Does a time monitoring deficit influence older adults' delayed retrieval shift during skill acquisition?

By: Christopher Hertzog, Dayna R. Touron, & Jarrod C. Hines

Hertzog, C., Touron, D. R., & Hines, J. (2007). Does a time monitoring deficit influence older adults' delayed retrieval shift during skill acquisition? Psychology and Aging, 22, 607-624. DOI: 10.1037/0882-7974.22.3.607

Abstract:
The authors evaluated age-related time-monitoring deficits and their contribution to older adults' reluctance to shift to memory retrieval in the noun-pair lookup (NP) task. Older adults (M = 67 years) showed slower rates of response time (RT) improvements than younger adults (M = 19 years), because of a delayed strategy shift. Older adults estimated scanning latencies as being faster than they actually were and showed poor resolution in discriminating short from long RTs early in practice. The difference in estimated RT for retrieval and scanning strategies predicted retrieval use, independent of actual RT differences. Separate scanning and recognition memory tasks revealed larger time-monitoring differences for older adults than in the NP task. Apparently, the context of heterogeneous RTs as a result of strategy use in the NP task improved older adults' accuracy of RT estimates. RT feedback had complex effects on time-monitoring accuracy, although it generally improved absolute and relative accuracy of RT estimates. Feedback caused older adults to shift more rapidly to the retrieval strategy in the NP task. Results suggest that deficient time monitoring plays a role in older adults' delayed retrieval shift, although other factors (e.g., confidence in the retrieval strategy) also play a role.

Article:
Monitoring internal states and behavioral outcomes is a critical component of adaptive self-regulation in cognitive tasks (Koriat, Goldsmith, & Pansky, 2000; Nelson, 1996; Shallice & Burgess, 1991). Research on metamemory and aging indicates that older adults monitor encoding and retrieval processes with equivalent accuracy compared with young adults, despite major age differences in episodic memory performance itself (e.g., Bäckman & Karlsson, 1985; Butterfield, Nelson, & Peck, 1988; Connor, Dunlosky, & Hertzog, 1997; Eakin & Hertzog, 2006; Hertzog, Kidder, Powell-Moman, & Dunlosky, 2002; Lovelace & Marsh, 1985; but see Souchay, Isingrini, Clarys, Taconnat, & Eustache, 2004). Studies of memory monitoring typically correlate metacognitive judgments with recall or recognition to assess monitoring accuracy. Relatively little attention has been paid to adults' accuracy in monitoring the dynamics of memory, such as the time taken to recall or recognize previously learned material.

The literature on monitoring temporal duration has found robust age differences in time monitoring on a variety of time estimation tasks, most emphasizing the duration of simple perceptual events or performed actions (e.g., McCormack, Brown, Maylor, Richardson, & Darby, 2002; Salthouse, Wright, & Ellis, 1979; Wearden, Wearden, & Rabbitt, 1997). Block, Zakay, and Hancock's (1998) meta-analysis concluded that older adults overestimated temporal durations relative to younger adults. Craik and Hay (1999) argued that Block et al.'s (1998) inferences applied to simple tasks with blank or unfilled intervals between onset and duration judgment rather than filled intervals in which people were actively engaged in other tasks. They found that verbal estimates of temporal judgments with filled intervals underestimated temporal duration, whereas prospective production (requiring a response after a fixed amount of time) typically overestimated duration. Older adults' time estimates diverged more from actual durations in both types of tasks (see also McCormack et al., 2002). Craik and Hay (1999) suggested that aging affects both an internal clock and attention devoted to the passage of time that is constrained under divided attention or other resource demands.

Time-based prospective memory tasks, which typically require participants to monitor time while engaged in a demanding primary task, have also indicated age differences in time-monitoring accuracy (e.g., Einstein, McDaniel, Richardson, Guynn, & Cunfer, 1995; Logie, Maylor, & Della Salla, 2004; Martin & Schumann-Hengsteler, 2001; Park, Hertzog, Kidder, Morrell, & Mayhorn, 1997). One interpretation of these findings is that absorption in the primary task eliminates the intention to monitor time from current awareness and that an executive failure to reinstate clock checking is the culprit for older adults' poorer performance. However, a simple deficit in time monitoring per se may also be involved (Park et al., 1997).

It is an open question whether this age-related deficit in monitoring processing time has a functional impact on older adults' performance in speeded tasks. Robinson, Hertzog, and Dunlosky (2006) asked younger and older adults to learn paired-associate items using the interactive imagery strategy. Participants pressed a key as soon as the image had been formed. Individuals also made a judgment of learning (JOL), scaling their confidence in the likelihood of later item recall, and then estimated how long it had taken them to form the image. There were reliable age differences in the absolute accuracy of the latency estimates, with older adults showing larger differences between subjective and actual latency. However, the ordinal correlation of subjective latency with JOLs was equivalent for the two groups. This result suggests that (a) older adults have a relatively accurate sense of ordinal differences in image formation time (i.e., they can discriminate long from short imagery latencies) and (b) this level of monitoring accuracy was sufficient to inform JOL ratings. Given that age differences in latency estimates are often small in magnitude (e.g., McCormack et al., 2002), the functional impact of age differences in time-monitoring accuracy latency has yet to be demonstrated.

This study evaluates a task in which age differences in latency monitoring could have a material impact on age differences in performance: Ackerman and Woltz's (1994) noun-pair (NP) lookup task. The NP task requires individuals to determine whether a target pair of nouns is matched by one of a set of pairs shown in a lookup table at the top of the computer screen. After repeated exposure to the same item pairs, individuals can recognize the correct pairing and respond on the basis of memory retrieval without searching the lookup table for the matching pair. Shifting to memory retrieval yields major benefits for the latency of correct match responses, because successful match decisions based on memory retrieval occur, on average, more than 1 s faster than successful matches based on visual scanning.

Rogers and Gilbert (1997) reported that older adults were slower to shift to the memory retrieval strategy (see also Rogers, Hertzog, & Fisk, 2000). Some older adults relied on scanning even after hundreds of repetitions of the same NPs. Touron and Hertzog (2004a, 2004b) showed that this delay was not due to impaired associative learning alone; some older adults chose to scan even when they correctly recognized the pairings (see Touron, 2006). Moreover, monetary incentives hastened older adults' retrieval shift (Touron, Swaim, & Hertzog, 2007).

Why would older adults avoid retrieval even when their level of item learning affords its use? This study evaluates whether age differences in the accuracy of monitoring response latency might contribute to the phenomenon. A time-monitoring deficit could contribute to delayed retrieval shift under the assumption that individuals choose to retrieve or scan on the basis of the relative perceived costs and benefits of each strategy. Scanning is highly accurate but slow, whereas retrieval (especially early in practice) is fast but risks errors. To be willing to risk shifting to the potentially more error-prone retrieval strategy, an individual must (a) perceive that response times (RTs) are faster when he or she uses memory retrieval than when he or she uses visual scanning, (b) generate an accurate task mental model that choosing to retrieve will result in substantial improvements in performance efficiency, and (c) opt to achieve greater efficiency by using memory retrieval whenever possible. By this account, accurate latency monitoring is needed for perceiving the retrieval strategy's superior efficiency.[1]

We tested the hypotheses that (a) older adults are deficient in the accuracy of response latency monitoring and (b) deficient latency monitoring is associated with a delayed shift to the memory retrieval strategy in the NP task. To do so, we asked individuals to estimate the latency of individual RTs and to report the strategy they had just used on that trial. We also asked individuals to estimate their average response latency for each strategy at the end of a trial block. The latency-monitoring hypothesis predicts that older adults will be less accurate in estimating response latencies and that this inaccuracy will predict delayed retrieval shift.

Experiment 1

Experiment 1 assessed the accuracy of two measures of RT monitoring during the NP task. First, individuals were asked to monitor RT for individual NP trials, estimating how long it had taken them to respond on a given trial. Second, at the end of a trial block, individuals were asked to estimate that block's RT for each type of strategy, visual scanning and memory retrieval. Research on performance monitoring shows that aggregate estimates (postdictions) of performance often behave differently from item-level confidence judgments (Dunlosky & Hertzog, 2000), supporting the distinction between monitoring item retrieval accuracy and making inferences about aggregate accuracy over a set of trials. One could argue that item-level RT estimates measure elementary time monitoring, whereas block-level RT estimates for each strategy are relevant to participants' inferences about the relative efficiency of the two strategies.

Experiment 1 also evaluated the effects of feedback on shaping RT monitoring accuracy. On alternating trials, individuals assigned to the feedback condition were either prompted to make latency judgments or given feedback about their actual RT latency. We hypothesized that feedback would improve the accuracy of RT monitoring in both age groups. One question of interest was whether it would differentially improve the RT monitoring of older adults.

Method

Design

The experiment was a 2 (age: young, old) × 2 (feedback: given, withheld) × 40 (presentation block) mixed design, with age and feedback on actual item RT as between-subjects factors and block as the within-subject factor.

Participants

Younger adult participants were undergraduates from Appalachian State University who received course credit for participation. Older adults were recruited from the community in and around Boone, North Carolina, and received a $40 honorarium for their participation. The sample consisted of 52 younger adults (ages 18 to 25 years) and 51 older adults (ages 60 to 75). Participants were screened for basic health issues that could impair performance, such as arthritis. Near visual acuity of at least 20/50, uncorrected, was required for participation. Participants' demographics and cognitive background measures are provided in Table 1. Typical patterns of age differences in these measures were observed. Older adults performed better on a vocabulary test but worse on a measure of speed, the Digit–Symbol Substitution Task (Wechsler, 1981), and on a test of incidental memory for the digit–symbol pairings used in the speed test.

Materials and Procedure

Participants were tested in groups of up to 3 persons, seated in a room with personal computers. The NP lookup task was programmed in Visual Basic 6.0. Stimuli were presented in a 15-point Arial font on a 15-in. LCD monitor with a resolution of 1024 × 768 pixels. Participants were seated at a height and distance that optimized their screen viewing and comfort.

The stimulus set contained 24 unrelated, concrete nouns. The lookup table, presented at the top of the screen, consisted of 12 pairs configured into four columns by three rows. Pairings in the lookup table did not change, although pair locations changed randomly between trials. The target pair was matched (i.e., identical) to one of the pairs in the lookup table for half the trials in each block of 24 trials, selected at random. For the other 12 trials in a block, an unmatched pair was constructed by random selection of two unmatched nouns from the table. Matching target pairs were always presented in an invariant order corresponding to the order in the lookup table (e.g., *IVY–BIRD* was never presented as *BIRD–IVY*). Forty blocks of 24 NP trials were administered; thus, each pair served as a target item 80 times during the experiment.

Table 1

Means and Standard Errors of Participant Characteristics for All Experiments

Measure	Young$_{feedback}$ *M*	Young$_{feedback}$ *SE*	Young$_{no\ feedback}$ *M*	Young$_{no\ feedback}$ *SE*	Old$_{feedback}$ *M*	Old$_{feedback}$ *SE*	Old$_{no\ feedback}$ *M*	Old$_{no\ feedback}$ *SE*
				Experiment 1				
Age (years)[a]	18.9	0.19	19.7	0.40	67.2	0.86	67.7	0.83
Education (years)[a,b]	12.8	0.22	13.3	0.26	15.6	0.61	16.9	0.35
No. medications[a]	0.9	0.21	0.7	0.20	3.0	0.40	3.2	0.43
Vocabulary[a,c]	31.2	0.45	30.1	0.63	34.6	0.70	36.6	0.43
Digit–Symbol[a,c]	67.0	1.76	63.3	1.60	44.8	1.90	49.1	2.02
Digit–Symbol memory[a]	7.6	0.36	7.0	0.42	5.1	0.41	5.3	0.49
				Experiment 2				
Age (years)[a]	20.7	0.45	20.6	0.46	66.2	0.98	66.0	0.97
Education (years)[a,b]	15.0	0.46	14.7	0.36	15.0	0.58	15.8	0.64
No. medications[a]	0.9	0.21	0.8	0.29	3.5	0.69	3.1	0.58
Vocabulary[a,c]	30.0	0.70	30.3	0.93	35.4	0.63	36.0	0.69
Digit–Symbol[a,c]	62.8	3.64	65.5	3.94	41.9	3.32	43.5	3.71
Digit–Symbol memory[a]	7.0	0.46	8.15	0.28	5.4	0.57	4.43	0.65
				Experiment 3				
Age (years)[a]	22.2	1.08	21.3	0.49	66.7	0.79	67.0	0.82
Education (years)[a]	14.7	0.21	15.1	0.23	16.7	0.52	15.8	0.64
No. medications[a]	1.0	0.25	1.38	0.27	2.4	0.29	2.3	0.39
Vocabulary[a]	29.5	0.74	29.2	0.79	35.2	0.82	35.2	0.58
Digit–Symbol[a,c]	65.6	1.69	69.1	2.24	52.1	2.46	46.3	2.45
Digit–Symbol memory[a]	7.0	0.39	7.65	0.30	5.7	0.50	5.68	0.47

Note. Vocabulary = number correct out of 40 on the Shipley Vocabulary Test (Zachary, 1986); Digit–Symbol = Wechsler Adult Intelligence Scale Digit–Symbol subtest (Wechsler, 1981); Digit–Symbol memory = symbol recall memory following the Wechsler Adult Intelligence Scale Digit–Symbol subtest (Wechsler, 1981).

[a] Age main effect, $p < .05$. [b] Feedback main effect, $p < .05$. [c] Age × Feedback interaction, $p < .05$.

A 500-ms fixation point initiated each trial, replaced by the target NP in the same location, concurrently with the lookup table. Participants responded by pressing keys labeled *Y* and *N* (i.e., "Yes, the target pair matches a pair in the table" or "No, it does not"). After that response, participants reported how they had made their decision by pressing keys corresponding to (a) visual scanning, (b) memory retrieval, (c) both scanning and memory retrieval, and (d) other strategies. The latter two categories were infrequently used (fewer than 10% of correct responses) and are ignored in our analyses.

Participants in both conditions provided RT estimates on half the trials. Following each odd-numbered trial, participants in the RT feedback condition were given their actual RT in seconds (e.g., "Your speed was 2.1 seconds") for that trial. Participants in the no-feedback condition received only a blank time interval of the same length. After each even-numbered trial, all participants estimated their RT for the preceding trial immediately after the strategy report. They did so by choosing one of nine response options, which were divided into half-second intervals beginning with "less than .50 seconds" and ending with "greater than 4.0 seconds."

At the end of each block, all participants estimated their average RTs, separately for the scanning trials and the retrieval trials within the preceding block. Block-level RT estimates used the same 1–9 scale that had been used to collect the trial-level RT estimates, with instructions for participants to respond "0" if they never used the strategy during the preceding block. After making block-level RT estimates, participants were given response accuracy feedback detailing the percentage of correct matching trials achieved during the previous block. Participants were asked to maintain 90% accuracy during the NP task. If participants were 100% accurate during a block, they were told to speed up; if they had more than two inaccurate responses out of 24 trials during a block, they were told to slow down. Finally, participants were shown their percentage of memory usage and their average overall RT for the preceding block of trials.

A posttest questionnaire was given following the completion of the NP task. Participants first estimated their general confidence in their ability to recall the second word in each pair when presented with the first word. They scaled their confidence by circling one of the options of 0% confidence, 10% confidence, 20% confidence, and so on through 100% confidence. They were also asked to estimate the percentage of pairs memorized during the computer task using the same scale. Participants were then given a list of first-word cues and asked to make a JOL for each cue. They were asked, “How confident are you that you can recall the second word of the pair when prompted with the first word?” They responded by circling an option between 0% and 100% in 10% increments. After giving a JOL for all items, participants were presented with the list of cues and asked for cued recall, writing the second word in each pair after its associated cue.

Finally, participants answered the following questions by responding on a 5-point Likert scale (labels for the endpoints of 1 and 5 were provided): (a) “Were you confident to use your memory?” (1 = *yes*, 5 = *no*), (b) “How much effort did it require for you to memorize the word pairs?” (1 = *automatic*, 5 = *effortful*), and (c) “How much does using memory improve performance on this task?” (1 = *very much*, 5 = *not at all*).

Results

If not otherwise specified, data were analyzed in a 2 (age) × 2 (feedback) × 30 (block) general linear model on each dependent variable. Because some participants did not complete all 40 blocks of training and all participants approached asymptote after Block 30, Blocks 31–40 were omitted from analysis. We report multivariate F tests to test repeated measures block-related effects (McCall & Appelbaum, 1973) because sphericity and circularity assumptions were usually violated.

Response Accuracy

Older adults had a higher percentage of correct responses (M = 94.4%, SE = 0.15%) than younger adults (M = 91.9%, SE = 0.16%), $F(1, 91) = 10.74$, $MSE = 379$, $p < .01$. There was no reliable effect of feedback, $F(1, 91) = 3.77$, $MSE = 379$, $p > .05$. Accuracy varied as a function of practice block, $F(29, 2639) = 1.53$, $MSE = 36$, $p < .05$, but this was not obviously a function of a systematic trend with increasing practice (lower order polynomial trends were not reliable). No other main effects or interactions were found.

Actual RTs

The patterns in RT improvement replicated those in our earlier studies with the NP task (e.g., Touron & Hertzog, 2004a). Younger adults responded faster than did older adults, $F(1, 91) = 88.80$, $MSE = 15{,}195{,}157$, $p < .01$. RTs decreased with practice, multivariate $F(29, 63) = 28.39$, $p < .01$. A significant Block × Age interaction also emerged, $F(29, 63) = 3.05$, $MSE = 439{,}128.8$, $p < .01$. This interaction was associated with reliable quadratic, $F(1, 91) = 9.92$, $MSE = 665{,}789$, $p < .01$, and cubic, $F(1, 91) = 7.40$, $MSE = 263{,}634$, $p < .01$, trends, with younger adults' RT curve decreasing more quickly and approaching asymptote earlier in practice. Participants who received RT feedback responded faster than those who did not, $F(1, 91) = 5.63$, $MSE = 15{,}195{,}157$, $p < .05$. No other significant main effects or interactions were obtained.

Strategy Use

Mean proportions of trials with reported retrieval strategy use are shown in Figure 1. Younger adults reported using the retrieval strategy more than older adults, $F(1, 91) = 35.73$, $MSE = 11{,}184$, $p < .01$. Retrieval use increased with practice, multivariate $F(28, 64) = 23.74$, $p < .01$. Younger adults shifted to the retrieval strategy faster than older adults during the course of NP training, which led to a significant Block × Age interaction, multivariate $F(28, 64) = 2.45$, $p < .01$. This interaction was associated with reliable quadratic, $F(1, 91) = 10.10$, $MSE = 1{,}129$, $p < .01$, and cubic, $F(1, 91) = 19.83$, $MSE = 626$, $p < .01$, trends. Younger adults' retrieval use curve increased more quickly and approached asymptote earlier in practice. Those who received RT feedback reported the use of retrieval more frequently than those who did not receive feedback, $F(1, 91) = 5.60$, $MSE = 11{,}184$, $p < .05$. No additional main effects or interactions were found. As in our earlier work, RT improvements were more a function of strategy shifts than a result of increases in processing efficiency for each type of strategy. Finding the delayed shift to the retrieval strategy by older adults set the stage for evaluating whether this difference was associated with age differences in latency-monitoring accuracy.

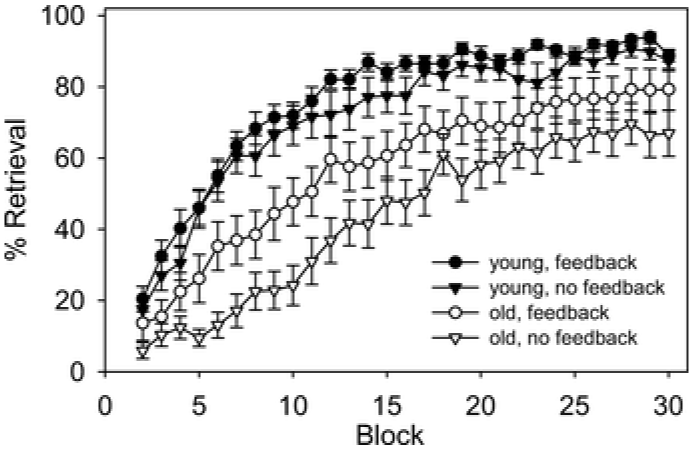
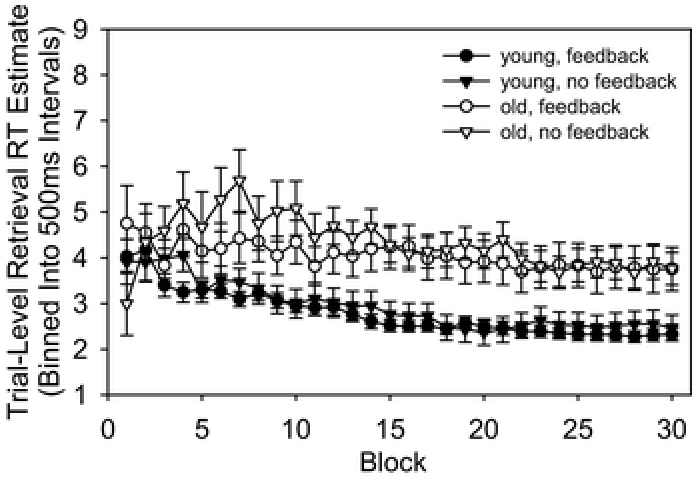


Figure 1. Percentage of retrieval strategy use over blocks, separated by age and feedback condition for Experiment 1. Error bars = 1 standard error of measurement.

Trial-Level Estimated RTs
We evaluated the average trial-level RT estimates from the 9-point ordinal RT scale by block for each feedback condition. We first show the behavior of the RT estimates alone and then consider the absolute accuracy of the RT estimates.

Estimated RT
Figure 2 reports the estimated RTs, aggregated into blocks for each reported strategy, retrieval and scanning. Because retrieval strategy use increased dramatically with practice, there was greater precision early in practice for RT estimates for scanning and greater precision late in practice for RT estimates for the retrieval strategy (note in particular the wider standard error bars and greater mean variability late in practice for scanning). Indeed, 25 young adult participants and 3 older adult participants reported 100% retrieval use at least once between Blocks 25 and 30 and hence contributed few or no data to the scanning RT estimates late in practice. For this reason, we used mixed model analysis (with SAS PROC MIXED; Littell, Milliken, Stroup, & Wolfinger, 1996) to include all available data when evaluating changes with practice.

The mixed model analysis detected main effects of age (younger adults estimated shorter RTs), $F(1, 3823) = 77.63$, $p < .01$; feedback (participants who did not receive trial-level RT feedback estimated shorter RTs), $F(1, 3823) = 6.13$, $p < .05$; and strategy (participants estimated shorter RTs for retrieval responses), $F(1, 3823) = 245.55$, $p < .01$, on trial-level RT estimates. Younger adults' estimates were more differentiated between strategies, which led to a significant Age × Strategy interaction, $F(1, 3823) = 37.55$, $p < .01$. Younger adults' estimated RTs were also more influenced by feedback condition, which led to a significant Age × Feedback interaction, $F(1, 3823) = 15.32$, $p < .01$. Feedback led to slightly shorter estimates of retrieval RT but much longer estimates of scanning RT (a significant Feedback × Strategy interaction), $F(1, 3823) = 50.87$, $p < .01$. Practice-related decreases in trial-level estimates were larger for the retrieval strategy than for the scanning strategy, particularly for younger adults, which led to significant Block × Strategy, $F(1, 3838) = 10.13$, $p < .01$, and Block × Age × Strategy, $F(1, 3838) = 11.80$, $p < .01$, interactions. The general impression from these results is that younger adults' estimates showed greater differentiation between strategies. Feedback had a slightly greater effect for younger adults than for older adults, but only for scanning RT estimates.



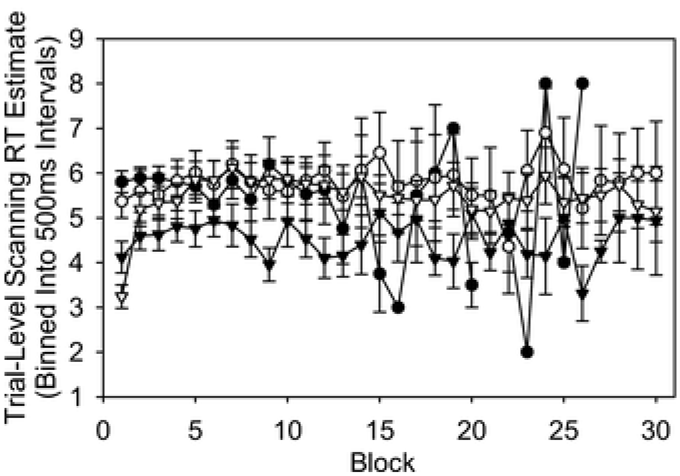


Figure 2. Mean trial-level retrieval (top) and scanning (bottom) noun-pair response time (RT) estimates over blocks, separated by age and feedback condition for Experiment 1. Error bars = 1 standard error of measurement.

Absolute accuracy of RT estimates

We next evaluated absolute accuracy of the RT estimates for the two strategies. Absolute accuracy was scaled by the number of 500-ms bins that estimated RTs deviated from actual RTs. If the actual RT was within the estimate range, deviation was zero. If the actual RT was within the next highest estimate range, the deviation was −1, indicating underestimation. If the actual RT was within the next lowest estimate range, the deviation was 1, indicating overestimation. Figure 3 shows these data. For the faster retrieval strategy, younger adults converged rapidly to accurate estimation, with older adults showing a large underestimation early in practice that also decreased rapidly over blocks. The largest differences were found for the scanning strategy. Older adults substantially underestimated the time taken to scan the NP lookup table, an age difference that was not eliminated by practice or feedback.

The mixed model analysis of trial-level RT estimates revealed differences by age (younger adults' estimates were more accurate), $F(1, 3820) = 50.58$, $p < .01$; feedback (estimates were generally more accurate by participants given trial-level feedback), $F(1, 3820) = 91.56$, $p < .01$; and strategy (retrieval estimates were more accurate), $F(1, 3820) = 164.71$, $p < .01$. The accuracy of younger adults' estimated RTs improved more with feedback, which led to a significant Age × Feedback interaction, $F(1, 3820) = 33.88$, $p < .01$. Feedback improved latency-monitoring accuracy, particularly for the scanning strategy and early in training, as reflected in the significant Block × Feedback × Strategy interaction, $F(1, 3839) = 17.56$, $p < .01$. All two-way interactions associated with these factors were also reliable.

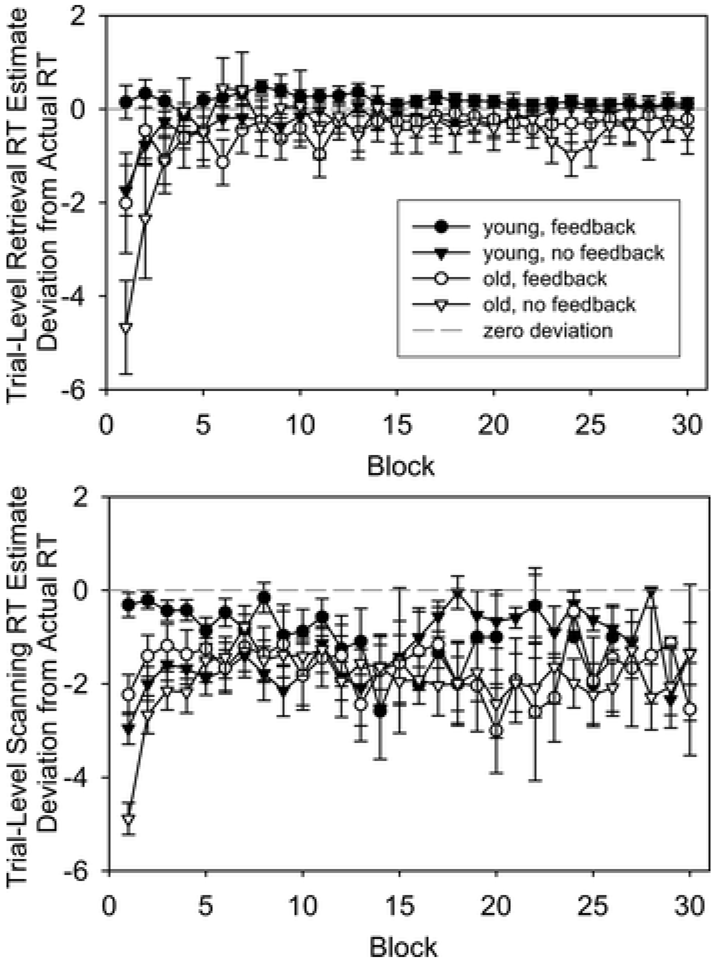


Figure 3. Deviation of mean trial-level retrieval (top) and scanning (bottom) noun-pair task response time (RT) estimates from actual trial-level RT over blocks, separated by age and feedback condition for Experiment 1. Error bars = 1 standard error of measurement.

In sum, RT estimates were more accurate for retrieval trials than scanning trials throughout practice. The effects of feedback were most notable for scanning trials early in practice. Nevertheless, older adults underestimated RTs for both strategies, even with feedback, and were noticeably worse in estimating scanning strategy latencies.

Block-Level Estimated RTs

Figure 4 graphs the block-level RT estimates across blocks for each feedback condition, separated by strategy. Note that feedback appeared to have a bigger and more persistent effect on younger adults' block-level estimates, compared to their trial-level estimates. Feedback raised scanning RT estimates and lowered retrieval RT estimates for younger adults. Feedback effects emerged later in practice for older adults and were consistent in raising RT estimates relative to estimates given without feedback.

The mixed model analysis found main effects of age (younger adults estimated shorter RTs), $F(1, 5239) = 184.71$, $p < .01$; feedback condition (participants not provided with RT feedback estimated shorter RTs), $F(1, 5239) = 37.40$, $p < .01$; and strategy (participants estimated shorter RTs for retrieval responses), $F(1, 5239) = 246.92$, $p < .01$. Younger adults' estimated RTs better differentiated between the two strategies, which led to a significant Age × Strategy interaction, $F(1, 5239) = 33.22$, $p < .01$. Feedback effects were larger for scanning estimates, particularly for younger adults, which led to significant Condition × Strategy, $F(1, 5239) = 45.24$, $p < .01$, and Age × Condition × Strategy interactions, $F(1, 5239) = 4.60$, $p < .05$. Practice-related changes in block-level estimates occurred earlier for the retrieval strategy than for the scanning strategy, particularly for younger adults, reflected in significant Block × Strategy, $F(1, 5248) = 8.8$, $p < .01$, and Block × Age × Strategy interactions, $F(1, 5248) = 21.58$, $p < .01$. Low early retrieval estimates by participants in the no-feedback condition led to less overall practice-related change for these data, reflected in a Block × Condition × Strategy interaction, $F(1, 5248) = 16.21$, $p < .01$. When we compared Figures 2 and 4, it appeared that block-level estimates changed more across blocks than did trial-level estimates.

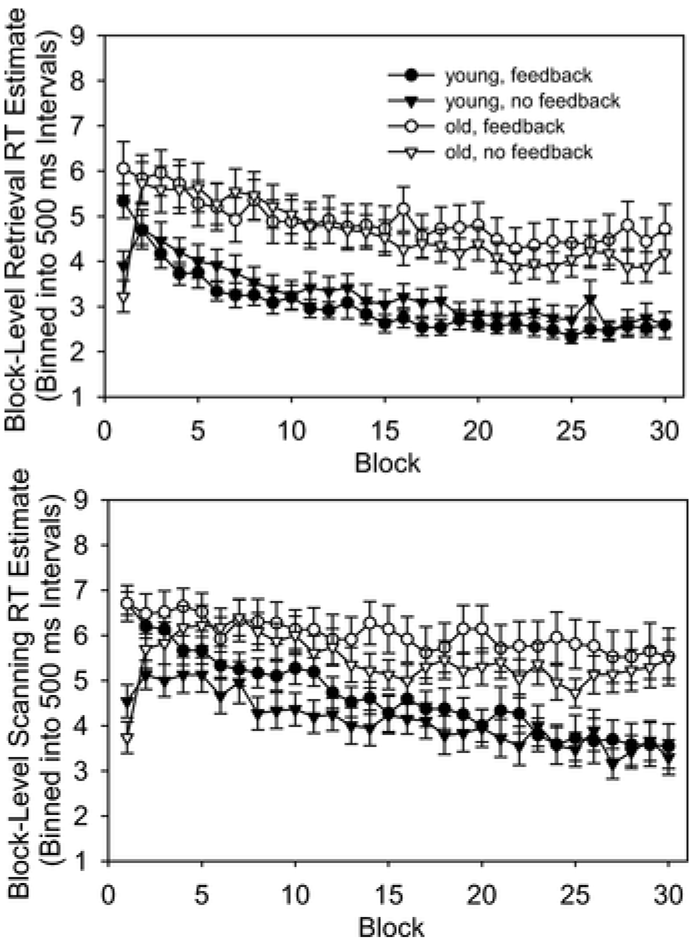


Figure 4. Mean block-level retrieval (top) and scanning (bottom) noun-pair response time (RT) estimates over blocks, separated by age and feedback condition for Experiment 1. Error bars = 1 standard error of measurement.

Absolute Accuracy of Block-Level RT Estimates

To determine whether the patterns in Figure 4 translated into age differences in accuracy of the block-level RT estimates, we analyzed block-level deviations between actual and estimated RTs (see Figure 5). Note that the decreasing trend in block-level scanning estimates in Figure 4 translated into decreases in accuracy that were actually greater for younger adults and for persons given trial-level feedback.

It is surprising that younger adults' estimates tended to be less accurate overall, $F(1, 4804) = 3.54$, $p = .06$. The effect of feedback also ran counter to expectations, with RT estimates being more accurate for participants who were not provided feedback, $F(1, 4804) = 36.89$, $p < .01$. Retrieval estimates were more accurate than scanning estimates, $F(1, 4804) = 322.48$, $p < .01$. Older adults' estimation accuracy was more influenced by feedback, leading to a significant Age × Condition interaction, $F(1, 4804) = 6.46$, $p < .01$. Retrieval RTs were overestimated slightly more by older adults, but scanning RTs were overestimated more by younger adults, leading to a significant Age × Strategy interaction, $F(1, 4804) = 17.03$, $p < .01$. Feedback led to greater overestimation of block-level RT, particularly for the scanning strategy, leading to a significant Condition × Strategy interaction, $F(1, 4804) = 14.92$, $p < .01$. Practice-related changes in estimation accuracy differed by strategy, with relatively stable and accurate retrieval estimates after moderate training but increasingly overestimated scanning estimates, which yielded a Block × Strategy interaction, $F(1, 4839) = 249.18$, $p < .01$.

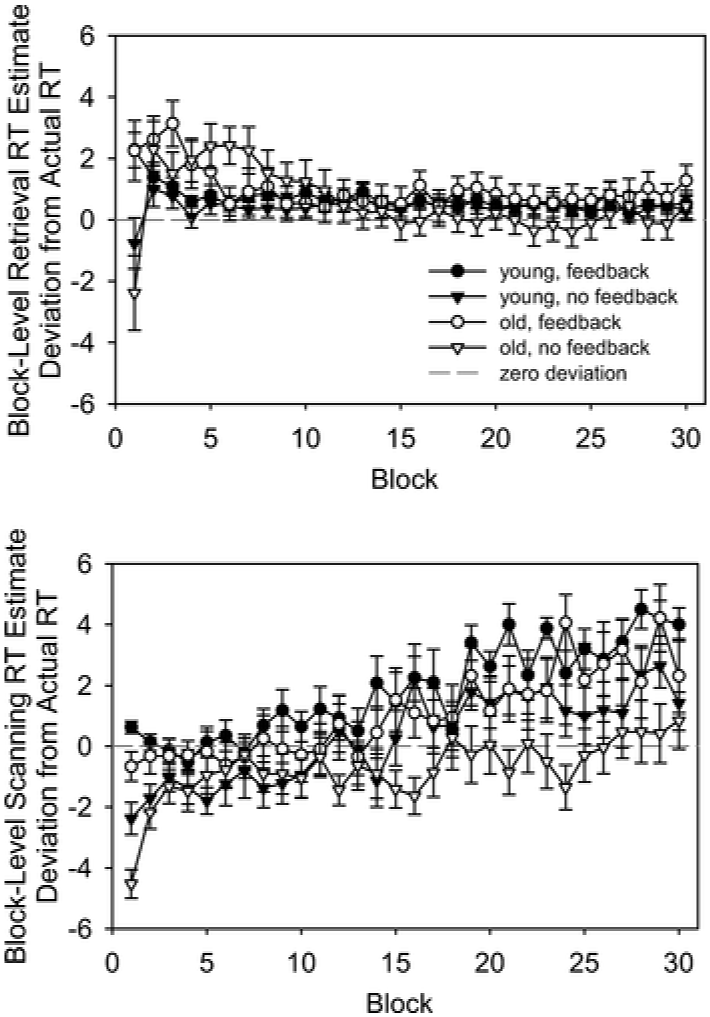


Figure 5. Deviation of mean block-level retrieval (top) and scanning (bottom) noun-pair response time (RT) estimates from actual block-level RT over blocks, separated by age and feedback condition for Experiment 1. Error bars = 1 standard error of measurement.

The differences between trial-level and block-level RT estimates merit further evaluation. In particular, distinct trends can be seen in estimation accuracy for the scanning strategy. Whereas trial-level scanning RTs were generally underestimated, estimates of block-level scanning RTs overestimated actual RT to an increasing degree with task practice, particularly for young adults and participants who received trial-level RT feedback. This disparity suggests a fundamental difference in the mechanisms involved in estimating trial- versus block-level RTs. It is possible that the process of comparison with aggregate subjective retrieval RT led to overestimation of scanning RT, with perceptions of fast retrieval inflating scanning estimates. Indeed, young adults who received trial-level RT feedback estimated the shortest aggregate retrieval RTs and also overestimated scanning RT to the greatest degree.

Relative Accuracy of RT Estimates

An important issue concerning RT estimates is whether individuals perceive accurate ordinal separation of RTs, even if their absolute estimates of temporal duration are distorted. These data are particularly important in light of arguments that absolute accuracy of RT estimates could be worse for older adults solely as a function of general slowing affecting both RT and an internal clock used to monitor RT. Ordinal correlations of RTs with RT estimates should not be affected by the mean baseline RT of each group (see Nelson, 1984). We therefore computed Goodman–Kruskal gamma correlations between each individual's actual and subjective trial-level RTs. To get more stable estimates of these correlations, we pooled data into three 10-block bins (early, middle, and late practice).

Table 2 reports the gamma correlations as a function of age and feedback group, first for all RTs and then separately by strategy type. For the aggregate gamma correlations, participants who received feedback had higher relative accuracy throughout training than those who did not receive feedback, $F(1, 76) = 9.70$, $MSE = 0.16$, $p < .01$. Although no overall age difference in relative accuracy was noted, relative accuracy improved with training for older adults but decreased for younger adults, leading to a significant Block × Age interaction, multivariate $F(2, 75) = 7.68$, $p < .01$. Older adults had reliably lower resolution for the first block, $F(1, 92) = 18.56$, $p < .01$. No other significant main effects or interactions were obtained.

Table 2

Gamma Correlations Between Actual and Estimated Response Time by Age, Feedback Condition, and Response Strategy for All Experiments

	$\text{Young}_{\text{feedback}}$		$\text{Young}_{\text{no feedback}}$		$\text{Old}_{\text{feedback}}$		$\text{Old}_{\text{no feedback}}$	
Block	γ	*SE*	γ	*SE*	γ	*SE*	γ	*SE*
Experiment 1: Overall								
Early	.75	.03	.46	.08	.47	.07	.19	.08
Middle	.66	.09	.49	.08	.62	.06	.50	.08
Late	.62	.12	.39	.14	.71	.06	.58	.07
Experiment 1: Retrieval strategy								
Early	.55	.06	.41	.07	.40	.10	.08	.12
Middle	.60	.09	.50	.07	.57	.07	.36	.09
Late	.60	.12	.29	.16	.67	.06	.39	.09
Experiment 1: Scanning strategy								
Early	.67	.04	.27	.07	.38	.08	.22	.09
Middle	.49	.25	.37	.18	.50	.16	.51	.10
Late	1.00	0.00	.64	.18	.44	.30	.48	.19
Experiment 2								
Early	.62	.04	.52	.04	.57	.05	.53	.06
Middle	.62	.04	.58	.05	.66	.04	.61	.05
Late	.58	.06	.57	.06	.69	.05	.70	.04
Experiment 3								
Overall	.63	.05	.71	.05	.50	.09	.48	.11

When we separated gamma by strategies, distinctive patterns emerged. Note, however, that these results should be evaluated cautiously, as participants with an early strategy shift contributed little to the scanning strategy data late in practice and participants with a late strategy shift contributed little to the retrieval strategy data early in practice. Retrieval strategy gammas appeared largely similar to the aggregate gammas. Resolution differed by feedback condition, $F(1, 70) = 10.88$, $MSE = 0.26$, $p < .01$. The main effect of block was significant, multivariate $F(2, 69) = 3.20$, $p < .05$, driven by a significant linear trend, $F(1, 70) = 4.16$, $p < .05$, with resolution increasing over blocks. There appeared to be lower initial resolution by older adults, which then increased with practice, contrasted against stable or decreasing resolution by younger adults. However, the Block × Age interaction was nonsignificant ($p = .15$). In contrast, scanning strategy gammas suggested larger increases in resolution by younger adults than by older adults, but this might have been, in part, a function of limited data and unstable scanning gammas for young adults late in practice (when the gamma correlation was perfect). No main effects or interactions were significant for the scanning strategy gammas (all *p*s > .15).

Although the patterns of change differed between response strategies, the level of relative accuracy was generally similar. Most critical, however, participants who did not receive latency feedback were generally more deficient in monitoring retrieval latency than scanning latency. This outcome stands in contrast to the absolute accuracy effect, which showed deficient monitoring of scanning latency but not retrieval latency.

Posttask Questionnaire

Questionnaire measures (see Table 3) were subjected to separate (univariate) 2 × 2 (Age × Feedback) analyses of variance. Older adults were reliably less confident in their use of the memory retrieval strategy after practice,

$F(1, 99) = 34.38$, $MSE = 15{,}482$, $p < .001$. They reported that memorizing the pairings required greater effort, $F(1, 98) = 10.03$, $MSE = 11$, $p < .01$, though they also perceived greater performance improvements from memory use, $F(1, 99) = 9.35$, $MSE = 10$, $p < .01$. Older adults also had a lower level of global confidence in their use of retrieval during the NP task, $F(1, 99) = 32.71$, $MSE = 25{,}835$, $p < .01$. In addition, they had lower general confidence in their ability to remember the noun pairings on the posttask cued-recall test, $F(1, 99) = 47.55$, $MSE = 26{,}177$, $p < .001$, and gave lower JOLs prior to cued recall, $F(1, 99) = 9.42$, $MSE = 62$, $p < .01$. These ratings were consistent with age differences in performance, given that younger adults recalled more NPs, $F(1, 99) = 34.69$, $MSE = 32{,}967$, $p < .01$. Gamma correlations comparing JOLs and recall performance did not demonstrate age differences in relative metamemory accuracy ($M_{old} = 0.66$, $SE_{old} = 0.12$; $M_{young} = 0.63$, $SE_{young} = 0.10$), $F(1, 63) = 0.03$, $p = .87$.

Table 3
Means and Standard Errors of Posttest Measures for Experiment 1

Measure	$Young_{feedback}$ *M*	$Young_{feedback}$ *SE*	$Young_{no\ feedback}$ *M*	$Young_{no\ feedback}$ *SE*	$Old_{feedback}$ *M*	$Old_{feedback}$ *SE*	$Old_{no\ feedback}$ *M*	$Old_{no\ feedback}$ *SE*
General recall confidence	79.2	3.65	79.6	3.60	63.2	5.15	46.9	4.22
General memory estimate	81.6	3.73	85.9	2.74	54.4	6.71	49.6	4.62
Average JOL	84.6	4.13	84.7	3.87	51.8	7.44	54.1	6.11
Recall memory	71.0	5.09	79.0	4.61	38.3	7.45	41.4	6.53
Global confidence rating	1.8	0.23	1.9	0.21	2.9	0.31	3.9	0.91
Memorization effort	2.8	0.22	2.7	0.21	3.0	0.24	3.8	0.18
Perceived improvement	1.2	0.07	1.4	0.13	1.8	0.26	2.0	0.28

Note. JOL = judgment of learning.

Providing RT feedback affected older adults' memory confidence. Reliable Age × Feedback interactions for global retrieval confidence, $F(1, 99) = 3.99$, $MSE = 1{,}795$, $p < .05$, and perceived effort of memorization questions, $F(1, 98) = 4.07$, $MSE = 5$, $p < .05$, reflected higher confidence in older adults given RT duration feedback (the two younger feedback groups did not differ). Older adults who received feedback also reported greater general confidence in their recall ability, although this effect did not reach significance, $F(1, 99) = 3.52$, $MSE = 1{,}584$, $p = .07$.

Multilevel Models of Retrieval Use

Trial-level RT estimates

The purpose of this analysis was to determine whether latency monitoring influenced the rate at which individuals shifted to the retrieval strategy. We aggregated trial-level RTs and RT estimates for each strategy and each block of trials, computing the difference between the two strategies in actual RT and in estimated RT for each block. These data were entered into a multilevel regression model (see Singer, 1998). The Level 1 (block-level) equation modeled the increases in retrieval strategy use across blocks, using a log transform to linearize the block function. The model predicted the intercept (centered at Block 8)[2] and the slope of the retrieval use function from objective and subjective RT differences between the retrieval and scanning strategies for each block. Level 2 (person-level) covariates included age and feedback conditions.

Our initial results with age, feedback, and block as predictors replicated the prior linear model and demonstrated significant random effects (individual differences) in rates of change over blocks, likelihood-ratio $\chi^2(2) = 43.4$, $p < .001$. When random effects for blocks were estimated, the Age × Block effect on retrieval shift was not reliable. We then estimated a model that included (a) the actual RT difference between scanning and retrieval and (b) the estimated RT difference between these strategies as predictors of retrieval use. Table 4 (top half) reports the estimated coefficients from the model. Actual and estimated RT differences both predicted retrieval use at Block 8 ($p < .05$). The significance of estimated RT strategy effects in the presence of actual RT strategy effects demonstrated that perceived temporal duration had a unique effect, controlling on actual RT. When we removed actual RT differences, the effect of the estimated RT difference was 1.22 ($SE = 0.43$), $t(938) = 2.82$, $p < .01$. Thus, the effect was modest in magnitude—retrieval use was increased by just over 1% per unit of 500-

ms perceived difference between the strategies. Controlling on actual and estimated RT differences between the strategies reduced the estimated age difference of 24% retrieval use in the initial model to 20% use, indicating that some age-related variance in retrieval shift was associated with age differences in perceived RT difference between the strategies. Subsequent models found a trend for an Estimated RT × Block interaction ($p > .05$) and no Age × Estimated RT difference interaction effect ($p > .25$). Thus, it seems to be the case that estimated RT differences between strategies affected retrieval shift in both age groups, with the age differences in degree of estimation accuracy accounting for the age-related effects of estimated RT on retrieval strategy shift.

Table 4
Multilevel Regression Analysis Predicting Retrieval Use From Trial-Level Response Time (RT) Estimates and Block-Level RT Estimates

Predictor	*df*	Estimate	*SE*	*t*
	Trial-level estimate			
Intercept	91	84.4	8.3	10.11**
Age	91	−19.9	3.8	−5.29**
Feedback	91	−7.8	3.8	−2.08*
Block	937	26.4	1.3	20.45**
Estimated RT difference	937	0.91	0.45	2.04*
Actual RT difference	937	0.69	0.29	2.35*
	Block-level estimates			
Intercept	91	97.4	9.6	10.18**
Age	91	−24.6	4.4	−5.64**
Feedback	91	−10.5	4.4	−2.41*
Block	1181	28.7	0.7	38.76**
Estimated RT difference	1181	1.60	0.40	3.98**
Actual RT difference	1181	0.69	0.27	2.60**

Note. Block was log transformed.
* $p < .05$. ** $p < .01$.

Block-level RT estimates

We repeated the multilevel analysis, this time using the aggregate RT estimates provided at the end of each block to compute the estimated RT difference for scanning and retrieval strategies. One issue addressed in the model is that, given that the block-level estimates were provided at the end of a block, it was more meaningful to use them to predict retrieval use in the following block. We did so using the lag function in SAS. Table 4 (bottom half) reports the model solution. The results closely parallel the results with trial-level RT estimates. Interactions of age and estimated RT with blocks were not reliable. Estimated RT differences between scanning and retrieval strategies predicted retrieval use on the next block independent of the actual RT difference.[3]

Discussion

The results from Experiment 1 indicate that separate processes govern elementary latency monitoring (as revealed by trial-level estimates) and aggregate latency monitoring (as revealed by block-level estimates for each strategy) and that these separate processes are differentially affected by age. In general, older adults were less accurate than young adults in estimating latency for individual trials but more accurate than young adults in providing aggregate latency estimates for separate response strategies following blocks of trials.

Regarding trial-level estimation, older adults showed less differentiation of RT estimates for the two strategies, had lower absolute accuracy, and had initially lower relative accuracy than younger adults. Both age groups were more accurate in estimating latency on trials in which they had employed the retrieval strategy. Both groups underestimated trial-level scanning latencies, with older adults doing so by a much greater amount. When we put these data together, it appears that older adults are prone to underestimate the benefits of the retrieval strategy relative to the scanning strategy.

Providing RT feedback on alternate trials improved the absolute accuracy of trial-level RT estimates in the first few blocks of practice, but there was little differentiation between feedback and nonfeedback conditions after the first few blocks. However, feedback dramatically improved the relative accuracy of RT estimates for both age groups, and these differences persisted across practice.

The multilevel regression results indicate that the estimated RT differences between the scanning and retrieval strategies predicted the rate of shift to the retrieval strategy. Independent of actual RT, larger perceived RT benefit for retrieval predicted the probability of retrieval use. This analysis establishes the functional benefit of latency monitoring in promoting the shift to the more efficient retrieval strategy. The effect was modest in magnitude, but it could be the case that these results underestimate the full effects of latency monitoring on retrieval shift in the standard NP task. All participants received feedback about overall RT level (ignoring strategy) at the end of each block, which might have calibrated their trial-level RT estimates. Furthermore, we used a conservative measure of RT monitoring, estimating latency in 500-ms bins, which might have limited the predictive strength of estimated RT differences for retrieval use. The measure we used was sensitive enough to detect relationships, but additional research with alternative measures of estimating latency is needed to determine whether our measure constrained the relationship of latency monitoring to strategy choice in the NP task.

Providing RT feedback had little effect on younger adults' retrieval use or posttask confidence in the memory retrieval strategy. In contrast, it increased older adults' retrieval strategy use and their confidence in the memory strategy. Given that individuals were randomly assigned to feedback conditions, this outcome suggests that better latency-monitoring accuracy early in practice increased retrieval strategy use by older adults. It also suggests that posttask confidence in the retrieval strategy was, in part, a function of awareness of its benefits on RT, as induced by feedback. This is precisely what would be expected if older adults' poor time monitoring impeded formation of an accurate mental model for the task, in which the retrieval strategy was regarded as the optimal means for achieving efficient performance.

Block-level latency estimates produced an intriguingly distinct pattern of results, as compared to trial-level estimation. In this case, individuals were explicitly asked to estimate the average RT for each strategy, scanning and retrieval. Over practice, individuals increasingly overestimated the aggregate latency of scanning trials, an effect that was larger for younger adults than older adults, but were highly accurate in estimates for retrieval trials. This pattern of latency monitoring for scanning trials is not consistent with conclusions from the time-monitoring literature, which suggest that in mixed conditions (when processes of different temporal durations are monitored), individuals tend to underestimate the duration of longer, more complex processes, especially under concurrent task load (Craik & Hay, 1999). Estimation errors take a qualitatively different and more exaggerated form for the block-level RT estimates, with overestimation of the longer scanning latencies and with the accuracy of scanning latency estimates actually degrading with practice. One possible explanation is that when they were making the aggregate judgments, individuals did not sample instances of subjective RT latencies from memory but rather relied on general inferences about the two different strategies, which were subject to contextual bias. Such inferences could distort estimates of how long scanning actually takes, and such distortions may have greater impact for younger than for older adults, resulting in the reversed pattern of age differences in accuracy of block-level RT estimates. We consider this issue further in the General Discussion. Note that, despite these differences in block-level and item-level latency-monitoring accuracy, the multilevel regression model showed that differences in block-level estimates for the two strategies did predict degree of retrieval use. Individuals who perceived greater RT discrepancy between the two strategies were more likely to use retrieval.

An interesting question is whether the finding of online underestimation of scanning RT versus retrospective overestimation of aggregate scanning RT (particularly when estimates of aggregate retrieval RT were short) is specific to the comparison of RTs for heterogeneous strategies differing in average completion times. We examined this issue in Experiments 2 and 3 by separately evaluating accuracy of RT monitoring of a single strategy. Age differences in monitoring scanning latencies were evaluated in Experiment 2, whereas age differences in memory retrieval latency monitoring were evaluated in Experiment 3.

Experiment 2

Experiment 2 used a varied-mapping condition, in which the noun pairings were randomly shuffled from trial to trial. With varied mapping, there are no consistent pairings to learn (Ackerman & Woltz, 1994), and RT estimates must be based on monitoring the duration of visual search. We still expected age differences in time-monitoring accuracy, but we predicted that the substantial trial-level underestimation of scanning RTs would be attenuated relative to Experiment 1. Furthermore, we expected that block-level scanning RT would no longer be overestimated, especially late in practice, given the absence of subjective retrieval RTs as a relative point of contrast.

Method

Design

The experiment was a 2 (age: young, old) × 2 (RT feedback: given, withheld) × 30 (presentation block) mixed design, with age and feedback as between-subjects factors and block as the within-subject factor.

Participants

Recruitment methods and compensation were consistent with Experiment 1. Participants were tested in groups of up to 3: 42 younger adults between 18 and 25 years of age, and 42 older adults between 60 and 75 years of age. Participant characteristics are provided in Table 1.

Materials and Procedure

Experiment 2 utilized the same 24 unrelated, concrete NPs that were used in the first experiment. The NP task was identical to that of Experiment 1 with four exceptions. First, participants completed only 30 blocks of 24 NP trials, given the small RT improvements after Block 30 in Experiment 1. Second, the lookup table used in Experiment 2 consisted of 12 varied mapping NPs rather than consistently mapped pairs—that is, pairings in the lookup table were randomly assigned on each trial. Third, strategy reports were eliminated because scanning was the only strategy possible in this version of the task. Fourth, participants were not given instructions to maintain a specific level of response accuracy for the matching trials, though response accuracy feedback was provided at the end of each block.

Results

Response Accuracy

Older adults had a higher percentage of correct responses (M = 95.8%, SE = 0.14) than younger adults (M = 92.2%, SE = 0.27), $F(1, 80) = 11.87$, MSE = 682, $p < .01$. Those who received feedback responded less accurately (M = 92.6%, SE = 0.25) than those who did not (M = 95.4%, SE = 0.17), $F(1, 80) = 7.23$, MSE = 682, $p < .01$. The main effect of block was not significant, nor were any of the interactions.

Actual RTs

Figure 6 reports the actual visual scanning RTs for each group. Younger adults responded faster than did older adults, $F(1, 80) = 71.35$, MSE = 9,904,408, $p < .01$. Those who received feedback tended to respond faster than those who did not, although this difference was not statistically reliable, $F(1, 80) = 3.23$, MSE = 9,904,408, $p = .08$. Note that the pattern was similar to that in Experiment 1, with individuals who received feedback tending to have shorter RTs. RTs declined with practice, multivariate $F(29, 52) = 5.65$, $p < .01$. The trend analysis revealed reliable linear, quadratic, cubic, and quartic trends ($p < .01$), with the linear trend having the largest effect, $F(1, 80) = 58.06$, MSE = 814,350, $p < .01$. RTs became more efficient with practice, but the improvement was less than in the CM condition of Experiment 1, in which strategy shift was possible. No other significant main effects or interactions were obtained.

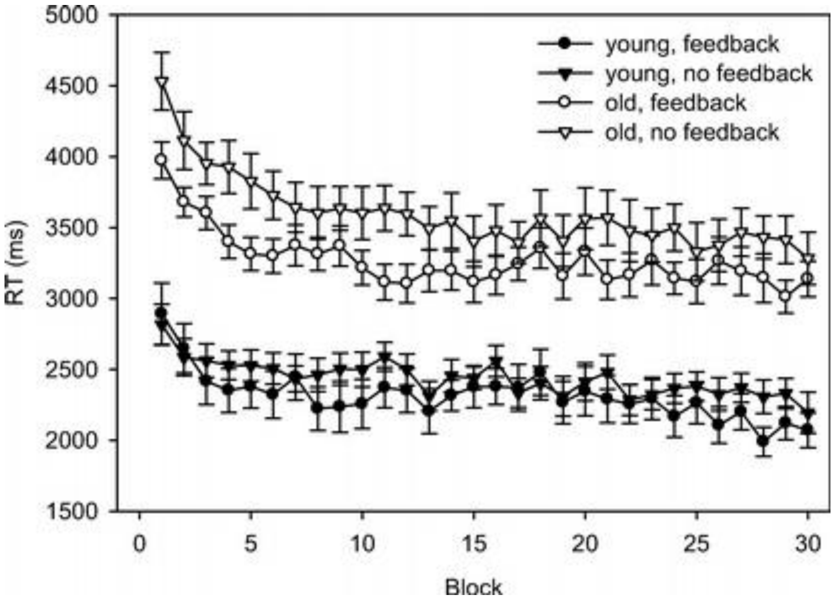


Figure 6. Mean trial-level noun-pair task response time (RT) over blocks, separated by age and feedback condition for Experiment 2. Error bars = 1 standard error of measurement.

Absolute Accuracy of Trial-Level RT Estimates

Absolute accuracy of the RT estimates was assessed as an estimate's deviation from its corresponding RT. RT estimates did not change reliably over blocks, so one might expect changes in absolute accuracy over time. This was indeed the case (see Figure 7, upper panel). Younger adults made more accurate estimates than older adults, $F(1, 78) = 8.80$, $MSE = 57$, $p < .01$. Those who received feedback made more accurate estimates than those who did not receive feedback, $F(1, 78) = 46.10$, $MSE = 57$, $p < .01$. These main effects were qualified by an Age × Feedback interaction, $F(1, 78) = 11.92$, $MSE = 57$, $p < .01$. Feedback essentially eliminated age differences in estimation accuracy, but without feedback the age differences in accuracy were substantial. In fact, older adults' latency monitoring was surprisingly bad. Without feedback, older adults underestimated scanning latencies by roughly 2 s. Deviations from actual RT decreased with practice, reflecting increased absolute accuracy, multivariate $F(29, 50) = 3.95$, $p < .01$. The trend analysis revealed reliable linear, quadratic, cubic, and quartic trends ($p < .01$), with the largest effect associated with the quadratic curvature, $F(1, 78) = 35.84$, $MSE = 1.38$, $p < .01$. No other significant main effects or interactions were obtained. Note that the pattern of improvement was qualitatively similar for all four groups in Figure 7. However, the two groups receiving feedback were highly accurate in their average trial-level RT estimate after about 10 blocks, but the underestimation of scanning times persisted for both groups not given feedback.

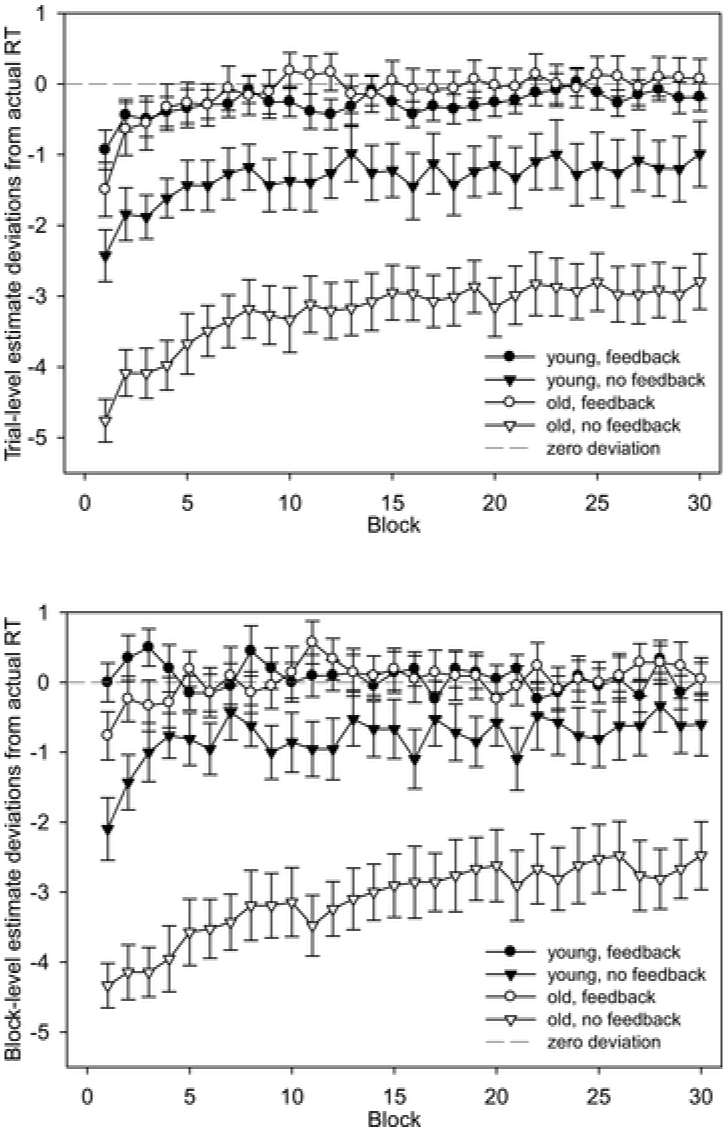


Figure 7. Deviation of mean trial- and block-level noun-pair task response time (RT) estimates from actual trial-level (top) and block-level (bottom) RT over blocks, separated by age and feedback condition for Experiment 2. Error bars = 1 standard error of measurement.

Absolute Accuracy of Block-Level RT Estimates

The absolute accuracy of block-level estimates was also qualitatively similar to trial-level estimates' accuracy (see Figure 7, lower panel). In particular, note the difference in the pattern for scanning estimates in Experiment 2 relative to Experiment 1 (see Figure 4). Older adults' estimates were more deviant than those of younger adults, $F(1, 77) = 14.19$, $MSE = 61$, $p < .01$, and estimates by participants who received feedback were more accurate, $F(1, 77) = 39.25$, $MSE = 61$, $p < .01$. An Age × Feedback interaction, $F(1, 77) = 13.16$, $MSE = 61$, $p < .01$, indicated that older adults who did not receive feedback were highly inaccurate relative to the other three groups. Again, feedback essentially eliminated age differences in the accuracy of scanning latency estimates. Unlike Experiment 1, block-level RT estimates became more accurate with practice, multivariate $F(29, 49) = 1.94$, $p = .02$. The trend analysis revealed reliable linear, quadratic, and cubic trends ($p < .01$), with the largest effect associated with linear change, $F(1, 77) = 13.28$, $MSE = 5.25$, $p < .01$. No other significant effects were obtained.

Relative Accuracy of RT Estimates

Gamma correlations were again used to examine the relative accuracy of item-level RT estimates (see Table 2). Relative accuracy improved more with training for older adults than for younger adults, leading to a significant Block × Age interaction, multivariate $F(2, 79) = 4.37$, $p < .01$. The main effect of training was also significant,

$F(2, 160) = 5.72$, $MSE = 0.1301643$, $p < .01$. Both the main effect and the interaction were significant only for the linear trend, $F(1, 80) = 8.48$, $MSE = 0.03$, $p < .01$, and $F(1, 80) = 7.64$, $MSE = 0.03$, $p < .01$, respectively. No other significant main effects or interactions were obtained.

Note that these patterns of change differed from those shown for scanning in Experiment 1. Older adults' relative accuracy was higher and improved more with training in Experiment 2. The mixture of different strategies in Experiment 1 might have adversely affected older adults' RT monitoring resolution.

Discussion

Perhaps the most surprising aspect of Experiment 2 is the substantial time-monitoring deficit in older adults who did not receive trial-level RT feedback. For these individuals, both trial-level and block-level RT estimates underestimated actual RTs by about 2 s on average. It appears that there were major age differences in absolute accuracy of scanning latency monitoring that were obscured by the context of shorter retrieval latencies in Experiment 1. The fact that feedback effects eliminated age differences in absolute accuracy of both types of latency estimates suggests that older adults, like younger adults, were able to adjust their estimates on the basis of accurate information about actual latency but were unable to do so when they relied only on internal time-monitoring processes. Without feedback, older adults' inaccuracy persisted despite substantial practice on the visual search task. Such findings rule out age differences in scaling behavior (i.e., ineffective use of the latency rating scale) as an explanation for older adults' poor latency-monitoring accuracy when they were not given feedback.

In addition, Experiment 2 produced another surprise. Age differences in relative accuracy of the RT estimates essentially disappeared. The discrepancy between absolute and relative accuracy is not necessarily surprising. Other metacognitive research has shown that absolute and relative accuracy behave differently and that age differences in absolute accuracy can be produced in the same data that generate age equivalence in relative accuracy (e.g., Connor et al., 1997; Hertzog et al., 2002). What is surprising is that older adults' poorer resolution (especially without RT feedback) early in Experiment 1 appears to have been a function of monitoring-strategy-heterogeneous RTs. Relative accuracy in Experiment 1 was influenced by two independent sources of variance: (a) discriminations of latency for the two different strategies, and (b) discriminations of short and long latencies for a given strategy. Given this fact, the results suggest that older adults' diminished relative accuracy in Experiment 1 reflects lowered sensitivity to (and influence of) the type of strategy used on the latency estimates. This effect parallels the literature on monitoring memory accuracy; the strategy used for encoding often has less impact on JOLs than observable stimulus features (Lovelace, 1990).

The data support the major hypotheses of Experiment 2. First, older adults showed a reliable time-monitoring deficit, manifested as less accurate RT estimates, as well as a differential effect of latency feedback on older adults' estimation accuracy. Second, the increasing overestimation of block-level scanning latencies across blocks found in Experiment 1 disappeared in Experiment 2. Instead, the absolute accuracy of block-level RT estimates was largely consistent with the trial-level estimates. Thus, the contrast involved in monitoring scanning with retrieval latencies apparently created the overestimation of block-level scanning estimates found late in practice in Experiment 1.

Experiment 3

Experiment 2 indicated that the contrast of retrieval and scanning strategies influenced the accuracy of latency monitoring for visual scanning in Experiment 1. Experiment 3 evaluated the accuracy of latency monitoring in an associative recognition memory task without concurrent visual search. We hypothesized that there would be reliable age differences in estimated retrieval RTs. We chose the associative recognition task to be similar to the type of associative learning that occurs in the CM NP task but restricted the task to a single study–test trial. Hence, only trial-level RT estimates were collected during the test phase of the task.

Method

Participants

Recruitment methods and compensation were consistent with previous experiments. Participants were tested in groups of up to 3 persons. The sample consisted of 46 younger adults between 18 and 25 years of age and 40 older adults between 60 and 75 years of age. Participant characteristics are provided in Table 1.

Design

The experiment was a 2 (age: young, old) × 2 (RT feedback: given, withheld) design, with age and feedback as between-subjects variables.

Materials and Procedure

The associative recognition task was programmed in Visual Basic 6.0. Stimuli were presented in a 15-point Arial font on a 15-in. LCD monitor with a resolution of 1,024 × 768 pixels. Participants were seated at a height and distance that optimized their screen viewing and comfort.

The stimulus set contained 120 semantically unrelated, concrete nouns. During the study phase, 60 word pairs were presented individually and in a random order. Stimuli were displayed for 4 s for young adults and 6 s for older adults. Immediately after studying each pair, participants made a JOL, estimating the probability (on a continuous scale from 0% to 100%) that they would remember the pair just studied 10 min later. A 250-ms blank screen separated each JOL screen from the next word pair. In the associative recognition test, participants were shown 60 word pairs in a random order. Half of the test word pairs were identical to those studied, whereas the other half were mismatched pairs, which we constructed by combining the same first word originally studied with a second word from another pair. Participants responded by pressing keys labeled *Y* and *N* (i.e., “Yes, the target pair matches a pair that was presented during study” or “No, it does not”).

Participants in both conditions provided RT estimates for retrieval times on half the test trials. Following each odd-numbered trial, participants in the RT feedback condition were provided with their actual RT (in seconds) for the preceding trial. Participants in the no-feedback condition were shown a blank time interval of the same duration. After each even-numbered trial, all participants estimated their RT for the preceding trial immediately afterward. They did so by choosing one of nine options, which were divided into half-second intervals beginning with “less than .05 second” and ending with “greater than 4.0 seconds.” Following feedback, estimates, or a blank interval, participants made a confidence judgment regarding the accuracy of their preceding yes–no response on a scale from 0% to 100% confidence. To evaluate changes in latency monitoring during the test phase, we arbitrarily divided the order of tested items into six blocks of 10 trials, with block treated as the within-subject factor.

Results

Associative Recognition Memory

For these analyses, group differences were assessed in 2 (age: young, old) × 2 (feedback: given, withheld) GLMs on each dependent variable. Associative recognition accuracy did not differ between younger (M = 83.1%, SE = 1.37) and older adults (M = 82.8%, SE = 1.41) in this experiment, $F(1, 84) = 0.27$, MSE = 110.93482, $p = .61$. Likewise, feedback had no effect on accuracy, $F(1, 84) = 0.06$, MSE = 23.35789, $p = .81$.

JOLs and JOL Resolution

There were no age (young: M = 50.34, SE = 2.64; old: M = 48.00, SE = 2.57) or feedback group differences in mean JOLs, $F(1, 84) = 0.51$, MSE = 608.0332, $p = .48$, and $F(1, 84) = 0.05$, MSE = 61.3299, $p = .82$, respectively. Likewise, gamma correlations between JOLs and item-level recognition accuracy revealed no age (young: marginal M = 0.33, SE = 0.05; old: marginal M = 0.28, SE = 0.05), $F(1, 84) = 0.48$, $p = .49$, or feedback differences, $F(1, 84) < .01$, $p = .96$ (see Table 2). JOLs were equally diagnostic of recall for all groups.

Recognition Confidence Judgments (CJs)

CJs did not differ reliably with age, $F(1, 84) = 2.42$, $p > .10$ (young: $M = 76.80$, $SE = 3.07$; old: $M = 68.80$, $SE = 3.64$), or feedback condition ($F < 1$). Gamma correlations of CJs with recognition memory accuracy revealed a tendency for an age difference in the resolution, although it was not significant, $F(1, 82) = 3.33$, $p > .05$. Younger adults ($M = 0.70$, $SE = 0.04$) tended toward better resolution than older adults ($M = 0.58$, $SE = 0.05$). No main effect was found for feedback condition ($F < 1$).

Associative Recognition Latencies

Figure 8 shows the associative recognition RTs over blocks of 10 trials. Younger adults performed faster than did older adults, $F(1, 83) = 50.25$, $MSE = 49{,}489{,}699.51$, $p < .01$. RTs also improved reliably with practice, $F(5, 415) = 28.27$, $MSE = 20{,}875{,}301.37$, $p < .01$, and a statistically significant Age × Block interaction was found, such that older adults' RTs improved more than those of younger adults, $F(5, 415) = 5.27$, $MSE = 1{,}182{,}548.28$, $p < .01$.

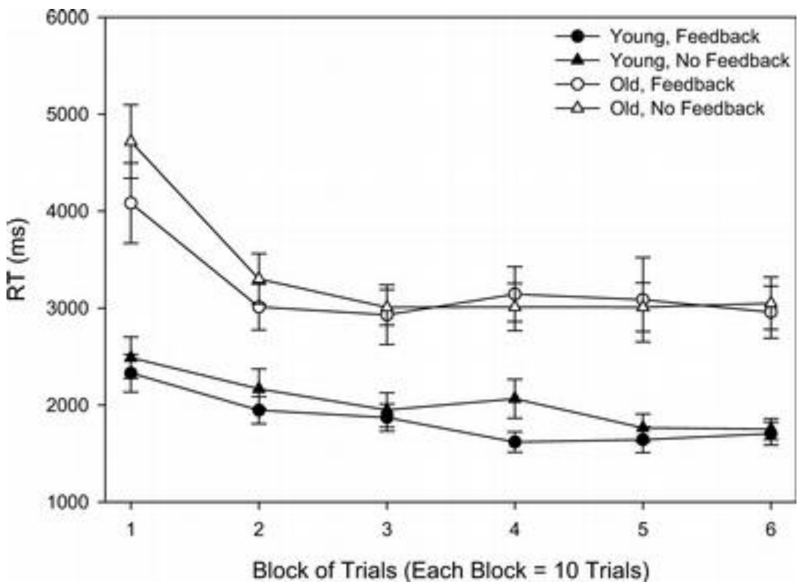


Figure 8. Mean associative recognition response time (RT) over test blocks, separated by age and feedback condition for Experiment 3. Error bars = 1 standard error of measurement.

Accuracy of RT Estimates

Figure 9 plots the absolute accuracy of the latency estimates for associative recognition. Because estimates were provided on every other trial, blocks represent an aggregation over five RT estimates, compared to their five corresponding actual RTs. As can be seen, there were robust age differences in the accuracy of latency monitoring, $F(1, 82) = 37.99$, $p < .01$. A reliable main effect of feedback was found as well, $F(1, 82) = 26.69$, $p < .01$. Absolute accuracy improved reliably with practice, $F(5, 410) = 19.96$, $p < .01$, with older adults' accuracy improving more than that of younger adults, $F(5, 410) = 3.50$, $p < .01$.

We also evaluated relative accuracy of the RT estimates with actual RTs (see Table 2). In contrast to equivalent JOL resolution in predicting recognition performance, there were reliable age differences in the resolution of RT estimates (young: $M = 0.67$, $SE = 0.05$; old: $M = 0.49$, $SE = 0.05$), $F(1, 81) = 5.40$, $p < .05$. Resolution did not vary as a function of feedback condition.

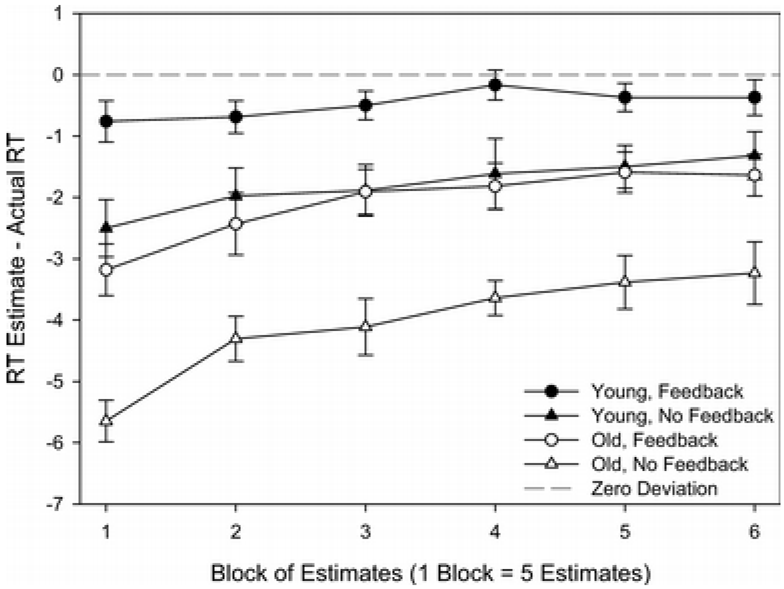


Figure 9. Deviation of mean associative recognition response time (RT) estimates from actual RT over test blocks, separated by age and feedback condition for Experiment 3. Error bars = 1 standard error of measurement.

Discussion

Experiment 3 demonstrates two important points. First, there were substantial age differences in the accuracy of monitoring associative recognition RTs. Older adults who had not been given feedback systematically underestimated RTs by over 2 s. The improvements in their accuracy were determined by improvements in RT, not by changes in RT estimates over blocks. Combined with the findings of Experiment 2, these results indicate that older adults in Experiment 1 produced more accurate estimates of scanning and retrieval latencies early in practice because the mixture of strategies produced greater variability in RT. Older adults' good accuracy in RT monitoring in Experiment 1 does not imply that they accurately monitored retrieval task latencies in general.

Second, we found substantial latency-monitoring deficits in the same older adults who showed no difference from younger adults in the accuracy of their JOLs for forecasting associative recognition. Although there was a trend for older adults to have lower relative accuracy of their CJs after recognition, the effect was smaller than the age difference in RT estimate accuracy. These results suggest, then, that aging affects the monitoring of memory task RT to a greater degree than the monitoring of encoding and recognition processes that correlate with memory performance accuracy.

General Discussion

The present experiments demonstrate a time-monitoring deficit in older adults, consistent with earlier literature (Block et al., 1998; Craik & Hay, 1999). In particular, the results support Craik and Hay's (1999) contention that older adults are prone to underestimate response latencies in cognitively demanding tasks. It does not seem to be the case that age differences in temporal duration are proportional to baseline RT, with the degree of underestimation being strictly a function of how long it takes younger and older adults to respond. In a comparison of the results from Experiments 2 and 3, the degree of underestimation was comparable and perhaps greater in the associative recognition task than in the visual scanning task, despite somewhat shorter RTs in the recognition memory task. In the no-feedback groups, baseline RTs for young adults in the scanning and retrieval conditions differed by almost 700 ms (2,654 ms vs. 1,968 ms), but their binned RT estimates underestimated actual RT by similar amounts (−2.0 vs. −1.90). Likewise, for older adults, actual scanning and retrieval RTs differed by a full second (4,198 ms vs. 3,148 ms), but the absolute accuracies of the binned RT estimates were almost identical (−4.31 vs. −4.12). Such findings weigh against the sufficiency of the hypothesis that age-related slowing in internal clock rates produces slower subjective time and, hence, shorter time estimates by older adults (Block et al., 1998; Salthouse et al., 1979; Schroots & Birren, 1990). Instead, older adults' deficits in temporal duration estimation also might have been influenced by their degree of absorption in the ongoing task and the process heterogeneity of RTs in the monitoring task (see Craik & Hay, 1999). We must acknowledge, however, that the current study was not designed to test the proportional clock-slowing hypothesis and that a careful examination of it would benefit from more precision in the RT estimates than is available with our binned ordinal rating scale.

In Experiment 1, older adults' trial-level latency monitoring was less accurate than younger adults' in terms of both relative and absolute accuracy. The feedback manipulation differentially benefited older adults' latency estimate accuracy (given on alternating trials to those who requested latency estimates), demonstrating that older adults could make more accurate latency estimates if they were provided with valid information about their speed of response. Yet, even with feedback, it was still the case that older adults dramatically underestimated the RTs for trials in which they used the scanning strategy. Experiments 2 and 3 demonstrated even larger age deficits in the absolute accuracy of monitoring visual scanning or memory retrieval as homogeneous processes within the task environment. Thus, it appears that older adults benefited from the contrast between retrieval and scanning latencies in Experiment 1, improving their absolute accuracy in estimating scanning response latencies, even though they did not fully perceive the relative efficiency of the retrieval strategy.

It is notable that the process of trial-level RT estimation seemed to differ from the block-level RT estimation process. This difference was revealed in Experiment 1, in which latencies were estimated for two response strategies differing in average latency. Scanning latencies were overestimated by both younger and older adults, especially late in practice, and the effect was more pronounced for younger adults. This pattern was qualitatively different than that observed for trial-level RT estimates. Experiment 2 showed that this pattern was eliminated when people estimated latencies of only the visual scanning process.

How can this effect by explained? The pattern of results rules out the hypothesis that aggregate block-level estimates are made by exhaustive sampling of actual response latencies for the preceding block from memory. Instead, individuals appeared to make inferences about aggregate latencies on the basis of a more limited and perhaps biased weighting of latencies in the preceding block. One possibility is that making aggregate estimates involves a representativeness heuristic, in which latencies of a few trials are recalled and used to make the aggregate judgments. In Experiment 1, such an effect might have mixed with a source recollection problem—individuals might not have been able to accurately recall whether the item latency they sampled was associated with a scanning strategy or a retrieval strategy. However, given that older adults have known deficits in source memory processes, which are required for recollection of strategy-specific detail (i.e., Henkel, Johnson, & De Leonardis, 1998; Spencer & Raz, 1995), a source memory account per se predicts that older adults would show greater distortion of block-level RT estimates than younger adults, and the opposite pattern was observed in Experiment 1. Older adults' block-level estimates were actually more accurate.

Source memory problems could contribute to the process in the following manner. Assume that individuals achieve accurate monitoring of their overall level of RT but that they do not explicitly form quantitative estimates of the differences in latency between strategies until asked to do so by the block-level RT estimates. Then poor source recollection by both young and old adults would force people, in general, to rely on inferences about differential latency for the two strategies instead of sampling RTs from memory and classifying them accurately according to original strategy. Dunlosky and Hertzog (2000) showed that both younger and older adults radically underestimated the memory benefits of a superior encoding strategy after task performance, possibly because they could not always explicitly recall the strategy they had used at the time of recall. By this account, aggregate judgments are anchored to accurate monitoring of overall RT level but are adjusted for scanning and retrieval latencies on the basis of an implicit theory about how much faster retrieval is than scanning. In the NP task, retrieval became the dominant strategy late in practice. Perhaps for this reason, the accurate estimate of overall latency was used as the anchor for estimating retrieval latencies, which were then also highly accurate. The greater distortion of scanning latencies for younger adults late in practice should then reflect the consequence of their experience-based implicit theory that scanning is much slower than retrieval. Both age groups discounted the RT for scanning, knowing that it was a slower strategy, which led to systematic distortion (overestimation) of scanning latencies that increased with task practice. However, young adults showed a greater degradation in absolute accuracy for block-level scanning latencies that was, by this account, a consequence of their more accurate implicit theory of the lower efficiency of the scanning strategy.

Alternatively, the observed pattern could be influenced by the fact that younger adults were more likely than older adults to retrieve rather than scan in the NP task. Given this fact, younger adults might have had more limited evidence in each block about actual scanning latencies, which would limit the quality of their aggregate estimates to a greater degree than was the case for older adults, especially late in practice, when scanning became rare. Such a mechanism could act in concert with the anchoring plus inference account detailed above, providing another mechanism by which younger adults were more vulnerable to distortion in the block-level RT estimates.

The major new contribution of the present research is that older adults' time-monitoring deficit had functional significance for skill acquisition. Some of older adults' well-documented retrieval shift reluctance (e.g., Touron & Hertzog, 2004a, 2004b) can be attributed to a failure to accurately monitor the relative efficiency of the retrieval strategy. Failing to perceive the RT benefits of the retrieval strategy predicts older adults' delayed retrieval shift. Use of the retrieval strategy was related to the perceived block-level difference in RTs between strategies, such that those who perceived longer RTs for the scanning strategy were more likely to use the retrieval strategy. However, the multilevel regression results show that other factors not analyzed in this study accounted for most of the age differences in retrieval shift. Older adults may be generically less willing to shift strategies, showing a kind of behavioral inertia that is difficult to overcome (e.g., Spieler, Mayr, & LaGrone, 2006), even when they are given explicit instructions to shift strategies (Touron et al., in press). Stable suboptimal performance by young adults has been observed in other tasks and may occur for multiple reasons, including simple habit strength (e.g., Fu & Gray, 2004). We have also shown, in these data and in other studies (Touron & Hertzog, 2004a, 2004b), that older adults profess less confidence in their ability to successfully use memory retrieval and that this poor memory self-concept correlates with retrieval strategy use (Hertzog & Touron, 2006). Nevertheless, the present data implicate deficient RT monitoring as one contributing factor to older adults' delayed retrieval shift.

Hertzog and Touron (2006) recently developed a task mental model questionnaire, which they gave at the end of NP task practice—in particular, one that asked people to rate benefits of the two strategies for short RT and accuracy early and late in practice. A measure of rated efficiency benefits of the retrieval strategy correlated with its use in both younger and older adults. Older adults perceived much less benefit of retrieval for efficient responding, however. The present results suggest that deficient time monitoring in the NP task may contribute to age differences in such mental models, which, in turn, predict age differences in strategy shift. We are currently planning a study to directly evaluate this hypothesized role of mental models as the mediator of time monitoring and retrieval use connections.

We cannot rule out the hypothesis that the relationships between latency monitoring and retrieval strategy use observed in Experiment 1 could be accounted for by some form of motivated (top-down) distortion in RT estimates. It is possible that these relationships reflect the fact that older adults who lack confidence in their memory are motivated to perceive less difference in the efficiency of the two strategies. Mather and colleagues (e.g., Mather, Knight, & McCaffrey, 2005) reported that post hoc choice attribution heuristics, in which reconstructive processes fill memory gaps about past choice options with information consistent with individuals' beliefs, are more prevalent in older adults. The present findings could be influenced by an a priori bias to perceive higher costs and lower benefits of a nonpreferred strategy. Although we consider the hypothesis that inaccurate latency monitoring affects the perception of efficiency benefits for the two strategies to be more plausible, the distortion hypothesis can only be addressed by new research involving (a) explicit measurement of memory beliefs and an NP mental task model before the start of actual NP task performance, combined with (b) the measurement of latency monitoring during the NP task itself.

The present results suggest a potential dissociation of age differences in the accuracy of monitoring memory products and monitoring memory dynamics. Older adults were equally accurate in their JOLs after extended NP task practice (Experiment 1) and in JOLs immediately after study in Experiment 3. The latter are based more on monitoring encoding, whereas the former are essentially cue-only delayed judgments known to be influenced by monitoring retrieval outcomes. Consistent with other literature (e.g., Connor et al., 1997), we observed equally accurate delayed JOLs in this associative recognition task. In contrast, our participants were less accurate in monitoring retrieval latencies, especially in the absence of RT feedback. However, it is premature to argue for a pure qualitative dissociation between monitoring memory accuracy and memory dynamics. We detected a trend for older adults to have lower relative accuracy of recognition memory CJs. Kelley and Sahakyan (2003) also found reduced accuracy of CJs in one of two experiments, and Souchay and colleagues (e.g., Souchay et al., 2004) have reported an age deficit in monitoring episodic (but not semantic) feeling-of-knowing accuracy. Other evidence indicates no general age deficit in monitoring episodic retrieval, however. In addition to equally accurate delayed JOLs, both older and younger adults have highly accurate confidence judgments after cued recall (e.g., Dunlosky & Hertzog, 2000), and Touron and Hertzog (2004a) found equal relative accuracy of recognition CJs in the NP task itself. In any case, our findings with monitoring memory retrieval latency argue that metacognitive monitoring is not fully spared by aging (Hertzog & Hultsch, 2000). Instead, some aspects are apparently impaired, and, in the case of latency monitoring, this impairment has functional consequences.

As noted earlier, an age deficit in time monitoring may be implicated in age deficits in laboratory tasks, such as time-based prospective memory (Park et al., 1997). It may also influence self-regulation in concurrent task performance (e.g., Salthouse, Hambrick, Lukas, & Dell, 1996), in which accurate estimation of process duration and response duration may be needed for contention scheduling and conflict monitoring by central executive processes (Shallice & Burgess, 1991). RT monitoring might also substantially influence the execution of a wide variety of real-world tasks, such as using a computer (Fu & Gray, 2004), driving a car, planning a sequence of activities, or performing household chores, that are known to be predicted by speeded cognitive functioning, including attention and the useful field of view (Wood et al., 2005). One might speculate, on the basis of the time-monitoring literature, that age deficits would be minimal for tasks that require timed motor responses and more substantial when conscious assessment (verbal reports) of latency are required for strategic self-regulation. Future research should examine the relationships among latency monitoring, formation of an accurate mental task model, and age differences in strategic task performance in other task environments.

Footnotes

1 We do not claim that all individuals form such a comparative mental model of strategy costs and benefits during the NP task nor that strategy shift requires conscious awareness that retrieval is the more efficient strategy. We do not focus on NP task mental models in this study. However, we argue that for time monitoring to influence strategy shift, individuals must use the information from monitoring RTs as one influence on their strategy choice, and that a mental model about relative efficiency is likely to be the means of doing so.

2 We set the intercept at Block 8 to ensure that the intercept (which associated with young adults in the feedback condition) was below asymptote (i.e., 100% retrieval use).

3 Using same-block retrieval use rather than lagged use reduced the effects of block-level estimates. They were only significant in the presence of actual RT differences; when the latter variable was dropped from the model, the effect of estimated RT differences between strategies was nonsignificant. We considered these outcomes empirical justification for focusing on the lagged regression results we report.

References:

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.

Bäckman, L., & Karlsson, T. (1985). The relation between level of general knowledge and feeling-of-knowing: An adult age study. *Scandinavian Journal of Psychology*, *26*, 249–258.

Block, R. A., Zakay, D., & Hancock, P. A. (1998). Human aging and duration judgments: A meta-analytic review. *Psychology and Aging*, *13*, 584–596.

Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, *24*, 654–663.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, *12*, 50–71.

Craik, F. I. M., & Hay, J. F. (1999). Aging and judgments of duration: Effects of task complexity and method of estimation. *Perception & Psychophysics*, *61*, 549–560.

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about strategy effectiveness: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging*, *15*, 462–474.

Eakin, D. K., & Hertzog, C. (2006). Release from implicit interference in memory and metamemory: Older adults know that they can't let go. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *61*, P340–P347.

Einstein, G. O., McDaniel, M. A., Richardson, S. L., Guynn, M. J., & Cunfer, A. R. (1995). Aging and prospective memory: Examining the influences of self-initiated retrieval processes. *Journal of Experimental Psychology*, *21*, 996–1007.

Fu, W.-T., & Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science*, *28*, 901–935.

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.

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., & Touron, D. R. (2006, April). *Aging and individual differences in algorithm to retrieval shift with task practice*. Paper presented at the Cognitive Aging Conference, Atlanta, GA.

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., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481–537.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute, Inc.

Logie, R. H., Maylor, E. A., & Della Salla, S. (2004). Working memory in event- and time-based prospective memory tasks: Effects of secondary demand and age. *European Journal of Cognitive Psychology*, *16*, 441–456.

Lovelace, E. A. (1990). Aging and metacognition concerning memory function. In E. A.Lovelace (Ed.), *Aging and cognition: Mental processes, self-awareness and interventions* (pp. 157–187). Amsterdam: North-Holland.

Lovelace, E. A., & Marsh, G. (1985). Predictions and evaluation of memory performance by young and old adults. *Journal of Gerontology*, *40*, 192–197.

Martin, M., & Schumann-Hengsteler, R. (2001). How task demands influence time-based prospective memory performance in young and older adults. *International Journal of Behavioral Development*, *25*, 386–391.

Mather, M., Knight, M., & McCafffrey, M. (2005). The allure of the alignable: Younger and older adults' false memories of choice features. *Journal of Experimental Psychology: General*, *134*, 38–51.

McCall, R. B., & Appelbaum, M. I. (1973). Bias in the analysis of repeated-measures designs: Some alternative approaches. *Child Development*, *44*, 401–415.

McCormack, T., Brown, G. D., Maylor, E. A., Richardson, L. B. N., & Darby, R. J. (2002). Effects of aging on absolute identification of duration. *Psychology and Aging*, *17*, 363–378.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *84*, 93–116.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*, 102–116.

Park, D. C., Hertzog, C., Kidder, D. P., Morrell, R. W., & Mayhorn, C. B. (1997). Effect of age on event-based and time-based prospective memory. *Psychology and Aging*, *12*, 314–327.

Robinson, A. E., Hertzog, C., & Dunlosky, J. (2006). Aging, encoding fluency, and metacognitive monitoring. *Aging, Neuropsychology, and Cognition*, *13*, 458–478.

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.

Salthouse, T. A., Hambrick, D. Z., Lukas, K., & Dell, T. D. (1996). Determinants of adult age differences in synthetic work performance. *Journal of Experimental Psychology: Applied*, *4*, 305–329.

Salthouse, T. A., Wright, R., & Ellis, C. L. (1979). Adult age and the rate of an internal clock. *Journal of Gerontology*, *34*, 53–57.

Schroots, J. J. F., & Birren, J. E. (1990). Concepts of time and aging in science. In J. E.Birren & K. W.Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., (pp. 45–64). San Diego, CA: Academic Press.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe deficits in man. *Brain*, *114*, 727–741.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *23*, 323–355.

Souchay, C., Isingrini, M., Clarys, D., Taconnat, L., & Eustache, F. (2004). Executive functioning and judgment-of-learning versus feeling-of-knowing in older adults. *Experimental Aging Research*, *30*, 47–62.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory context: A meta-analysis. *Psychology and Aging*, *10*, 527–539.

Spieler, D. H., Mayr, U., & LaGrone, S. (2006). Outsourcing cognitive control to the environment: Adult age differences in the use of task cues, *Psychonomic Bulletin & Review*, *13*, 787–793.

Touron, D. R. (2006). Are item-level strategy shifts abrupt and collective? Age differences in cognitive skill acquisition. *Psychonomic Bulletin & Review*, *13*, 781–786.

Touron, D. R., & Hertzog, C. (2004a). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging*, *19*, 452–466.

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

Touron, D. R., Swaim, E., & Hertzog, C. (2007). Moderation of older adults' retrieval reluctance through task instructions and monetary incentives. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *62*, 149–155.

Wearden, J. H., Wearden, A. J., & Rabbitt, P. M. A. (1997). Age and IQ effects on stimulus and response timing. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 962–979.

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

Wood, K. M., Edwards, J. D., Clay, O. J., Wadley, V. G., Roenker, D. L., & Ball, K. K. (2005). Sensory and cognitive factors influencing functional ability of older adults. *Gerontology*, *51*, 131–141.

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