

Metacognitive Influences on Study Time Allocation in an Associative Recognition Task : An Analysis of Adult Age Differences

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

Hines, J., [Touron, D. R.](#), & Hertzog, C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, 24(2), 462-475. DOI: 10.1037/a0014417

Made available courtesy of the American Psychological Association: <http://dx.doi.org/10.1037/a0014417>

This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

*****Note: Figures may be missing from this format of the document**

Abstract:

The current study evaluated a metacognitive account of study time allocation, which argues that metacognitive monitoring of recognition test accuracy and latency influences subsequent strategic control and regulation. The authors examined judgments of learning (JOLs), recognition test confidence judgments (CJs), and subjective response time (RT) judgments by younger and older adults in an associative recognition task involving 2 study-test phases, with self-paced study in Phase 2. Multilevel regression analyses assessed the degree to which age and metacognitive variables predicted Phase 2 study time independent of actual test accuracy and RT. Outcomes supported the metacognitive account—JOLs and CJs predicted study time independent of recognition accuracy. For older adults with errant RT judgments, subjective retrieval fluency influenced response confidence as well as (mediated through confidence) subsequent study time allocation. Older adults studied items that had been assigned lower CJs longer, suggesting no age deficit in using memory monitoring to control learning.

Article:

An individual attempting to learn new information in preparation for a test must decide how much effort to spend studying and restudying that information. Self-regulation of study behavior involves selectively choosing information to study and choosing how—and for how long—to study that information (e.g., Winne & Perry, 2000). Research on metacognition and self-regulated learning examines how individuals use monitoring and control mechanisms to optimize learning (e.g., Nelson, 1993, 1996). A general premise is that learners' selection of items for study or allocation of study time is influenced by their current state of learning, goals for future learning, and beliefs about themselves as learners (Cull & Zechmeister, 1994; Dunlosky & Hertzog, 1998; Metcalfe & Kornell, 2005; Thiede & Dunlosky, 1999). For example, individuals may choose to devote more study time to information they have not yet learned (e.g., Nelson & Narens, 1990). Typically, rates of learning are fastest when individuals devote more effort to studying items they have not yet mastered.

This study evaluates whether there are age differences in self-regulation of learning based on metacognitive monitoring. Although previous research indicates that older adults are capable of accurate encoding and retrieval monitoring (see Hertzog & Hulstsch, 2000), the limited evidence available suggests older adults may be less likely to use their accurate monitoring to effectively regulate learning and test performance. In self-paced testing, older adults do not judge readiness to recall as accurately as younger adults (Murphy, Schmitt, Caruso, & Sanders, 1987; Riggs, Lachman, & Wingfield, 1997), in part because they do not spontaneously use self-testing strategies to monitor recall readiness (but see Bottiroli, Dunlosky, Guerini, Cavallini, & Hertzog, 2008). Older adults, like younger adults, are more likely to select items they have not yet learned for restudy when forced to choose a limited subset of items (Dunlosky & Hertzog, 1997). However, given the opportunity to restudy all items, older adults' self-paced study time may not be optimally allocated. Dunlosky and Connor (1997) found larger negative gamma correlations between recall and study time for younger adults than for older adults, suggesting that younger adults devote more time and effort to items they have not yet learned, relative to older adults. The difference might arise because older adults do not use the more optimal control

strategy of differential allocation of study time to unlearned items or because they invest more time on maintenance rehearsal of already learned items, fearful that otherwise these items will be forgotten. In either case, a simultaneous evaluation of metacognitive monitoring and control over study is needed to ensure that any age differences that are observed are consistent with the hypothesis of spared monitoring but impaired control (Hertzog & Hultsch, 2000). Dunlosky and Connor supported the spared monitoring, impaired control hypothesis because they found no age differences in the accuracy of delayed judgments of learning but age differences in the correlations of delayed judgments of learning with subsequent study time allocation.

The present study allows for a more thorough examination of age differences in monitoring mechanisms that are relevant to self-regulated learning because it explicitly separates two classes of monitoring that are potentially relevant: (a) monitoring effectiveness of study and (b) monitoring retrieval at the time of test. To frame the problem in terms of which types of monitoring could be related to study time allocation, we turn next to a discussion of the relevant monitoring mechanisms.

Monitoring Study and Test Performance

A metacognitive perspective argues that monitoring at different stages of learning can be used for self-regulatory control (e.g., Nelson & Narens, 1990). Judgments of learning (JOLs)—confidence in future recall after information has been studied—correlate with later memory performance and with study time allocation in tasks like paired-associate learning (e.g., Nelson, 1993; Dunlosky & Connor, 1997). JOLs made immediately after each item is studied, but prior to any memory test for that item, assess monitoring of encoding. JOLs made with a delay after study, but still prior to any memory test, measure monitoring of accessibility of target information when cued (Dunlosky & Nelson, 1994; Nelson & Dunlosky, 1991; Weaver & Kelemen, 2003). Delayed JOLs are typically more accurate in predicting memory performance, but immediate JOLs may reflect aspects of monitoring encoding that persist over learning trials and that influence study time.

Much of the experimental literature on monitoring and study time allocation has evaluated JOLs (e.g., Koriat, Ma'ayan, & Nussinson, 2006; Metcalfe & Finn, 2008). Ironically, JOLs may not capture the most important aspect of monitoring relevant to study time allocation. Although delayed JOLs predict subsequent study behaviors, including item selection for restudy and study time allocation (e.g., Dunlosky & Connor, 1997; Nelson, Dunlosky, Graf, & Narens, 1994), correlations are typically stronger between study behavior and memory performance itself (Dunlosky & Connor, 1997).

If actual memory performance is the best predictor of study time allocation, then metacognitive monitoring of cognitive states during test, including both target accessibility and accuracy of memory test performance, may be the more important sources of feedback for self-regulatory control, including study time allocation. Indeed, individuals discover which items are well learned versus unlearned principally on the basis of test experience (e.g., Dunlosky & Hertzog, 2000; Finn & Metcalfe, 2007; Hertzog, Dixon, & Hultsch, 1990; Hertzog, Price, & Dunlosky, 2008; Koriat & Bjork, 2006). Confidence judgments (CJs)—rated confidence in the accuracy of responses following memory tests—are proximal indicators of performance monitoring (Dunlosky & Hertzog, 2000; Higham, 2002). One could even argue that the relationship of JOLs to study time could be a causally spurious correlation created by the fact that both JOLs and study time allocation correlate with memory performance monitoring (as measured by CJs) and memory (but see Metcalfe & Finn, 2008).

The potential importance of performance monitoring, as measured by CJs, leaves open the possibility that age differences in study time allocation are due to age-related deficits in the accuracy of performance monitoring rather than the monitoring mechanisms captured by JOLs. Older adults are prone to high-confidence errors in some recognition memory tasks (e.g., Dodson, Bawa, & Krueger, 2007; Kelley & Sahakyan, 2003), and such monitoring failures could contribute to suboptimal study time allocation.

In this study, we applied a metacognitive perspective to study time allocation by evaluating the influence of JOLs and CJs on strategic control in an associative recognition task. In general, a metacognitive perspective posits the following: If a metacognitive variable (e.g., CJs) measures monitoring that is potentially relevant to

control of study, and if that monitoring influences subsequent study behavior, then the metacognitive variable should predict study behaviors even when statistically controlling for the objective performance that was monitored (e.g., recognition accuracy). Despite the relevance of performance monitoring for self-regulated study, the experimental literature has focused exclusively on JOLs, not CJs, in evaluating study time allocation. This could be due, in part, to the fact that CJs are highly correlated with cued-recall accuracy (e.g., Dunlosky & Hertzog, 2000), the memory task typically used in study time allocation studies. The current study used a recognition memory task for paired-associate items. Recognition CJs are not as highly correlated with successful recognition as they are with recall (e.g., Benjamin, 2003; Kelley & Sahakyan, 2003), so we were more likely to discern unique associations of CJs on subsequent study time allocation, independent of memory performance itself.

A Model for Study Time Allocation

Figure 1 contains a model for self-paced study time allocation as examined in this experiment. The first study opportunity (Phase 1) was experimenter paced and followed by an associative recognition memory test. Two aspects of actual memory performance, the accuracy and latency of Phase 1 recognition test responses (see Figure 1 performance variables: Phase 1 recognition accuracy and Phase 1 recognition response time [RT]), were sources of evidence, or cues (Koriat, 1997), that could be monitored to guide subsequent self-paced study time allocation (Phase 2 study time allocation). Figure 1 also shows three metacognitive variables that could guide study time allocation in Phase 2: Phase 1 JOLs, Phase 1 CJs, and Phase 1 RT estimates. As noted above, Phase 1 JOLs capture learning confidence during Phase 1 study, whereas Phase 1 CJs capture performance monitoring during the recognition memory test in Phase 1. Given that a metacognitive account of study time allocation assumes an influence of item monitoring independent from item difficulty (as reflected in memory performance), we hypothesized that Phase 1 JOLs and CJs would predict Phase 2 study time allocation independent of Phase 1 recognition performance. An open question was whether JOLs would predict Phase 2 study time allocation independent of Phase 1 CJs. If study time allocation was driven only by performance monitoring, then JOLs would correlate with study time allocation, but this effect would be eliminated when controlling for CJs.

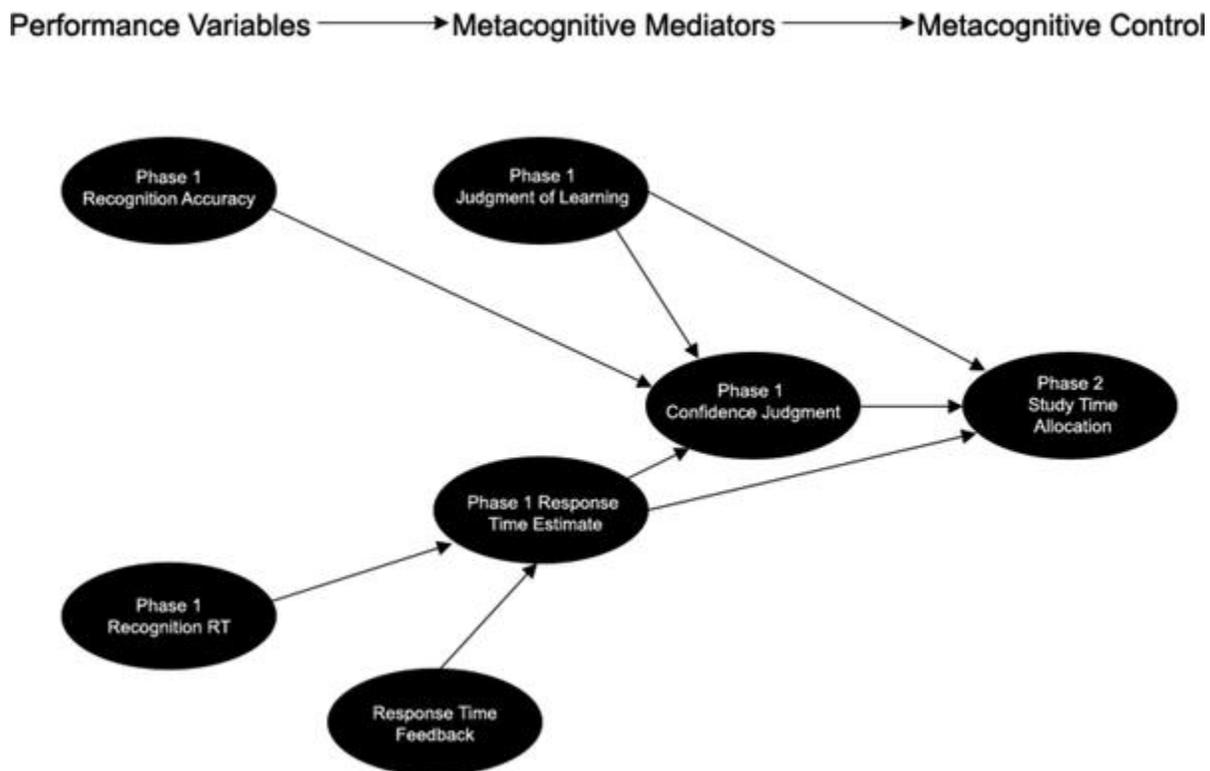


Figure 1. Flow diagram demonstrating precursors of Phase 2 study time. RT = response time.

Whereas CJs reflect monitoring of recognition memory accuracy, memory test RTs can also be diagnostic of the degree to which an item has been learned. Recognition memory RTs reflect accessibility of information from episodic memory. Fast, correct recognition responses are also associated with higher confidence (CJs) about memory accuracy (e.g., Benjamin & Bjork, 2000; Doshier, 1982). Recognition memory RTs decrease as a function of repeated testing and enhanced learning (e.g., Touron & Hertzog, 2004), indicating that retrieval fluency is a valid cue for the probability of subsequent remembering.

Given the relevance of retrieval fluency as a metacognitive cue, we sought to directly measure monitoring of Phase 1 recognition RTs. Monitoring of performance latency, or subjective (i.e., perceived) fluency, reflects how quickly an individual believes he or she evaluated the recognition test item. It can be measured by requesting explicit RT estimates after each recognition trial. If subjective fluency is a metacognitive cue that guides strategic control, then subjective fluency should influence Phase 1 CJs and Phase 2 study time allocation, independent of actual fluency, as measured by actual RTs. The next section further examines subjective fluency as a metacognitive variable.

Subjective Fluency as a Metacognitive Cue

Memory researchers have shown that illusions of processing fluency can be misattributed to memory for prior events, which has been labeled a fluency heuristic (e.g., Kelly & Jacoby, 1998; Whittlesea & Leboe, 2003). Such research has used various experimental manipulations to influence perceptions of fluency and examined effects on perceptions of familiarity but has not explicitly asked individuals to estimate the latency of their responses to measure subjective fluency. There is a large literature on monitoring temporal duration (e.g., Block & Zakay, 1997; Craik & Hay, 1999). However, metacognitive research has typically measured subjective confidence in memory accuracy but not subjective experiences of processing fluency. Instead, measures of actual fluency (i.e., actual rather than perceived RT) have been studied with respect to their influence on metacognitive judgments (e.g., Hertzog, Dunlosky, Robinson, & Kidder, 2003; Koriatic & Ma'ayan, 2005).

Explicit measurement of subjective fluency is important for evaluating metacognition and cognitive control because actual and subjective fluency are separable constructs and behave in empirically distinct ways. Robinson, Hertzog, and Dunlosky (2006) found that estimates of encoding time (a type of subjective fluency) were more correlated with JOLs than actual encoding time (actual fluency). Actual fluency of encoding and retrieval influences memory judgments independent of its influences on memory performance (e.g., Benjamin, Bjork, & Schwartz, 1998; Hertzog et al., 2003; Koriatic & Ma'ayan, 2005), consistent with a metacognitive account of fluency effects. Older adults, like younger adults, show this effect (Robinson et al., 2006). Given that subjective encoding fluency influences JOLs independent of actual encoding RTs, we hypothesized that subjective fluency of recognition test responses would predict both recognition CJs and the allocation of further study time, independent of actual recognition RTs. It may be the case that what is perceived to have occurred (subjective RT) is of greater importance than what actually occurred (i.e., actual RT) in terms of allocating future study time.

Inaccurate perceptions of fluency, if used to assess learning, might adversely impact strategic control. For example, an inaccurate perception of a slow recognition test response may lead a person to study an item longer than warranted on the basis of the actual retrieval latency. Age differences in the accuracy of RT monitoring therefore represent another as-yet-untested potential source of age differences in study time allocation. Older adults are less accurate than younger adults when monitoring latency in a variety of experimental tasks (e.g., Block, Zakay, & Hancock, 1998; Craik & Hay, 1999; Hertzog, Touron, & Hines, 2007). Particularly germane to this study, older adults are less accurate in estimates of retrieval RTs in associative recognition tasks compared with younger adults (Hertzog et al., 2007). However, Hertzog et al. (2007) showed that these deficits can be at least partially ameliorated by explicit RT feedback and training.

We therefore manipulated the accuracy of subjective fluency in this experiment by providing actual RT feedback on odd trials, alternating with trials that requested RT estimates. The advantage of this procedure is that we expected RT feedback to influence subjective fluency but not actual RTs, allowing us to consider the

contribution of RT estimate accuracy for subsequent monitoring and control. It is important to note in our Figure 1 model that feedback directly influences only subjective RT, not downstream variables such as CJs. The hypothesis of age differences in study time regulation can be tested by evaluating interactions of age with JOLs, CJs, and RT estimates on subsequent study time allocation, as we discuss in more detail shortly.

Intact and Rearranged Pairings in Associative Recognition

The associative recognition test we used required participants to discriminate between intact pairs (same pairings as originally studied) and rearranged pairs (words from originally studied items recombined into new pairings). This approach prevents correct responding solely on the basis of recognizing words as old, without recovering the studied associations. Cohn and Moscovitch (2007) showed that strategies for correct rejection of rearranged lures are more complex because one can recall both or only one of the original associations and use that information to reject the rearranged lure. Moreover, age differences in associative recognition appear to be driven largely by a higher likelihood of false positives on rearranged items (Cohn, Emrich, & Moscovitch, 2008; Healy, Light, & Chung, 2005), possibly implicating age differences in retrieval-based inferences at test. For this reason, CJs for rearranged pairs may not reflect performance monitoring for both items involved in the rearrangement. A test of metacognitive control in the associative recognition task is therefore best assessed by examining relations of CJs for intact pairs to subsequent study time for those pairs. More broadly, it was important to evaluate whether the type of recognition test item (intact vs. rearranged) influenced metacognitive monitoring accuracy and subsequent study time allocation.

Summary of Major Hypotheses

The current research examined a metacognitive account of study time allocation in a two-phase associative recognition task, evaluating the influence of cues such as JOLs, CJs, and subjective test fluency. A key feature of the design is that there are multiple, correlated cues that potentially influence subsequent study time allocation. To evaluate the more fundamental mediational hypotheses stated above, the typical approach in the experimental literature—calculation of within-person correlations of a single metacognitive variable (e.g., JOLs) with study time—is inadequate for isolating influences of one cue, controlling for others. Instead, we used multilevel regression analyses (Singer, 1998; Snijders & Bosker, 1999), which allow for the simultaneous evaluation of multiple variables' influence on measures of interest (Hoffman & Rovine, 2007). Our analytic approach is described more fully in the Results section.

We anticipated that these multilevel regressions would reveal that metacognitive judgments (CJs and JOLs) account for independent variance in strategic control above and beyond objective performance measures. We predicted stronger relationships of CJs than of JOLs to study time allocation, given the likely importance of performance monitoring for governing later study behavior. Furthermore, we expected that subjective fluency would impact metacognitive judgments and study allocation beyond any effects of actual fluency. Finally, we examined age differences in predictive relationships to address the possibility that older adults do not effectively use monitoring for the self-regulation of study behavior. In particular, subjective RT effects should be more robust for older adults because they are expected to display a greater discrepancy between actual and estimated recognition test RT—and this is especially true for those who do not receive objective RT feedback to aid in calibrating their RT estimates.

Method

Design

The experiment was a 2 (age: young, old) \times 2 (RT feedback: given, withheld) \times 2 (study–test phase: 1, 2) mixed design, with age and feedback as between-subjects independent variables and study–test phase as the within-subjects variable.

Participants

Adults between the ages of 18 and 25 years were classified as young, whereas our older group consisted of adults between the ages of 60 and 75. Younger adult participants ($n = 39$, $M_{\text{age}} = 21.8$ years) were undergraduates from Appalachian State University who received course credit for their participation. Older

adults ($n = 35$, $M_{\text{age}} = 66.8$ years) were recruited from the nearby community and were paid for their participation. Approximately half of the participants in each age group were tested within each feedback condition. All participants were prescreened for basic health issues that could impede participation and were required to have good corrected visual acuity (20/50 or better). No participants were removed from our analyses as a result of these exclusions.

Materials and Procedure

Informed consent was established prior to beginning the experimental session. Participants then completed a brief personal data questionnaire, the Shipley Institute of Living Scale Revised vocabulary test (Zachary, 1986), and the Digit–Symbol subtest of the Wechsler Adult Intelligence Scale–Revised (Wechsler, 1981). Data pertaining to participant characteristics are presented in Table 1.

Table 1
Means (and Standard Deviations) of Participant Characteristics

Measure	Younger adults		Older adults	
	Feedback	No feedback	Feedback	No feedback
Age (years) ^a	22.2 (5.4)	21.3 (2.5)	66.7 (3.9)	67.0 (3.8)
Number of medications ^a	1.0 (1.3)	1.4 (1.4)	2.4 (1.4)	2.3 (1.8)
Vocabulary ^a	29.5 (3.7)	29.2 (4.0)	35.2 (4.0)	35.2 (2.7)
Digit–Symbol ^{a,b}	65.6 (8.5)	69.1 (11.4)	52.1 (12.0)	46.3 (11.5)
Digit–Symbol memory ^a	7.0 (2.0)	7.7 (1.5)	5.7 (2.5)	5.7 (2.2)

Note. Number of medications = self-reported number of daily medications taken; Vocabulary = number correct out of 40 on the Shipley Institute of Living Scale Revised vocabulary test (Zachary, 1986); Digit–Symbol = Wechsler Adult Intelligence Scale–Revised Digit–Symbol subtest (Wechsler, 1981). Digit–Symbol memory = symbol recall memory following the Wechsler Adult Intelligence Scale–Revised Digit–Symbol subtest.

^a Age main effect, $p < .05$. ^b Age \times Feedback interaction, $p < .05$.

The associative recognition task was programmed in Visual Basic 6.0. Stimuli were presented using a 15-point Arial font on a 15-inch (38.1-cm) LCD monitor with a resolution of 1024×768 . Participants were seated at a height and distance that optimized their screen viewing and comfort. Self-paced instructions preceded each portion of the task.

The computer task consisted of two consecutive study–test phases, which were identical in format and stimuli save for one key difference: Study was experimenter paced for Phase 1 (word pairs were presented for 4 s for younger adults or 6 s for older adults) and self-paced for Phase 2 (participants terminated item-level study by pressing the spacebar). The computer task included the following sequence: (a) experimenter-paced Study 1, (b) Test 1, (c) self-paced Study 2, and (d) Test 2.

During each of the two study phases, 60 word pairs (e.g., *IVY–BIRD*) were presented individually in a randomized order. These word pairs were constructed from 120 semantically unrelated nouns, and all study items remained unchanged between Study Phases 1 and 2. After studying each item, participants gave a JOL rating their ability to remember the item approximately 10 min later on a continuous scale from 0% to 100% confidence. After viewing the final study item in each phase, participants began self-paced testing for that phase, wherein, because of a programming error, 31 (instead of an even 30 or half) of the test stimuli were intact study items (e.g., an item presented as *IVY–BIRD* during study would be presented as *IVY–BIRD* during test) and the other 29 items were rearranged study items (e.g., *IVY–BIRD* and *BARREL–STAR* might be presented at test as *IVY–STAR* and *BARREL–BIRD*). There was no overlap between the 31 matched and 29 unmatched word pairs, and items were selected randomly to be presented either intact or rearranged. The foils were created by merely swapping the second word in each word pair foil with the second word from another foil. Test items were generated randomly prior to the test portion of each phase. Word pair constituents were

always presented in capital letters and always retained their position on the left or right side of the word pair. Prior to the presentation of each test item, a “+” was presented on the screen to orient participants to the location of the upcoming word pair. Upon the presentation of the stimulus, participants were instructed to press one of two keys indicating whether or not that word pair had been studied previously.

Following odd-numbered trials, participants in the RT feedback condition were provided with their actual RT (in seconds) for the preceding trial. Participants in the no-feedback condition received no additional information after each *yes–no* response but were presented with a blank screen for a time interval of the same length as that of the feedback screen (250 ms). Participants were given no rationale for why we asked them to estimate their trial-level recognition RTs; they were merely asked to supply their best guesses. Although this measure may be less intuitive than a JOL or CJ, there was no evidence from our sample that individuals misunderstood what we were asking them to supply to us. Following even-numbered trials, participants estimated their recognition RT by choosing one of nine RT estimate categories, which divided their potential responses into half-second intervals beginning with “less than 0.5 second” and ending with “greater than 4.0 seconds.” Next, participants provided a CJ rating their confidence in the accuracy of the previous recognition trial on a continuous scale from 0% to 100% confidence by entering any value within in that range. After all items were tested, participants were offered a short break prior to the beginning of the next study–test phase.

Results

We examined the following measures: recognition memory accuracy, recognition memory RT, RT estimates, JOLs and CJs, and Phase 2 self-paced study time. First, we present analyses of age differences in basic variables, including Goodman–Kruskal gamma correlations for resolution (relative accuracy) of metacognitive judgments. This approach is common in the metacognitive literature, thereby facilitating comparisons with other studies. In particular, we report aggregated statistics (such as gamma correlations or median RTs) that are absent in multilevel regression models that use item-level data directly. Second, we present the relationships of multiple variables with CJs and study time allocation via multilevel regression models.

General Linear Model (GLM) Analyses

All variables were analyzed in a GLM with age group (young, old) and RT feedback condition (given, withheld) as between-subjects variables and study–test phase (1, 2) and item type (intact, rearranged) as within-subjects variables.

Recognition memory accuracy

Younger and older adults had similar levels of recognition accuracy (see Table 2). Of interest, intact test items produced lower accuracy than did rearranged items, $F(1, 70) = 4.00, p < .05$, but this difference was small and only reliable during Phase 1, $F(1, 70) = 8.17, p < .01$. Both age groups’ recognition accuracy improved at Phase 2, $F(1, 70) = 128.19, p < .001$. There was no age difference in learning over phases, $F(1, 70) = 1.57, p = .22$. Thus, varying encoding time for older and younger adults was sufficient to equate them on recognition memory, a useful feature for testing hypotheses about age differences in monitoring accuracy and study allocation.

Table 2
Mean (and Standard Deviation) Recognition Memory Accuracy by Age, Test Phase, and Word Pair Type

Word pair type	Younger adults		Older adults	
	Phase 1	Phase 2	Phase 1	Phase 2
Intact	0.73 (0.18)	0.88 (0.19)	0.73 (0.17)	0.86 (0.15)
Rearranged	0.79 (0.18)	0.90 (0.15)	0.78 (0.17)	0.85 (0.21)

Recognition memory RT

Analysis of RTs was restricted to correct recognition trials only, using median RTs to reduce the influence of extreme outliers. Younger adults responded faster than older adults to memory test items, $F(1, 70) = 46.87, p < .001$ (see Table 3). Intact items were responded to more quickly than were rearranged items, $F(1, 70) = 82.61, p < .001$, and this difference was greater for older adults, $F(1, 70) = 18.79, p < .001$. The magnitude of the difference was substantial, consistent with the argument that rearranged pairs invoke additional processes, such as recall to reject or recall to accept, that increase processing time (e.g., Cohn & Moscovitch, 2007). Reflecting improved learning, participants responded faster at Phase 2, $F(1, 70) = 96.90, p < .001$, with older adult RTs decreasing more overall, $F(1, 70) = 4.90, p < .05$.

Table 3
Mean (and Standard Deviation) Recognition Test Response Time (in Milliseconds) by Age, Test Phase, and Word Pair Type

Word pair type	Younger adults		Older adults	
	Phase 1	Phase 2	Phase 1	Phase 2
Intact	1,754 (585)	1,206 (351)	2,932 (961)	2,045 (769)
Rearranged	2,015 (718)	1,475 (473)	3,666 (1,599)	2,808 (1,107)

RT estimate accuracy

The absolute accuracy of RT estimates was calculated as the difference between estimated RT and actual RT (binned for analysis into 500-ms intervals to match the scale used for RT estimates; see Table 4). Replicating previous work (Hertzog et al., 2007), older adults underestimated their RTs to a greater degree than did younger adults, $F(1, 70) = 22.36, p < .001$, and those who received RT feedback estimated their RTs more accurately than did those who did not, $F(1, 70) = 20.97, p < .001$ ($M_{\text{feedback}} = -0.85, SD_{\text{feedback}} = 1.55; M_{\text{no feedback}} = -2.39, SD_{\text{no feedback}} = 2.03$). More accurate RT estimates were given for intact items, $F(1, 70) = 30.66, p < .001$, and the difference in RT estimation accuracy between item types was greater for older adults, $F(1, 70) = 7.50, p < .01$. This may be a consequence of faster RTs for intact items; past research has shown that older adults estimate time less accurately as intervals increase (e.g., Craik & Hay, 1999; but see Hertzog et al., 2007). RT monitoring accuracy improved in Phase 2, $F(1, 70) = 84.81, p < .001$, with older adults showing greater improvement, $F(1, 70) = 13.23, p < .01$. This result, however, was qualified by older adults' greater potential for improvement; both groups may have comparably improved had they displayed similar initial monitoring accuracy. Regardless, a clear and substantial age difference in RT monitoring accuracy was found. Although older adults' mean actual RT was approximately twice that of younger adults, their estimated RT displayed but a fraction of that difference.

Table 4
Mean (and Standard Deviation) Absolute Accuracy of Recognition Test Response Time Estimates by Age, Test Phase, Response Time Feedback Condition, and Word Pair Type

Variable	Younger adults		Older adults	
	Phase 1	Phase 2	Phase 1	Phase 2
Response time feedback				
Given	-0.43 (0.93)	-0.05 (1.01)	-2.11 (1.74)	-0.87 (1.13)
Withheld	-1.91 (2.03)	-1.12 (1.76)	-3.91 (1.53)	-2.66 (1.54)
Word pair type				
Intact	-0.95 (1.64)	-0.41 (1.37)	-2.71 (1.78)	-1.29 (1.50)
Rearranged	-1.15 (1.75)	-0.67 (1.56)	-3.27 (2.02)	-2.20 (1.98)

Note. Negative values indicate underestimation of test response times. A mean value of zero would indicate perfect monitoring accuracy or no difference between actual and estimated response latency.

Phase 2 self-paced study time

Older adults took reliably more time to study items, which we consider in detail with the multilevel regression models for this variable. A main effect of item type was also found whereby items that were rearranged at Test 1 were studied longer during Phase 2 compared with previously intact items, $F(1, 67) = 10.95, p < .01$ (see Table 5). Items incorrectly recognized at Phase 1 were also studied longer than those that were correctly recognized, $F(1, 70) = 33.75, p < .001$. However, rearranged items were studied longer only if accurately recognized, $F(1, 67) = 19.01, p < .001$.

Table 5
Mean (and Standard Deviation) Phase 2 Study Time (in Milliseconds) by Age, Phase 1 Recognition Test Accuracy, and Word Pair Type

Word pair type	Younger adults		Older adults	
	Accurate	Inaccurate	Accurate	Inaccurate
Intact	2,044 (1,367)	3,866 (2,570)	4,512 (2,990)	8,096 (5,104)
Rearranged	3,410 (2,162)	3,629 (2,705)	6,809 (4,510)	7,746 (5,964)

Metacognitive judgments

No main effects of age were found for magnitudes of JOLs or CJs (see Table 6). Mean judgments increased during Phase 2—JOLs, $F(1, 70) = 96.50, p < .001$; CJs, $F(1, 70) = 46.36, p < .001$ —consistent with increases in recognition memory accuracy. Younger adults' JOLs increased more between phases than did those of older adults, $F(1, 70) = 4.43, p < .05$. Intact items were given higher recognition confidence ratings overall compared with rearranged items, $F(1, 70) = 30.87, p < .001$, but the difference was not reliable in Phase 2, $F(1, 70) = 6.44, p < .05$.

Table 6
Mean (and Standard Deviation) Metacognitive Memory Measures by Age, Test Phase, and Word Pair Type

Measure and word pair type	Younger adults		Older adults	
	Phase 1	Phase 2	Phase 1	Phase 2
Judgments of learning				
Intact	29.32 (21.89)	63.85 (29.68)	35.00 (23.26)	54.60 (27.61)
Rearranged	31.68 (24.45)	64.15 (31.01)	31.64 (22.19)	54.74 (26.20)
Confidence judgments				
Intact	82.14 (29.87)	91.56 (22.67)	72.36 (29.67)	84.66 (27.96)
Rearranged	73.12 (28.00)	87.51 (26.24)	55.83 (32.40)	78.54 (29.24)

Because of Nelson's (1984) work, Goodman–Kruskal gamma correlations have become the traditional method for assessing resolution or relative accuracy of metacognitive judgments, although this method of analysis is not always ideal.¹ These correlations are reported in Table 7; the degrees of freedom are lower than in other analyses because gammas cannot be calculated for individuals with recognition accuracy performance at floor or ceiling. No age differences in resolution were found for either JOLs, $F(1, 70) = 1.51, p > .25$, or CJs, $F(1, 70) = 1.68, p > .19$. As hypothesized, JOLs correlated more weakly with test performance than did CJs, but resolution in both cases was above chance (reliably greater than 0). Overall JOL resolution increased between phases, $F(1, 70) = 6.72, p < .05$. In addition, the relative accuracies of both JOLs and CJs (when collapsed across phase) were greater for intact items, $F(1, 70) = 25.72, p < .001$, and $F(1, 70) = 13.61, p < .001$, respectively.

To evaluate whether metacognitive judgments correlated with study time allocation, gamma correlations were computed between our metacognitive memory measures at Phase 1 and study time at Phase 2. As expected, correlations were negative, with higher confidence associated with less study time at the next study opportunity. No age differences in the study time gamma correlation were found for either JOLs, $F(1, 72) = 0.61, p > .4$, or CJs, $F(1, 72) = 0.39, p > .50$ (see Figure 2). The correlation of CJs with study time differed by item type, being dramatically lower for rearranged pairs, $F(1, 72) = 62.38, p < .001$. JOL–study time correlations were fairly low overall. The interaction between age and item type was not reliable ($.10 > p > .05$).

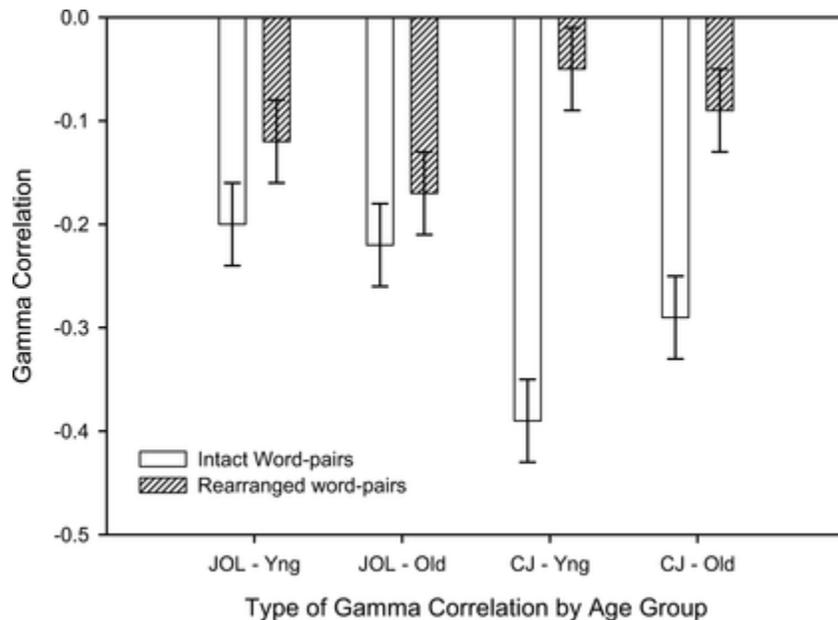


Figure 2. Gamma correlations between Phase 2 study time and metacognitive measures of memory performance collected in Phase 1 by age group and item type. JOL = judgment of learning; Yng = younger adults; Old = older adults; CJ = confidence judgment.

In summary, a common thread among GLM analyses was the distinction between intact and rearranged test trials. Because rearranging items impacted relationships between Phase 1 accuracy and Phase 1 CJs, as well as between these two variables and Phase 2 study time, we concluded that monitoring of rearranged pairs was indeed complicated by the involvement of two items in each rearrangement, possibly because of different recognition strategies. Therefore, the multilevel regression analyses reported below used only test trials for intact word pairs, which have a more definitive and straightforward interpretation as to how monitoring should affect control.

Multilevel Regression Analyses

Multilevel regression models allow for the simultaneous evaluation of multiple metacognitive variables relevant to hypotheses of statistical mediation (Hoffman & Rovine, 2007). The models were used to test metacognitive hypotheses of interest in the present study. For example, the hypothesis that subjective RTs completely mediate effects of actual RTs on CJs predicts that subjective RTs will significantly predict CJs when controlling for actual RT but that actual RTs will not predict CJs when controlling for subjective RTs. Incomplete or partial mediation would be revealed if both variables independently influence CJs, controlling for one another. The advantage of regression approaches is that correlated variables can be evaluated for their independent, direct, and indirect influences on downstream variables.

Multilevel regression also allows for simultaneous consideration of within-person and between-person effects.² For instance, it is possible to evaluate whether within-person (item) variation in CJs is associated with within-person variation in study time allocation (i.e., individuals allocate more study time to items they are not confident they know and less time to items they are confident they know), as reflected in gamma correlations of

these two variables. At the same time, one can also evaluate whether between-person differences in mean CJs are associated with between-person differences in average study time (i.e., individuals with overall lower confidence will allocate more overall study time compared with individuals with overall higher confidence). In principle, these influences can operate in different directions (as is often the case in speed–accuracy trade-offs, where the trade-off is a negative within-person correlation but speed and accuracy are positively correlated between-persons). In such cases, the two types of influence can suppress each other if they are not separated (Snijders & Boskers, 1999). It is important to note that within-person and between-person are standard terminology in multilevel regression and are not necessarily analogous to within-subjects and between-subjects variability components in comparisons such as analyses of variance.

We used multilevel regression models to evaluate predictors of Phase 1 recognition test confidence and Phase 2 study time. The relationships of interest can be seen in Figure 1. Items defined the first level, characterized by the variables Phase 1 recognition accuracy, Phase 1 recognition RT, Phase 1 RT estimate, Phase 1 JOL, Phase 1 CJ, and Phase 2 study time allocation. The second level was defined by participants and included between-subjects variables of feedback condition and age group. SAS PROC MIXED (Littell, Milliken, Stroup, & Wolfinger, 2000) was used to estimate unstandardized regression equations linking the variables.

Our general approach was to include all relevant within-person and between-person variables in a regression analysis and then trim nonsignificant effects to achieve a final regression model.³ We also routinely modeled random effects in intercepts and for test accuracy (and their correlation), reflecting between-subjects differences in the dependent variable; we report reliable random effects only ($p < .05$).⁴ As discussed in the GLM section of the Results, test trials that included intact word pairs were used exclusively in the regressions because of indications that the relationships between variables were less stable for trials that included rearranged word pairs. The final regression results for the two critical equations involving prediction of Phase 1 CJs and prediction of Phase 2 study time are reported below. The reasons for the inclusion of these equations are as follows: (a) Age differences in study time allocation were the driving force behind the current investigation, and (b) Phase 1 CJs were hypothesized to have a proximal and relatively large impact on Phase 2 study time, so it was important to determine which variables relate to recognition test confidence.

CJs
 For CJs, our initial regression equation included the variables of actual RT, estimated RT, recognition accuracy, JOLs, RT feedback, age, and selected interactions (Feedback \times Estimated RT, Age \times JOL, Age \times Actual RT, Age \times Estimated RT, and Age \times Recognition Accuracy). After trimming nonsignificant effects, we found reliable influences of both within-person and between-person effects. Table 8 reports the regression estimates and significance tests for reliable predictors, with the estimates scaled in the metric of the original variables. It is important to note that each effect is estimated while considering the effects of the other variables in the equation. Also note that age and RT feedback interactions, when reliable, are described with group-specific regression coefficients and *t* values within the text.

Table 8
 Multilevel Regression Analysis Predicting Phase 1 Confidence Judgments (Intact Word Pairs)

Effect	Estimate	SE	df	<i>t</i>
Intercept	64.48	5.20	68	12.41***
Age (younger)	9.57	5.64	68	1.70
Feedback condition (given)	0.58	6.11	68	0.10
Phase 1 judgment of learning (between-person)	0.50	0.14	68	3.63***
Phase 1 recognition accuracy (within-person)	20.36	3.04	1,004	6.69***
Phase 1 response time (within-person)	-0.00	0.00	1,004	-6.12***
Phase 1 Response Time (Within-Person) \times Age (Younger)	-0.00	0.00	1,004	-2.36*
Phase 1 response time estimate (within-person)	-3.41	0.82	1,004	-4.16***
Phase 1 Response Time Estimate (Within-Person) \times Feedback Condition (Given)	3.17	0.97	1,004	3.25**
Phase 1 response time estimate (between-person)	-7.05	2.78	68	-2.54*
Phase 1 Response Time Estimate (Between-Person) \times Age (Younger)	9.76	3.68	68	2.65*

* $p < .05$. ** $p < .01$. *** $p < .001$.

We describe and evaluate significant predictors of CJs beginning at the top of the table. Given dummy variable coding on the between-subjects variables of age and feedback, the intercept estimate corresponded to mean confidence of 64% in the older adults not given RT feedback. Age and feedback effects were not reliable ($p > .05$). An increase of 1% in a person's mean JOLs was associated with a 0.5% increase in test confidence, independent of any effect of actual recognition accuracy, though correct recognition responses were indeed rated higher in confidence than those that were not recognized correctly.

A reliable random effect of equation intercept was found, estimated variance = 567 ($SE = 103, Z = 5.51, p < .001$). We also detected reliable random effects for test accuracy (not shown in Table 8), indicating individual differences in the degree to which CJs were related to recognition memory performance, estimated variance = 213 ($SE = 57, Z = 3.75, p < .001$). Thus, individuals varied reliably in the resolution of their CJs. Given that participants were approximately 20% more confident in accurate compared with inaccurate responses, with a standard deviation of 26 for the random effect, it appears that many individuals had moderate to poor resolution of CJs with respect to recognition accuracy.⁵

Differences in RT were also related to test confidence. Actual RTs were scaled in milliseconds. Exploring the obtained within-person Age \times RT interaction further, we found that younger adults experienced a 6% drop in confidence per 1-s increase in RT ($\beta = .006$), $t(519) = -4.93, p < .001$, whereas older adults experienced a 3% drop in confidence per 1-s increase in RT ($\beta = .003$), $t(482) = -5.77, p < .001$. Younger adults' test confidence therefore appeared to display a greater sensitivity to items' actual RT. However, we cannot discount the possibility that this interaction is illusory, as it may be the case that younger and older adults were monitoring their RTs in a more relative than absolute sense. Because younger adult RTs were approximately half the magnitude of older adult RTs, a 6% drop in confidence per second increase in younger adult RTs could be equivalent to a 3% drop in confidence in older adult RTs.

Subjective RT estimates were scaled in 500-ms bins. There was a reliable effect of estimated RT (within-subjects) on CJs, with faster estimated RTs associated with higher confidence. This effect was qualified by an interaction with RT feedback condition. The effect of estimated RT was not reliable for those who received RT feedback ($\beta = -0.72$), $t(550) = -1.11, p > .25$, but there was an average within-person decrease of about 2% confidence per RT estimate bin for those who did not ($\beta = -2.43$), $t(451) = -2.57, p < .05$. Converting the binned estimates, those who did not receive feedback displayed a reliable 4% decrease in confidence per second increase in subjective fluency. Thus, feedback increased the correspondence of estimated RTs to actual RTs and thereby eliminated the effect of estimated RT on CJs. Absent feedback, subjective fluency and actual fluency were independently related to participants' confidence in their recognition memory responses at the item level. Between-person differences in RT monitoring were also related to CJs. An increase of one RT estimate bin (i.e., 0.5 perceived s) was associated with a 4.5% increase in younger adult CJs ($\beta = 4.52$), $t(35) = 2.18, p < .05$, and a 7.8% decrease in older adult CJs ($\beta = -7.78$), $t(31) = -2.27, p < .05$. Each age group may have interpreted lower response fluency differently (i.e., a more careful vs. a more difficult memory search).

To summarize, both actual performance and metacognitive factors displayed independent relationships with recognition test confidence during Phase 1. Within-person differences in actual RT were more related to younger adults' CJs, whereas differences in subjective or estimated RT were more related to older adults' CJs. For participants not given RT feedback, individual differences in subjective fluency were related to CJs and could also directly or indirectly influence future study allocation.

Study time allocation

Table 9 reports the final regression model predicting Phase 2 study time allocation. The intercept was scaled to older adults, who took on average about 6.2 s for self-paced study. We found reliable random effects for the equation intercept, estimated variance = 3,704,791 ($SE = 770,648, Z = 4.81, p < .001$). Younger adults studied

approximately 3 s less than older adults. Individuals studied items that were recognized correctly during Phase 1 about 1.2 s less than items that were unrecognized.

Table 9
Multilevel Regression Analysis Predicting Phase 2 Study Time Allocation (Intact Word Pairs)

Effect	Estimate	SE	df	t
Intercept	6,210.0	356.1	70	17.44***
Age (younger)	-3,158.1	488.7	70	-6.46***
Phase 1 recognition accuracy	-1,237.4	309.9	993	-3.99***
Phase 1 judgment of learning (within-person)	-17.3	5.9	993	-2.96**
Phase 1 judgment of learning (between-person)	61.2	17.1	70	3.58***
Phase 1 Judgment of Learning (Between-Person) × Age (Younger)	-90.6	23.5	70	-3.86***
Phase 1 confidence judgment (within-person)	-24.4	7.3	993	-3.36***
Phase 1 Confidence Judgment (Within-Person) × Age (Younger)	21.2	9.2	993	2.30*
Phase 1 response time (within-person)	0.3	0.1	993	4.10***
Phase 1 response time estimate (within-person)	233.0	120.2	993	1.94
Phase 1 Response Time Estimate (Within-Person) × Age (Younger)	-319.2	162.2	993	-1.97*

* $p < .05$. ** $p < .01$. *** $p < .001$.

Both within-person and between-person effects of Phase 1 study confidence (i.e., JOLs) were found, as were separate between-person effects by age group. Participants studied items 17 ms less during Phase 2 for each point increase in Phase 1 JOL; perceptions of initial item difficulty therefore persisted to Phase 2. Figure 3 plots the age-related interactions of our metacognitive variables and Phase 2 study time. A 1% increase in between-person JOL confidence was associated with a 29-ms decrease in study time for younger adults ($\beta = -29.39$, $t(37) = -2.34$, $p < .05$), but a 61-ms increase in study time for older adults ($\beta = 60.73$, $t(33) = 2.99$, $p < .01$). Whereas younger adults with overall higher JOLs predictably allocated less study time at Phase 2, older adults with overall higher JOLs allocated more study time at Phase 2. Perhaps older adults who were confident in their learning after Phase 1 study engaged in more maintenance rehearsal to maintain their perceived high level of item learning, or perhaps these older adults were more sensitive to recognition difficulty at Phase 1 test and consequently studied more at Phase 2.

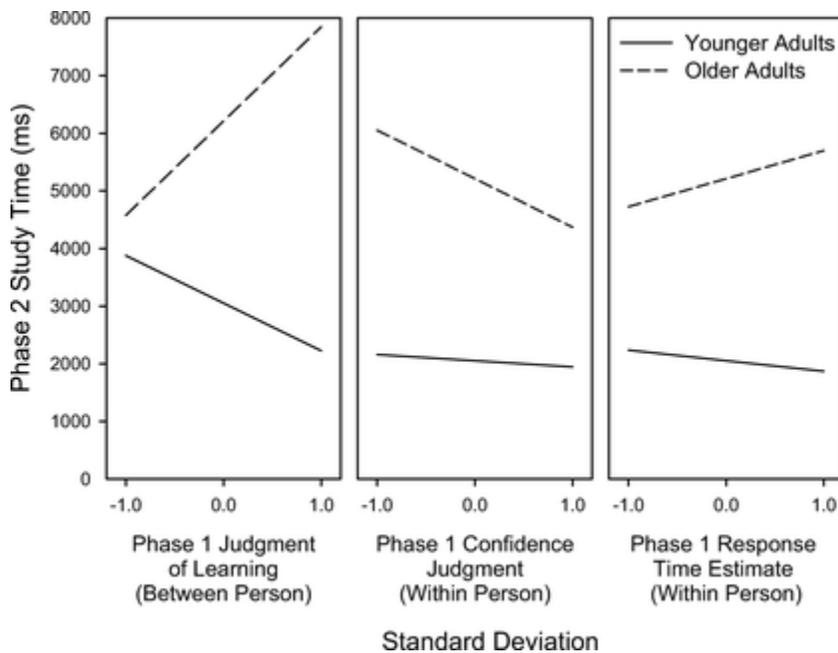


Figure 3. Age interactions with Phase 1 between-person judgments of learning, within-person confidence judgments, and within-person response time estimates on Phase 2 study time allocation.

Also critical to the metacognitive hypothesis, there was a reliable within-person effect of CJs, controlling for test accuracy, with higher CJs associated with shorter study time at Phase 2, as predicted. This effect was also qualified by a significant age interaction (see Figure 3). For each 1% increase in Phase 1 CJs, younger adults studied 3 ms less ($\beta = -2.98$, $t(517) = -1.36$, $p > .15$), whereas older adults studied 24 ms less ($\beta = -24.10$, $t(473) = -2.47$, $p < .05$). Items with lower confidence at Test 1 were indeed studied longer by older adults at Phase 2. The magnitude of the slopes was reliably larger for older adults, a finding inconsistent with the hypothesis of an age-related deficiency in the use of monitoring.

A within-person increase of 1 ms in Phase 1 recognition RT was associated with an increase of 3 ms in Phase 2 study time. Participants therefore discriminated between items on the basis of Phase 1 actual RT and allocated Phase 2 study time accordingly. There was no reliable main effect for Phase 1 within-person estimated RT, although there was a reliable interaction of this variable with age (see Figure 3). For each bin increase in estimated RT, younger adult study times decreased by 87 ms ($\beta = -86.76$, $t(517) = -.25$, $p > .80$), whereas older adult study times increased by 233 ms ($\beta = 232.86$, $t(473) = 1.40$, $p > .15$). Neither coefficient was reliable separately in the two age groups, so we cannot conclude there is an effect of subjective fluency on study time allocation, independent of other metacognitive judgments. Given the robust effects of RT estimates on CJs (see Table 8), it appears the principal source of an effect of RT monitoring accuracy on study time allocation was mediated through CJs. It is important to note, however, that the effects of RT estimates on CJs were only reliable when RT estimates were inaccurate (i.e., in the condition with no RT feedback) and that the effects of CJs on study time were reliable only for older adults. For example, an older adult who erroneously perceived greater response fluency at Phase 1 test may have been more confident in his or her item knowledge. The individual may have therefore studied this item less in Phase 2 than he or she would have with accurate performance monitoring. In this way, RT monitoring accuracy could influence both Phase 1 recognition confidence and (mediated by CJs) Phase 2 study time.

To summarize, both actual and metacognitive factors exerted independent effects on Phase 2 study time allocation. Phase 1 JOLs influenced Phase 2 study, independent of the influence noted on recognition confidence, with less study for items that had higher JOLs, and contrasting individual difference patterns by age group such that younger adults with high JOLs studied less whereas older adults with high JOLs studied more. Recognition CJs also showed different effects for the two age groups such that older adults (but not younger adults) studied items less if they had been confident in their recognition response. The actual RT for item recognition impacted item study time independent of the influence noted through CJs. Subjective fluency did not have a direct effect on study time allocation, independent of JOLs and CJs.

Discussion

These data provide clear evidence that study time allocation is related to metacognitive processes, independent of levels of performance. Both Phase 1 JOLs and CJs predicted Phase 2 study time allocation, controlling for Phase 1 recognition memory performance. Although the relationship between memory performance and study time allocation could reflect effects of item difficulty that are independent of metacognitive control (Koriat et al., 2006), our outcomes for JOLs and CJs indicate that study time is also influenced by metacognitive factors. Specifically, when individuals have lower confidence in (a) the quality of their initial item learning and (b) the accuracy of their recognition response, they study that item longer at Phase 2.

Older adults allocate study time during self-paced study on the basis of metacognitive monitoring in a manner similar to young adults. Consistent with the hypothesis of spared monitoring, there were no reliable age differences in the resolution of JOLs or CJs, which were similarly correlated with recognition memory accuracy for both age groups. However, our results did not support the hypothesis of impaired use of monitoring. Older adults showed a robust effect of JOLs and CJs on Phase 2 study time, and CJs actually had a greater influence on study time for older adults relative to younger adults. These results are inconsistent with a metacognitive control deficit, as exemplified by Dunlosky and Connor's (1997) findings regarding older adults' study time allocation. There are a number of differences in method between the studies, including the criterion memory test (cued recall in Dunlosky & Connor's [1997] study vs. associative recognition in this study). One advantage of

the present study is that our manipulation of presentation time—providing more study time for older adults in Phase 1—was successful in equating younger and older adults in recognition memory test performance, making comparisons of relations between recognition accuracy, CJs, and study time allocation more interpretable. Dunlosky and Connor's experiment produced substantial age differences in recall performance at Phase 1, consistent with the literature on slowed rates of learning in older populations (Kausler, 1994). Resolution of the discrepancy requires new experiments.

We note that the present results are consistent with other findings that older adults use a similar approach to selecting items for restudy based on past memory test outcomes (Dunlosky & Hertzog, 1997). It seems that age-related differences in metacognitive control of learning emerge under some conditions and not others. At present, we do not understand the moderating variables that determine whether age differences in metacognitive control will arise.

The present study differs from earlier work in its focus on performance monitoring and CJs. CJs had a higher gamma correlation with study time allocation than JOLs, corresponding to the hypothesis that performance monitoring may be most relevant to future study time allocation. However, when all variables were in the multilevel regression model, the regression effects for CJs and JOLs were comparable. Immediate JOLs can reflect a number of aspects of encoding that are related to study time, including whether a mediator was formed and the fluency of encoding (Hertzog et al., 2003; Robinson et al., 2006). Thus, although the present results justify evaluating performance monitoring via CJs, they also implicate other metacognitive influences that are independent from CJs or recognition memory accuracy. In contrast, the fact that recognition accuracy influences study time allocation independent of CJs and JOLs indicates that explicit metacognitive monitoring processes per se do not adequately capture all influences on study time (Koriat et al., 2006). The complex pattern of outcomes indicates the importance of considering multiple variables' influences in a simultaneous method, such as multilevel regression.

An interesting new finding in this study is the relationship of subjective RT estimates to CJs and study time allocation. Subjective fluency of recognition responses, as measured by Phase 1 recognition RT estimates, reliably predicted CJs independently from actual RTs (actual fluency) when participants are provided no feedback about actual RT. This finding is interesting for two reasons. First, it is, to our knowledge, the first demonstration that subjective retrieval fluency has an impact on response confidence and (mediated through confidence) on subsequent study time allocation. Second, it suggests that retrieval fluency effects that have been observed in the literature on memory and metamemory may be associated with conscious awareness and perhaps even explicit monitoring of retrieval fluency. Given that fluency is a construct that has had considerable impact on theorizing about metacognition (e.g., Benjamin et al., 1998; Hertzog et al., 2003; Koriat & Ma'ayan, 2005), the present result should focus attention on the benefits of explicitly measuring subjective fluency. For instance, an interesting question is whether aspects of fluency impact cognition and metacognition independent of any subjective awareness of fluency itself (e.g., Jameson, Narens, Goldfarb, & Nelson, 1994).

Another important contribution of the current study is the demonstration of an age-related reduction in the accuracy of estimated RTs during a recognition memory task (see also Hertzog et al., 2007). Both younger and older adults underestimated recognition test RT without corrective feedback, but the degree of underestimation was larger for older adults. This deficit was associated with a larger within-persons relationship between estimated RT and recognition CJs for younger adults but a larger between-persons relationship between estimated RT and CJs for older adults. The within-person outcome suggests that younger adults' CJs across items are more affected by variability in RT across items, whereas in the older group, mean CJs are more influenced by mean estimated RTs. This latter result raises the possibility that illusory overconfidence by older adults in recognition tests is influenced by a global false impression of fast responding. It is important to note that yes-no item recognition tests, which (a) are differentially influenced by fast RT for familiarity-based responses (Yonelinas & Jacoby, 1994) and (b) show greater age dependence on familiarity over recollection (e.g., Jennings & Jacoby, 1997; see Light, Prull, La Voie, & Healy, 2000), may be much more prone to illusory fluency effects than the type of associative recognition test we studied in this experiment.

Despite these age differences in subjective fluency, we found no age differences in the linkage of CJs to subsequent study time. This result parallels a finding of Robinson et al. (2006) regarding encoding fluency. In that study, older adults' subjective encoding fluency (measured by estimated encoding times) was less accurate than younger adults, but both age groups showed a similar effect of subjective encoding fluency on JOLs. In the present study, differential effects of retrieval fluency (as measured by estimated recognition RTs) on CJs did not result in differential effects of CJs on study time allocation for young and older adults. Such findings underscore the argument that one cannot merely rely on logical inference to assume that age differences in monitoring accuracy necessarily imply deficient metacognitive influences on control. An interesting question for future research is whether manipulations that influence the accuracy of CJs alter the correlation between memory performance and subsequent study time by strengthening the performance–CJ relationship.

It would be useful to replicate and extend the present findings regarding subjective fluency and study time allocation to a modified associative recognition task. A limitation of the present study is that RT estimates were obtained on only half the trials (because of the inclusion of RT feedback). We successfully showed feedback effects on estimated RT, which supports the argument that this measure can be treated as subjective fluency that is distinct from actual RT. The feedback manipulation also allowed us to demonstrate that inaccurate RT estimates have a distinct impact on recognition confidence. However, it would be useful to evaluate subjective fluency in which estimates were obtained on all trials. Moreover, the binned RT estimate approach may have affected the magnitude of the relationship of estimated RT to CJs and study time allocation.

Another limitation of the present task was the fact that participants were tested on intact pairs for only half the items in Phase 1. Rearranging study items at test had a substantial effect on CJs, and this effect appears to be of greater magnitude for older adults. Despite age equivalence in recognition accuracy, gamma correlations revealed that older adults were less confident in their recognition of word pairs that were rearranged at test. Additionally, Phase 1 CJs showed no correlation with Phase 2 study time for rearranged items. This outcome can be understood in part because successful rejection of rearranged pairs can be made on the basis of recalling either of the two original pairings rather than both of them (e.g., Cohn & Moscovitch, 2007). Hence, one could be confident in rejecting a rearranged pair but have low confidence that one has learned one of the two constituent pairs involved in the rearranged test probe. Such an effect would dilute CJ–study time correlations. New research would be needed to evaluate these and other reasons why CJs for rearranged pairs were less accurate and did not predict study time allocation.

In conclusion, this study shows that metacognitive states—as captured by JOLs, CJs, and subjective recognition test fluency—influence study time allocation independently of actual aspects of recognition memory performance, such as accuracy and latency. These phenomena support a model of metacognitive self-regulation positing that monitoring affects control (Nelson, 1996). Furthermore, the results indicate no age deficits in the use of metacognitive monitoring to control study time allocation. This result departs from other results in the literature and raises the possibility that age differences in self-regulation may depend on specific features of experimental tasks and the processes they invoke.

Footnotes:

1 Despite the common usage of the gamma correlation, it should be noted that such correlations involving study time for rearranged items should be interpreted with caution. Although it is only possible to correctly recognize an intact item by remembering the association from a study trial, it is possible to correctly reject a rearranged word pair by remembering the matching word for either the cue or target in the rearranged pair. Without knowing the source of the correct rejection, it is impossible to calculate a proper gamma. More general issues exist as well. First, gamma correlations focus on bivariate relationships (e.g., CJs and accuracy), ignoring all other influences on both variables. Second, the comparison of gamma correlations risks aggregation bias in both estimation and inference (Singer, 1998).

2 In the current study, we separated within-person and between-person influences by creating two orthogonal transformed predictors for each variable. First, within-person measures of Phase 1 recognition accuracy, RT,

estimated RT, confidence, and study time for each person were computed by subtracting each person's mean for each variable from that person's trial-level scores on each variable (thereby creating a difference score for each trial). Centering on individual-level means allowed examination of within-person variation on the dependent variable from within-person variation of its predictors. Second, each person's mean on the relevant variable was computed (e.g., mean CJs), and then a between-persons difference score was computed by subtracting the sample mean for that variable from the person's mean. Centering on group-level means captured individual differences in the average scores of each variable. Because these two transformed variables are orthogonal, both can be included simultaneously in a regression equation. For example, a reliable negative effect of RT, centered at the individual level, on CJs would indicate that people were more confident in the accuracy of items they recognized quickly versus those they recognized more slowly. In contrast, a reliable negative effect of group-mean-centered RT would indicate that people who were generally slower to respond were generally less confident.

3 It should be noted that an attempt was made to examine more directly the role of stimulus effects on the relationships presented in the above multilevel regression analyses. This was accomplished by adding stimulus and many different possible interaction terms, such as Stimulus \times Age and Stimulus \times Age \times Phase 1 Test Accuracy, as independent variables in each equation. No reliable main effects or interactions involving stimulus were found for Phase 1 test confidence, but a main effect of stimulus was found for Phase 2 study time. Upon closer examination, the latter effect was limited to a few word pairs that appeared to be less concrete than the others (e.g., *EXERCISE-BANK* and *MANAGER-BLOCK*), and the result was essentially that less concrete items were studied longer at Phase 2. However, because stimulus effects were limited to so few items and failed to substantively impact the relationships between other variables in the equations presented above, we chose to exclude stimulus effects from the above-presented equations.

4 Individual differences in predictor effects are captured in estimates of random effects in slopes. Readers may wish to think of effects labeled as random in our regression models as having varying intercepts or varying slopes (see Gelman & Hill, 2007).

5 Assuming a normal distribution, 68% of the distribution of regression coefficients lay between approximately -6 and 46 ; participants therefore varied considerably in the degree to which test accuracy related to confidence reports.

References:

- Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31*, 297–305.
- Benjamin, A. S., & Bjork, R. A. (2000). On the relationship between recognition speed and accuracy for words rehearsed via rote versus elaborative rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 638–648.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68.
- Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review, 4*, 184–197.
- Block, R. A., Zakay, D., & Hancock, P. A. (1998). Human aging and duration judgments: A meta-analytic review. *Psychology and Aging, 13*, 584–596.
- Bottiroli, S., Dunlosky, J., Guerini, K., Cavallini, E., & Hertzog, C. (2008). *Does strategy affordance moderate age-related deficits in strategy production?* Unpublished manuscript.
- Cohn, M., Emrich, S. M., & Moscovitch, M. (2008). Age-related deficits in associative memory: The influence of impaired strategic retrieval. *Psychology and Aging, 23*, 93–103.
- Cohn, M., & Moscovitch, M. (2007). Dissociating measures of associative memory: Evidence and theoretical implications. *Journal of Memory and Language, 57*, 437–454.
- 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.
- Cull, W. L., & Zechmeister, E. B. (1994). The learning ability paradox in adult metamemory research: Where are the metamemory differences between good and poor learners? *Memory & Cognition, 22*, 249–257.

- Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology and Aging, 22*, 122–133.
- Doshier, B. A. (1982). Effect of sentence size and network distance on retrieval speed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 173–207.
- Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25*, 691–700.
- Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multi-trial learning. *Journal of Gerontology: Psychological Sciences, 52*, 178–186.
- Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology and Aging, 13*, 597–607.
- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15*, 462–474.
- Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when JOLs occur? *Journal of Memory and Language, 33*, 545–565.
- Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 238–244.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, England: Cambridge University Press.
- Healy, M. R., Light, L. L., & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 768–788.
- Hertzog, C., Dixon, R. A., & Hulstsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task-performance in adults. *Psychology and Aging, 5*, 215–227.
- Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 22–34.
- Hertzog, C., & Hulstsch, D. F. (2000). Metacognition in adulthood and old age. In T. A. Salthouse & F. I. M. Craik (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 417–466). Mahwah, NJ: Erlbaum.
- Hertzog, C., Price, J., & Dunlosky, J. (2008). How is knowledge generated about memory encoding strategy effectiveness?. *Learning and Individual Differences, 18*, 430–445.
- Hertzog, C., Touron, D. R., & Hines, J. C. (2007). Does a time monitoring deficit influence older adults' delayed retrieval shift during skill acquisition? *Psychology and Aging, 22*, 607–624.
- Higham, P. A. (2002). Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory & Cognition, 30*, 67–80.
- Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods, 39*, 101–117.
- Jameson, K. A., Narens, L., Goldfarb, K., & Nelson, T. O. (1994). The influence of subthreshold priming on metamemory and recall. *Acta Psychologica, 73*, 55–68.
- Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology and Aging, 12*, 352–361.
- Kausler, D. H. (1994). *Learning and memory in normal aging*. San Diego, CA: Academic.
- Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language, 48*, 704–721.
- Kelly, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica, 98*, 127–140.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.
- Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learner's sensitivity to retrieval conditions at test. *Memory & Cognition, 34*, 959–972.
- Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language, 52*, 478–492.

- Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*, 36–69.
- Light, L. L., Prull, M. W., La Voie, D. J., & Healy, M. R. (2000). Dual-process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 238–300). London: Oxford University Press.
- Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (2000). *SAS system for mixed models* (4th ed.). Cary, NC: SAS Institute, Inc.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*, 174–179.
- Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*, 463–477.
- Murphy, M. D., Schmitt, F. A., Caruso, M. J., & Sanders, R. E. (1987). Metamemory in older adults: The role of monitoring in serial recall. *Psychology and Aging*, *2*, 331–339.
- 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. (1993). Judgments of learning and the allocation of study time. *Journal of Experimental Psychology: General*, *122*, 269–273.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*, 102–116.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science*, *2*, 267–270.
- Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, *5*, 207–213.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 501–508). San Diego, CA: Academic Press.
- Riggs, K. M., Lachman, M. E., & Wingfield, A. (1997). Taking charge of remembering: Locus of control and older adults' memory for speech. *Experimental Aging Research*, *23*, 237–256.
- Robinson, A. E., Hertzog, C., & Dunlosky, J. (2006). Aging, encoding fluency, and metacognitive monitoring. *Aging, Neuropsychology, and Cognition*, *13*, 458–478.
- Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *24*, 323–355.
- Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modelling*. London: Sage.
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1024–1037.
- Touron, D. R., & Hertzog, C. (2004). Strategy shift affordance and strategy choice in young and older adults. *Memory & Cognition*, *32*, 298–310.
- Weaver, C. A., & Kelemen, W. L. (2003). Processing similarity does not improve metamemory: Evidence against transfer-appropriate monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1058–1065.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: Psychological Corporation.
- Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, *49*, 62–79.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekarts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego, CA: Academic Press.
- Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and response speed. *Canadian Journal of Experimental Psychology*, *48*, 516–534.
- Zachary, R. (1986). *Shipley Institute of Living Scale Revised manual*. Los Angeles: Western Psychological Services.