young and older adults showed similar relative accuracy of confidence judgments for discriminating recognition memory outcomes. Goodman-Kruskal gamma correlations were used to assess the ordinal association of confidence judgments with recognition memory accuracy over the entire set of repetitions (Nelson, 1984). Gammas were high (and reliably different from zero) for both age groups (see Table 4). The mean gammas differed between age groups,  $F(1, 40) = 5.65$ ,  $MSE = 0.196$ ,  $p < .05$ , but this appeared to be caused by a few older adults with poor relative accuracy; median gammas (see Table 4) were .94 and .88 in young and older adults, respectively. Indeed, excluding 2 older adults with low gamma correlations ( $< .25$ ) reduced the age difference ( $M_{young} = 0.90$ ,  $SD = 0.02$ ;  $M_{old} = 0.83$ ,  $SD = 0.04$ ;  $p = .10$ ).

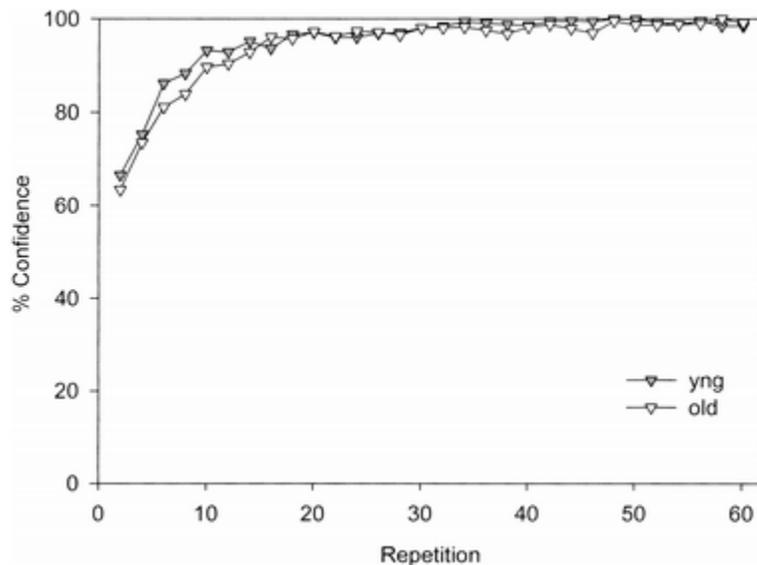


Figure 6. Experiment 3 mean confidence judgment by age, confidence judgment condition, and repetition. yng = young

**Table 4**  
*Means (and Standard Errors) and Medians for Correlations of Confidence With Recognition Memory Performance and Retrieval Strategy Use for Experiment 3*

Measure	Mean		Median	
	Young	Older	Young	Older
$\text{Gamma}_{\text{conf}(t)/\text{mem}(t)}^a$	0.90 (0.02)	0.76 (0.06)	0.94	0.88
$\text{Gamma}_{\text{conf}(t)/\text{ret}(t+1)}$	0.74 (0.04)	0.74 (0.09)	0.79	0.80

*Note.* For recognition memory performance, correlations were on the same trial for a given item; for retrieval strategy use, correlations were on the next trial for a given item.

<sup>a</sup> Age comparison significant at  $p < .05$ .

*Means (and Standard Errors) and Medians for Correlations of Confidence With Recognition Memory Performance and Retrieval Strategy Use for Experiment 3*

To assess age differences in strategy use based on recognition memory accuracy confidence, we also computed gamma correlations of confidence for an item with the likelihood of reported retrieval on the next repetition of the same item (see Table 4). These correlations were relatively high and reliably different from zero, indicating that both young and older adults were more likely to use a retrieval strategy for items that they had previously manifested high confidence in their recognition accuracy. There were no age differences in the magnitude of these correlations. The small age differences in the accuracy of confidence judgments did not translate into reliable age differences in the linkage of confidence to subsequent retrieval strategy use.

Thus, older adults showed little evidence of a deficit in relative accuracy for confidence judgments or correlations of confidence judgments with strategy use, in contrast to results from Kelley and Sahakyan (2003), in a task prone to memory illusions. Hence, age deficits in the accuracy of monitoring recognition memory accuracy cannot account for older adults' slower strategy shift in the noun-pair task.

Inserting confidence judgments into the noun-pair task increased older adults' rate of RT improvement and appeared to increase use of the memory retrieval strategy and to enhance item knowledge. The most plausible explanation is that reflecting on associative recognition memory confidence made older adults' incidental learning more salient to them, thereby increasing their use of the retrieval strategy. This effect exists above and beyond the benefit of providing memory probes, which also increased older adults' retrieval strategy use. It is important to note also that this effect is inconsistent with the learning deficit hypothesis.

### **Conditional Probability of Retrieval**

We again examined retrieval reluctance at the item level by comparing age differences in the probability of retrieval, conditional on correctly answering the previous memory probe for a given

noun pair. The conditional probability of retrieval was reliably lower for older adults than for young adults,  $F(1, 76) = 23.77$ ,  $MSE = 1.875$ ,  $p < .01$ ,  $r = .32$  (see Figure 7). Increases with practice in the conditional probability of retrieval were significant,  $F(2, 152) = 182.78$ ,  $MSE = 2.133$ ,  $p < .01$ , and occurred more rapidly for young adults,  $F(2, 152) = 3.29$ ,  $MSE = 0.038$ ,  $p < .04$ . Neither the main effect of confidence condition nor any interactions with the condition variable were significant ( $ps > .14$ ).

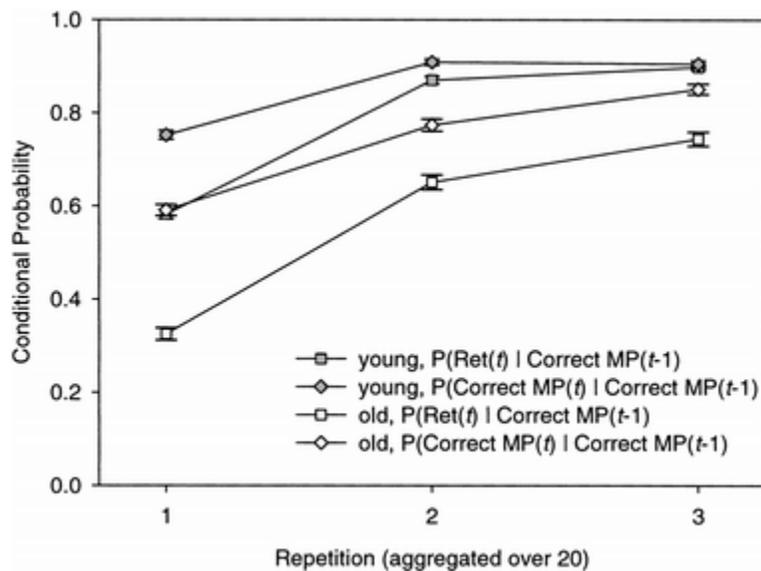


Figure 7. Experiment 3 mean probability of retrieval [ $P(\text{Ret}(t))$ ] and mean probability of a correct memory probe response [ $P(\text{Correct MP}(t))$ ], both conditional on success for an item's last memory probe [ $P(\text{Correct MP}(t - 1))$ ], by age and block (to increase precision, data were aggregated over 20 repetitions). Standard error bars are also shown

Age differences in consistency of recognition across trials were not responsible for this effect. Older adults were less stable in their item-level accuracy than were young adults,  $F(1, 76) = 15.82$ ,  $MSE = 0.573$ ,  $p < .04$ ,  $r = .24$ . Improvements in the consistency of memory-probe accuracy were significant,  $F(2, 152) = 151.50$ ,  $MSE = 0.659$ ,  $p < .01$ , and were more rapid for young adults,  $F(2, 152) = 10.00$ ,  $MSE = 0.044$ ,  $p < .01$ . However, age differences in the conditional probability of retrieval persisted when controlling on the covariate of consistency of recognition memory,  $F(1, 79) = 177.17$ ,  $MSE = 2.50$ ,  $p < .01$ . These results were fully consistent with those presented in Experiment 1, demonstrating retrieval reluctance in older adults that could not be accounted for by inconsistent access to noun-pair knowledge.

## Additional Metacognitive Variables

### Cued-recall JOLs

The cued-recall task provided additional evidence that older adults were able to monitor levels of noun-pair item learning. As expected, mean JOLs and cued-recall performance were reliably higher for young adults than for older adults (see Table 5). Recall tended to be better in the confidence condition than the no-confidence condition for older adults, and not young adults, but this effect was not reliable ( $p > .10$ ). Of most importance, gamma correlations were universally high, indicating good discrimination of levels of acquired item knowledge. The lower mean gamma correlation for older adults in the confidence condition was associated with an inflated standard error of estimate because of poor marginal distributions of recall or JOLs (see Hertzog et al., 2002, for further discussion of this phenomenon). Median gammas in each condition were 1.0 for all age groups. No reliable age differences in gamma were observed. Another interesting phenomenon emerged in the data. Older adults' recognition memory experience apparently produced substantial overconfidence in their cued-recall JOLs. Mean JOLs, reflecting subjective confidence in probability of cued recall, were higher than cued-recall probabilities for both age groups, reliably so for older adults. This is an unusual outcome because practice within the same task has been found to be associated with increasing underconfidence in JOLs in other recall tasks (Hertzog et al., 2002; Koriat, Sheffer, & Ma'ayan, 2002). Certainly, this effect is inconsistent with the hypothesis of underconfidence in learning by older adults.

Table 5  
*Age and Condition Differences in Means (and Standard Errors) of Metacognitive Variables for Experiment 3*

Measure	Young		Older	
	Control	Confidence	Control	Confidence
Memory beliefs scales				
PBMI Global	71.7 (3.39)	70.0 (2.88)	65.5 (3.81)	72.0 (2.80)
PBMI Relative	67.2 (1.88)	64.5 (2.37)	65.2 (2.24)	66.8 (1.75)
PBMI Control	67.8 (2.29)	69.8 (2.27)	72.6 (3.20)	73.0 (2.76)
Judgment of learning task				
<i>M</i> <sup>a</sup>	91.1 (3.00)	91.0 (2.09)	72.9 (5.52)	80.9 (5.07)
Cued recall <sup>a</sup>	87.5 (2.80)	86.6 (3.26)	58.3 (6.62)	70.4 (6.96)
Gamma <sup>b</sup>	0.86 (0.10)	0.71 (0.14)	0.91 (0.04)	0.54 (0.21)
Posttask ratings				
Effort <sup>a</sup>	2.3 (0.25)	2.6 (0.22)	3.3 (0.20)	3.2 (0.22)
Improvement	1.5 (0.25)	1.7 (0.28)	1.9 (0.28)	2.0 (0.30)
Global confidence <sup>a</sup>	1.5 (0.22)	1.8 (0.26)	2.7 (0.20)	2.5 (0.28)

*Note.* PBMI = Personal Beliefs About Memory Scale; Global = Global Memory Ability subscale; Relative = Relative Standing subscale; Control = Control Over Memory subscale; *M* (judgment of learning task) = average percent recall ability confidence, across items; Cued recall = average percent posttask second-word recall with first-word cue, across items; Gamma = mean intraindividual ordinal correlation between item judgment of learning and item recall; Effort = self-rated effort required for memorization; Improvement = self-rated perceived improvement offered by the memory strategy for task performance; Global confidence = self-rated confidence in using the memory strategy.

<sup>a</sup> Age comparison significant at  $p < .05$ . <sup>b</sup> The condition comparison is significant at  $p < .05$ ; this results from extreme (-1) values from a few participants in the confidence judgments conditions, as medians for all groups were 1.

## *Age and Condition Differences in Means (and Standard Errors) of Metacognitive Variables for Experiment 3*

### **Global and task-specific beliefs**

There were no reliable age or condition differences in global memory self-concept, relative standing, or perceived control over memory, as measured by the PBMI (see Table 3). The lack of age group differences was surprising, given previous work with these scales (Lineweaver & Hertzog, 1998). In contrast, older adults were less confident in their ability to rely on the memory retrieval strategy in the noun-pair task and rated using the memory retrieval strategy as more effortful than did young adults.

### **Metacognition-retrieval correlations**

To further evaluate the relationship between retrieval use and metacognition, we analyzed correlations of individual differences in retrieval use at the last block of training with memory ability and metacognitive reports (see Table 5). Given limited sample size, the correlations were computed in the aggregate sample (collapsing over age group) and should be treated with some caution. Retrieval use was not predicted by general beliefs about personal control over memory or memory self-efficacy, as measured by the three PBMI scales, Global Rating, Relative Standing, or Personal Control ( $r_s = -.01-.13$ ). These kinds of beliefs are related to strategic behavior in some tasks (e.g., Hertzog, Dunlosky, Robinson, & Kidder, 2003; Hertzog, Lineweaver, & McGuire, 1999) but not the strategy shift being studied here.

In contrast, confidence in ability to use the retrieval strategy did correlate with retrieval strategy use ( $r = -.36$ ), replicating results from Experiment 2. Individual differences in underlying item knowledge at the end of practice, as reflected in cued-recall performance, were strongly correlated with retrieval strategy use ( $r = .71$ ), and these differences were also reliably correlated with average confidence in underlying learning (mean JOLs) at the end of practice ( $r = .60$ ). The correlation of cued-recall performance with strategy use might be taken as evidence for the learning deficit hypothesis (strategy use correlated with level of learning). However, level of final learning (i.e., item knowledge) is also influenced by the degree of use of the retrieval strategy, which reinforces learning (Allen et al., 1969).

Thus, the metacognitive variables indicate that avoidance of the retrieval strategy is associated primarily with lower confidence in the ability to use that strategy in the noun-pair task. This outcome indicates that older adults' retrieval strategy aversion is not a generic manifestation of negative memory self-concept or personal control beliefs but rather is specific to the noun-pair task context.

## GENERAL DISCUSSION

The present set of experiments conclusively demonstrates that older adults show an aversion to using the memory retrieval strategy in the noun-pair lookup task. Certainly, rates of learning new associations between noun pairs were slower for older adults than for young adults, and an age-related associative learning deficit undoubtedly contributed to the slower strategy shift.

However, older adults transitioned from a noun-pair scanning strategy to a retrieval strategy more slowly than was warranted by their level of item knowledge in all three experiments, arguing against a simple learning deficit account of age differences in strategy shift.

Manipulations such as introducing recognition memory probes and prelearning of the associations accelerated the rate of a strategy shift for older adults, but robust age differences remained, even under these conditions. Age differences in strategy selection were reliably associated with stated lack of confidence in using the retrieval strategy, implicating subjective factors as playing a role in the delayed strategy shift.

The delayed strategy shift also cannot be attributed to metacognitive factors such as beliefs about memory (low memory self-concept, sense of personal control over memory) or to failures in monitoring recognition memory accuracy. Neither underconfidence in recognition memory responses nor reduced accuracy of retrieval monitoring, as manifested by correlations of confidence ratings with recognition outcomes, could account for older adults' slower strategy shift. The latter finding does not rule out the possibility that older adults may not spontaneously use their intact monitoring skills to guide strategy choice in standard noun-pair task conditions (i.e., when recognition memory probes and confidence ratings are not provided). That is, older adults may not attempt to monitor accuracy of memory retrieval unless the task requires it (as it did in Experiment 3). Moreover, we observed 2 older adults who had low correlations of confidence judgments with recognition accuracy and retrieval strategy use, and these participants tended to persist in scanning behavior throughout practice. Thus, failures in monitoring may account for delayed strategy shift in some older individuals, even if it cannot account for the general tendency for older adults to shift more slowly to the retrieval strategy.

It is also possible that metacognitive variables we have not yet measured are more critical for rapid strategy selection and would better explain older adults' retrieval strategy aversion. Confidence in the accuracy of recognition memory may be necessary but not sufficient for selecting the retrieval strategy on a given noun-pair lookup trial. Assuming that adaptive strategy choice (Siegler & Lemaire, 1997) occurs rapidly, a fast feeling of knowing on presentation of the target noun-pair may be the basis for choosing the retrieval strategy (Schunn et al., 1997). We have demonstrated that older adults have good discrimination of items that are known from items that are not known, on the basis of both JOLs and confidence judgments for recognition memory probes. However, these judgments do not simulate the real-time dynamics of adaptive strategy choice required by the noun-pair task. The next logical step is to evaluate the dynamics and accuracy of older adults' feeling-of-knowing judgments. It is possible, for example that their feeling-of-knowing judgments are less accurate in the noun-pair context (Souchay et al., 2000) or that the latency for feeling of knowing is sufficiently slow to cause older adults to routinely scan rather than wait for a high feeling of knowing to select the retrieval strategy. Moreover, Schunn et al. (1997) argued that strategy choice in such tasks is actually

implicit not explicit. It is therefore possible that unconscious activation of implicit beliefs about aging and memory (e.g., Hummert et al., 2002) in the task context, rather than the explicit memory beliefs measured by the scales from Lineweaver and Hertzog (1998), would predict retrieval aversion.

Future research should also focus on additional alternative explanations of the slower strategy shift. Older adults may differ in implicit task goals, resulting in different subjective values for the costs and benefits of the two strategies. For example, they may weigh errors as more costly and weigh fast RTs as being less important than accuracy, thereby avoiding the retrieval strategy because it risks errors early in practice. They may have a deficient mental model for noun-pair task improvements, failing to realize that the retrieval strategy produces more efficient performance. Finally, they may be more likely to routinely stick with a strategy that seems to work, even if other strategies might prove more beneficial (in essence, a kind of behavioral inertia).

Whatever explanatory variables are ultimately needed to account for the slower shift, it is clear that observed age differences in practice-related improvements in the noun-pair task overestimate the degree of an associative learning deficit in older adults (see Rogers et al., 2000). Older adults appear to have a more conservative criterion for selecting the retrieval strategy, and their continuing reliance on scanning slows their rate of noun-pair practice improvement. An interesting question, then, is whether the type of conservatism we have observed here generalizes to other experimental tasks and everyday contexts, such as driving behavior or willingness to learn new technologies, such as computer software. In general, gerontologists should be conservative in making inferences about the causes of skill-acquisition impairments in older adults. They must be wary of concluding that older adults are impaired in learning new skills solely because of neurobiological aging, given that their strategic behavior can be manipulated so as to improve rates of skill acquisition.

## FOOTNOTES

1. All RT presentations and analyses involve group means of participant medians. Medians are used to remove the influence of outliers at the participant level, which occur infrequently because of fast correct guessing, motivational lapses, or distractions.
2. Effect size correlations are provided as measures of comparison magnitudes independent of sample size. We have used the effect size correlation computed from Cohen's  $d$ , with difference standardized by pooled variance (see Rosenthal & Rosnow, 1991).
3. Because these functions are nonlinear and data were not smoothed, it is possible that the obtained effects of repetition indicate idiosyncratic changes rather than meaningful improvements. In consideration of this possibility, we test all such data for polynomial trends. For all analyses, the linear and quadratic trends accounted for greater than 75% of the variance in repetitions (unless otherwise noted), supporting the effects as fundamental differences in the shapes of the functions.
4. We initially intended for participants to respond to 80 repetitions per stimulus, but some were not able to complete 80 repetitions within the one session format. Analyses are constrained to the first 60 repetitions, as all participants completed at least that many.

## REFERENCES

- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*, 3–27.
- Ackerman, P. L., & Woltz, D. J. (1994). Determinants of learning and performance in an associative memory/substitution task: Task constraints, individual differences, volition, and motivation. *Journal of Educational Psychology*, *86*, 487–515.
- Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *8*, 463–470.
- Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, *51*, 355–365.
- Bahrick, H. P. (2000). Long-term maintenance of knowledge. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 347–362). London: Oxford University Press.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
- Berry, J. M. (1999). Memory self-efficacy in its social cognitive context. In T. Hess & F. Blanchard-Fields (Eds.), *Social cognition and aging* (pp. 69–96). San Diego, CA: Academic Press.
- Charness, N., & Campbell, J. I. D. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. *Journal of Experimental Psychology: General*, *117*, 115–129.

- Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12*, 50–71.
- Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13*, 597–607.
- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15*, 462–474.
- Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*, 725–747.
- Fisk, A. D., & Rogers, W. A. (1991). Toward understanding of age-related memory and visual search effects. *Journal of Experimental Psychology: General, 120*, 131–149.
- French, J. W., Price, L. A., & Thurstone, L. L. (1962). *French's kit of referenced tests*. Princeton, NJ: Educational Testing Service.
- Hertzog, C., Cooper, B. P., & Fisk, A. D. (1996). Aging and individual differences in the development of skilled memory search performance. *Psychology and Aging, 11*, 497–520.
- Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue utilized for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 22–34.
- Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., (pp. 417–466). Mahwah, NJ: Erlbaum.
- Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17*, 209–225.
- Hertzog, C., Lineweaver, T. T., & McGuire, C. L. (1999). Beliefs about memory and aging. In F. Blanchard-Fields & T. M. Hess (Eds.), *Social cognition and aging* (pp. 43–68). New York: Academic Press.
- Hertzog, C., Vernon, M. C., & Rypma, B. (1993). Age differences in mental rotation task performance: The influence of speed/accuracy tradeoffs. *Journals of Gerontology, 48*, 150–156.
- Hoyer, W. J., Cerella, J., & Onyper, S. V. (2004). Item learning in cognitive skill training: Effects of item difficulty. *Memory & Cognition, 31*, 1260–1270.
- Hummert, M. L., Garstka, T. A., & O'Brien, L. T. (2002). Using the implicit association test to measure age differences in implicit social cognitions. *Psychology and Aging, 17*, 482–495.

- Johnson, M. K., & Raye, C. L. (2000). Cognitive and brain mechanisms of false memories and beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 35–86). Cambridge, MA: Harvard University Press.
- Kausler, D. H. (1994). *Learning and memory in normal aging*. San Diego, CA: Academic Press.
- Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48*, 704–721.
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General, 131*, 147–162.
- Lachman, M. E., Bandura, M., Weaver, S. L., & Elliot, E. (1995). Assessing memory control beliefs: The Memory Controllability Inventory. *Aging and Cognition, 2*, 67–84.
- Lineweaver, T. T., & Hertzog, C. (1998). Adults' efficacy and control beliefs regarding memory and aging: Separating general from personal beliefs. *Aging, Neuropsychology, and Cognition, 5*, 264–296.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.
- Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 362–372.
- McDonald-Miszczak, L., Hertzog, C., & Hulstsch, D. F. (1995). Stability and accuracy of metamemory in adulthood and aging: A longitudinal analysis. *Psychology and Aging, 10*, 553–564.
- Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1170–1187.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109–133.
- Rabbitt, P., & Abson, V. (1990). Do older people know how good they are? *British Journal of Psychology, 82*, 137–151.
- Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16*, 323–341.
- Reder, M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 435–451.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, *126*, 288–310.

Rogers, W. A., & Gilbert, D. K. (1997). Do performance strategies mediate age-related differences in associative learning? *Psychology and Aging*, *12*, 620–633.

Rogers, W. A., Hertzog, C., & Fisk, A. D. (2000). An individual differences analysis of ability and strategy influences: Age-related differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 359–394.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw Hill.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Erlbaum.

Salthouse, T. A., & Dunlosky, J. (1995). Analyses of adult age differences in associative learning. *Zeitschrift fuer Psychologie*, *203*, 351–360.

Salthouse, T. A., & Somberg, B. L. (1982). Time-accuracy relationships in young and old adults. *Journal of Gerontology*, *37*, 349–353.

Schneider, W., & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 53–119). San Diego, CA: Academic Press.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. J. (1997). To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 3–29.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

Siegler, R. S., & Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experimental Psychology: General*, *126*, 71–92.

Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology*, *14*, 299–309.

Strayer, D. L., & Kramer, A. F. (1994). Aging and skill acquisition: Learning-performance distinctions. *Psychology and Aging*, *9*, 589–605.

Touron, D. R., & Hertzog, C. (2004). Strategy shift affordance and strategy choice in young and older adults. *Memory & Cognition*, *32*, 298–310.

Touron, D. R., Hoyer, W. J., & Cerella, J. (2001). Cognitive skill acquisition and transfer in younger and older adults. *Psychology and Aging*, *16*, 555–563.

Touron, D. R., Hoyer, W. J., & Cerella, J. (in press). Cognitive skill learning: Age-related differences in strategy shifts and speed of component operations. *Psychology and Aging*.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 582–600.

Vickers, D. (1979). *Decision processes in visual perception*. New York: Academic Press.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corporation.

Widner, R. L., Otani, H., & Smith, A. D. (2000). Hypermnesia: Age-related differences between young and older adults. *Memory & Cognition*, 28, 556–564.

Zachary, R. (1986). *Shipley Institute of Living Scale Revised manual*. Los Angeles: Western Psychological Services.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., (pp. 293–357). Mahwah, NJ: Erlbaum.