The Effectiveness of Updating Metacognitive Knowledge in the Elderly: Evidence From Metamnemonic Judgments of Word Frequency

Jonathan G. Tullis and Aaron S. Benjamin
University of Illinois at Urbana-Champaign

Accurate metacognitive knowledge is vital for optimal performance in self-regulated learning. Yet older adults have deficiencies in implementing effective learning strategies and knowledge updating and consequently may not learn as effectively from task experience as younger adults. Here we assess the ability of older adults to update metacognitive knowledge about the effects of word frequency on recognition. Young adults have been shown to correct their misconceptions through experience with the task, but the greater difficulty older adults have with knowledge updating makes it unclear whether task experience will be sufficient for older adults. The performance of older adults in this experiment qualitatively replicates the results of a comparison group of younger subjects, indicating that both groups are able to correct their metacognitive knowledge through task experience. Older adults seem to possess more effective and flexible metacognition than sometimes suggested.

Keywords: metacognition, monitoring, knowledge updating, word frequency, judgments of learning

The sophisticated use of metacognitive strategies is critical to being an effective learner (Finley, Tullis, & Benjamin, 2009). However, having control over learning is only effective when metacognitive judgments are accurate (Theide, Anderson, & Therriault, 2003; Tullis & Benjamin, 2011). Because learners' ideas about the memorability of varying materials or the effectiveness of varying study regimens are often incorrect, being able to learn from one's mistakes is particularly important. Correcting inappropriate metacognitive knowledge through direct task experience is especially important for older adults, who must adapt to changing environments and who may be particularly naïve about the differential effectiveness of many learning strategies (Brigham & Pressley, 1988; Hertzog & Hultsch, 2000). The goal of the present experiment is to evaluate whether task experience can help older adults correct a particular misconception about the effects of word frequency on recognition in a multiple study-test paradigm. This paradigm has been shown to lead younger adults to correctly appreciate the effects of word frequency, so those results can serve as a benchmark with which to assess the efficacy of metacognitive updating in the older population.

Metacognitive Accuracy and Effective Learning

The accuracy of metacognitive knowledge and monitoring is crucial to producing optimal performance during self-regulated learning because metacognitive knowledge directs metacognitive control over learning (Metcalfe & Finn, 2008; Nelson & Narens, 1990; Theide & Dunlosky, 1999). Errors in metacognitive knowledge and monitoring can, thus, lead to deficient use of control during study and suboptimal performance (Atkinson, 1972; Karpicke, 2009; Kornell & Bjork, 2008; Tullis & Benjamin, 2011). It is important to note that younger adults have shown some ability to correct metacognitive misconceptions through direct task experience (Benjamin, 2003; Brigham & Pressley, 1988; Finley & Benjamin, 2011). The ability of older adults to adapt their metacognitive knowledge to new situations through experience, however, is not as clear. Considerable evidence suggests that older adults do not update their metacognitive knowledge as effectively as younger adults (Bieman-Copland & Charness, 1994; Brigham & Pressley, 1988; Matvey, Dunlosky, Shaw, Parks, & Hertzog, 2002), as we review below.

Knowledge Updating in Elderly Learners

Brigham and Pressley (1988) investigated metacognitive ratings of the effectiveness of several encoding strategies during a vocabulary learning task. They found that both younger and older adults began with misconceptions about the effectiveness of different strategies for learning new word meanings, rating a "keyword" mnemonic strategy—in which subjects made up sentences that contained a word that sounds like the target and a synonym for the target—as less effective than a "semantic context" strategy, in which subjects made up sentences correctly using the target word. However, by the end of the task, younger adults had updated their knowledge and correctly rated the keyword mnemonic strategy as more effective than the semantic context strategy. Older adults' ratings of strategy effectiveness did not change across multiple study-test cycles, indicating a failure to update their knowledge through experience.

Similarly, Bieman-Copland and Charness (1994) reported that older adults failed to accurately update their knowledge regarding

differential cue effectiveness through experience in a cued recall task. After learning a list of word pairs and being tested on those pairs with different types of cues, younger adults modulated their ratings of cue effectiveness to correctly reflect the differential effectiveness of the types of cues used (rhymes, letters, or category cues); older adults did not adjust their ratings appropriately. Matvey et al. (2002) used this same cue-effectiveness paradigm and found a similar effect: Younger adults adjusted predictions in a second study-test cycle in a manner that reflected the effectiveness of that cue type during the first cycle, whereas older adults did not modulate their global predictions based on the type of cue. However, Matvey et al. also showed that both younger and older adults' relative predictive accuracy (gamma correlations) improved across cycles, suggesting that both younger and older adults did acquire some knowledge about how to predict performance in that task. Some evidence suggests that older learners may gain metacognitive knowledge through task experience in situations where learners do not need to compare the effectiveness of different learning strategies or classes of items. For instance, when learners bet on memory performance, older subjects learned to calibrate memory predictions as accurately as younger learners after several study–test trials (McGillivray & Castel, 2011).

Several theories have been proposed to explain why older adults modulate their metacognitive beliefs less effectively than younger adults. Both Brigham and Pressley (1988) and Bieman-Copland and Charness (1994) attributed the failure to update metacognitive knowledge to a reduction in the effectiveness of their metacognitive monitoring. They theorized that the cognitive demands of online monitoring may exceed the limited attentional and working memory capacities that older adults possess. Accurate monitoring requires that learners engage in the primary memory task while also generating immediate feedback about their performance, the variables that influenced their performance, and any discrepancies between their predictions and performance. Limited cognitive resources may impair older adults' monitoring because they reduce detection of the encoding variables that influence later performance. Matvey et al. (2002) posited that knowledge updating fails because older adults either fail to remember which strategy was used with which item, fail to accurately tally the items remembered from each strategy, or fail to differentiate each strategy's effectiveness from the overall amount recalled. Such difficulties might result from a general slowing in processing speed associated with aging and may prevent older adults from updating their beliefs through experience (Bieman-Copland & Charness, 1994; Price, Hertzog, & Dunlosky, 2008; see Salthouse, 1991). Others have suggested that older adults have significant deficits in associative learning, which prevent them from connecting each strategy with its effectiveness (Naveh-Benjamin, 2000; Price, Hertzog, & Dunlosky, 2008). The data reported here bear on the viability of these theoretical suggestions.

Certain conditions exist in which older adults maintain some ability to effectively update metacognitive knowledge. In one experiment (Dunlosky & Hertzog, 2000), older and younger adults learned word pairs either by repeating the items to themselves or by forming an interactive image across two cycles. Both age groups showed significant gains in between-person correlations between mean recall predictions and final cued recall across two study-test cycles, indicating that both groups updated their metacognitive judgments to more accurately reflect individual differ-

ences in overall memory performance. However, the interpretation of this result is confounded by their failure to find differences in either absolute accuracy or within-person judgments of learning across the two cycles. Dunlosky and Hertzog (2000) posited that the effects of deliberate attention might underlie the inability of older adults to update their knowledge. They suggested that older adults can update their metacognitive knowledge as effectively as young adults only under conditions in which the differential effect of the strategies is large enough to capture attention and when postdictions are used to draw attention to the relative effectiveness of strategies.

Very little research specifically addresses the conditions that promote metacognitive knowledge updating. In the only study that directly examined the effect of learning conditions on knowledge updating in the older population, Price et al. (2008) showed that, unlike younger learners, older adults did not benefit (or at least not much) from having differentially effective learning strategies blocked rather than intermixed.

To summarize these results, there is ample evidence that older adults do not update metacognitive knowledge as effectively as younger adults—in fact, there are to date *no* identified conditions in which older adults convincingly exhibit metacognitive knowledge updating—and there is little understanding of how different learning conditions affect the difficulty of knowledge updating. In this experiment, we investigated the ability of older adults to update their metacognitive knowledge in a task that is already well validated in younger adults and possesses certain key differences from the extant research discussed previously. The task is from Benjamin (2003; see also Guttentag & Carroll, 1998). We extend the purview of knowledge updating by examining not the effectiveness of different strategies or cues—as in all prior work—but rather the memorability of different types of items. Word frequency provides an ideal means of analyzing metacognitive knowledge updating for three reasons. First, word frequency is inextricably tied to the to-be-remembered stimulus, thus freeing learners from the burden of having to remember, for example, which stimulus was associated with which condition of learning. Second, novice learners often show misconceptions about its effect on recognition tasks (described in detail, below), thus providing a circumstance in which knowledge updating is possible. Finally, because predictions of memory performance are far more influenced by stimulus characteristics than by processing strategies used during learning (Koriat, 1997), and because word frequency has a large impact on memory predictions (Begg, Duft, LaLonde, Melnick, & Sanvito, 1989), the conditions are ones that align metacognitive tendencies with the potential for successful knowledge updating.

Novice learners fail to appreciate the fact that low-frequency (LF) words are easier to recognize than high-frequency (HF) words (Glanzer & Bowles, 1976; Gorman, 1961) and may sometimes even rate common words as more recognizable than less frequent items (Begg et al., 1989; Benjamin, 2003; Wixted, 1992). Because learners often have misconceptions about the relationship between word frequency and recognition, and because they are acutely sensitive to the effects of intrinsic characteristics like word frequency, one might expect to find evidence for metacognitive knowledge updating using these characteristics, even in older adults.

There is quite strong evidence for such updating in younger adults. In Benjamin (2003; Experiment 2), young learners predicted higher rates of recognition for HF items than for LF items during an initial study session. After a recognition test, subjects engaged in another cycle of learning and prediction followed by recognition. Predictions on the second study list correctly reflected the effect of word frequency on recognition, and LF words were rated as more memorable.[1]

Half of the subjects in that experiment were also asked to make postdictions during the first recognition test for only the items that they did not recognize. For these judgments, subjects rated how likely they *would have been* to recognize the item if they *had* studied it. Learners correctly postdicted that LF items would be more memorable. It is interesting to note that the learners who made postdictions during the first cycle appeared to have updated their metacognitive knowledge more dramatically: Their predictions in the second cycle revealed an even greater difference between LF and HF words than did those of the no-postdiction group (though this effect was not statistically significant). These two results indicate that younger subjects can update knowledge in this task, and that the act of making postdictions might aid in doing so.

The success of learners in that task can be contrasted with another recent example in which knowledge updating was not evident, even in younger learners. In an experiment with a similar two-cycle design (Diaz & Benjamin, 2011), subjects were exposed to the effects of proactive interference and release from proactive interference and did not learn to appreciate those effects (as revealed by memory predictions). The failure to learn under those conditions suggests that the factors enumerated previously in concern for the potential relationship between intrinsic factors and knowledge updating may be correct. The current experiment uses Benjamin's (2003) procedure to explore whether older adults can update their knowledge effectively.

In this experiment, younger and older adults predicted the likelihood of later recognizing HF and LF items during an initial study phase and then engaged in a recognition test. Subjects completed a standard recognition task, making "yes" or "no" responses to studied and new items. To more exactly replicate the prior work of Benjamin (2003), as well as to examine the secondary question of whether the act of making postdictions influences the magnitude of knowledge updating, half of the subjects made postdictions during the first recognition test. For every item they claimed not to recognize, they made judgments of their belief that they *would* have recognized it if it had been studied. Two study–test cycles were completed to investigate if learners changed their memorability predictions concerning HF and LF items and whether postdictions during the first cycle facilitated these changes.

Method

Subjects

Thirty-two community-dwelling older adults (age range = 60–84, *Mdn* age = 70, *SD* = 6.6) participated in exchange for nominal compensation. All older adults were high functioning: Performance on both the Mini-Mental State Exam (*M* = 28.5, *SD* = 1.3) and the Shipley Vocabulary scale (*M* = 34.9, *SD* = 3.8) was high. Seventy-six introductory-level psychology students from the Uni-

versity of Illinois at Urbana-Champaign also participated in exchange for partial course credit. The younger adults scored significantly lower on the Shipley Vocabulary scale than the older adults (*M* = 30.4, *SD* = 2.9), *t*(84) = 5.96.

Materials

The items used in this experiment were the same as those used by Benjamin (2003). The items were gathered from the compendium provided by Carroll, Davies, and Richman (1973) and consisted of 80 HF and 80 LF 4–8 letter nouns, verbs, and adjectives. The HF items averaged 100–270 on their scale, whereas the LF words averaged 5,000–5,230. Twenty HF items and 20 LF items were randomly intermixed in each study list. Each half of each study list contained 10 HF and 10 LF items and less than four items of a particular frequency in a row. The test list included all items studied in the preceding phase as well as 40 additional distractor items, such that each quarter of the test list contained an equal number of old and new items, as well as an equal number of HF and LF items. Presentation of the stimuli, as well as response recording, was done using the psychophysics toolbox extensions (Brainard, 1997) for MATLAB on PC computers.

Procedure

Subjects completed the experiment in individual rooms. Subjects read detailed instructions about the task, including a thorough explanation of the prediction ratings, recognition task, and the postdiction ratings, before they began. The directions emphasized that at the test, subjects would be given some previously studied and some new items, and they would have to judge if they had studied each item. An example of the recognition test and rating scale was given using the word *total*.

During the study phase, items were presented on the center of a computer monitor in black Times New Roman 80-point font for 4 s before being removed. Subjects predicted recall for each item immediately after its presentation on a scale that ranged from 1 (*I am sure that I will NOT remember this word*) to 9 (*I am sure that I WILL remember this word*) with interval gradations in between, and which was displayed at the bottom of the screen during all predictions. Predictions were self-paced: an item remained on screen until a subject made a prediction, at which point the next item was presented on the screen. Subjects made predictions on items until they cycled through the entire list.

Once they finished the study phase, subjects were instructed again in the details of the recognition task and given clear directions about the postdiction judgments that they needed to make whenever they did not endorse a word as previously studied. During the recognition test, subjects were asked if they previously studied each item and subjects responded with a simple "yes" or "no." For every "no" response during the test, subjects rated how likely they would be to recognize the word if they had studied it. Subjects made these individual postdictions on a scale that ranged from 1 (*I am sure I would NOT recognize this word*) to 9 (*I am sure I WOULD recognize this word*). The recognition test and the

---

[1] Though Benjamin (2003) did not report this particular test, the interaction between study–test cycle and word frequency on predictions was significant, *F*(1, 69) = 21.00.

postdictions were self-paced: Items remained on the screen until responses were made. After both judgments were made, a blank screen appeared for a 1-s interval before the next word was presented on screen. The study–prediction and test–postdiction cycle was repeated with the half of the items not used during the first cycle one minute after the completion of that cycle. Only half of the subjects made postdictions during the first cycle, and all subjects made postdictions during the test phase of the second cycle.

Results

The results of all inferential statistics reported below and throughout this article are reliable at the $\alpha < .05$ level by using two-tailed tests unless otherwise noted.

Recognition Performance

Performance on the recognition test is shown in Figure 1 for the younger subjects and in Figure 2 for the older subjects. In both study–test cycles and for all between-subjects conditions, the standard mirror effect for word frequency during recognition obtained (higher hits and lower false alarms for LF items). A 2 (study status: previously studied or new) $\times$ 2 (word frequency) $\times$ 2 (age) $\times$ 2 (postdiction group or not) repeated-measures analysis of variance (ANOVA) revealed a significant interaction between word frequency and study status on both tests, $Fs(1,104) = 177.62$ and 88.37 (Test 1 and Test 2, respectively), indicating a mirror effect.

The ANOVA also revealed a significant interaction between study status and age, such that younger learners had higher hit rates and lower false alarms than older learners on both tests, $Fs(1, 104) = 16.57$ and 20.19 (Test 1 and Test 2, respectively).

Knowledge Updating

We will next address the metacognitive measures, which are displayed in Figure 3 for the younger subjects and in Figure 4 for the older subjects. Of central importance is whether the effect of word frequency changed predictions in the appropriate direction between the two study-test cycles, and, secondarily, whether making postdictions moderated the size of that change. We present the results from the younger and older populations separately before comparing knowledge updating between the groups.

Metacognitive predictions in younger adults. A 2 (word frequency) $\times$ 2 (cycle) $\times$ 2 (postdiction or no-postdiction group) ANOVA on metacognitive predictions in younger learners revealed a significant interaction between frequency and cycle, $F(1, 74) = 11.36$, replicating previous results (Benjamin, 2003). This interaction shows that younger learners increase ratings of LF items more than HF items across study cycles, indicating that they update their knowledge about the mnemonic consequences of word frequency. Younger learners also showed significant interactions between cycle and postdiction group, $F(1, 74) = 3.96$, $p = .05$, and between frequency and postdiction group, $F(1, 74) = 8.95$, indicating that postdictions may play a significant role in
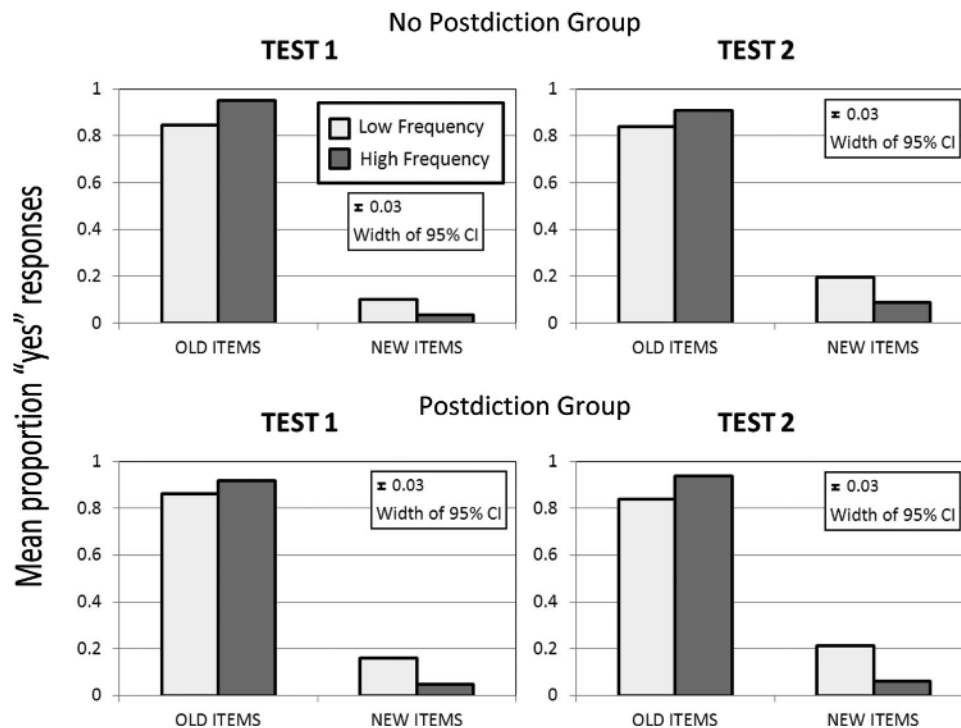

Figure 1. Proportion endorsed by younger subjects on the recognition test. Error bars show within subject 95% confidence intervals (CIs) for the interaction (see Benjamin, 2003; Loftus & Masson, 1994). Since the interaction variability does not provide the appropriate error term for pairwise comparisons, the error bars are not placed on the bars representing the means themselves.
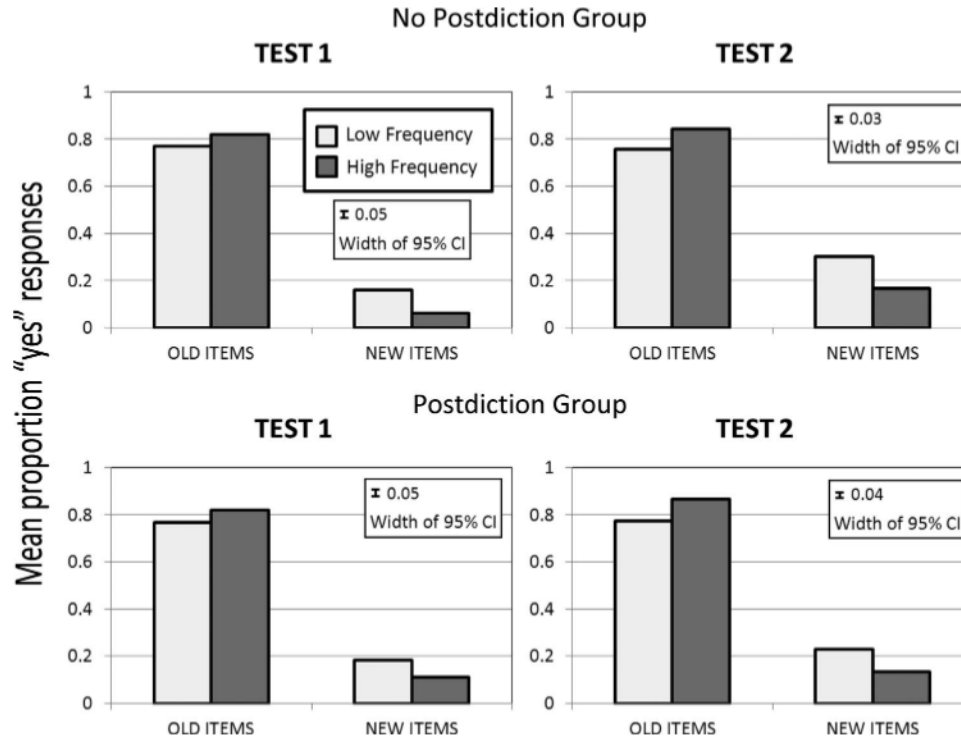
No Postdiction Group

Figure 2. Proportion endorsed by the older subjects on the recognition test. As described in Figure 1, error bars show within subject 95% confidence intervals (CIs) for the interaction (see Benjamin, 2003; Loftus & Masson, 1994).

knowledge updating in younger learners. Further, younger learners increased mnemonic predictions across cycles, $F(1, 74) = 22.79$. Unlike Benjamin's (2003) results, which showed that learners initially rate HF items as more memorable than LF items, young learners here rated LF items as more memorable than HF items across cycles, $F(1, 74) = 27.14$. Predictions of memorability in younger subjects for LF items were numerically higher than HF items during both the first and second cycles, $t(75) = 3.17$ and $t(75) = 6.66$, respectively.

Metacognitive predictions in older adults. A 2 (word frequency) $\times$ 2 (cycle) $\times$ 2 (postdiction or no-postdiction group) ANOVA on predictions in the older adult group revealed a significant interaction between frequency and cycle, $F(1, 30) = 16.77$, just as found in the sample of younger learners, indicating that older learners differentially shifted mnemonic predictions about HF and LF items across cycles. No other interactions or main effects reached significance. During the first study cycle, predictions of memorability across all older subjects for HF items were numerically higher than but did not significantly differ from LF items, $t(31) = 1.13$; $p > .25$. During the second study cycle, predictions of memorability across all older subjects for LF items were significantly higher than for HF items, $t(31) = 3.48$.

Metacognitive knowledge updating. In order to assess metacognitive knowledge updating across cycles, we computed the difference between LF item predictions and HF item predictions for each subject for each cycle. These results are summarized in Figure 5. A 2 (age) $\times$ 2 (cycle) $\times$ 2 (postdiction or no-postdiction group) ANOVA on the difference scores revealed a main effect of

cycle, $F(1, 104) = 26.92$, indicating that the difference in mnemonic predictions between LF and HF items increased across cycles. Additionally, the ANOVA showed a main effect of age, $F(1, 104) = 5.77$, indicating that young, more than old, learners rated LF items as more memorable than HF items across both cycles, and a main effect of postdiction group, $F(1, 104) = 4.58$, indicating that the postdiction group differentiated more between LF and HF items across cycles. No interactions reached significance. Critically, the magnitude of knowledge updating was actually slightly higher for older than younger adults.

The moderating effect of postdictions. Postdictions appear to have had a similar effect to that described earlier from Benjamin (2003): the numerical difference between low- and high-frequency words changes more across cycles in the postdiction group than in the no-postdiction group. The beneficial effect of postdictions for knowledge updating is most apparent in the younger population, who show a marginal three-way interaction between frequency, cycle, and postdiction group on their metacognitive predictions, $F(1, 74) = 3.66$, $p = .06$. When the data collected here is combined with the original young subject data from Benjamin (2003), the three-way interaction reaches significance, $F(1, 144) = 5.55$, indicating that knowledge is updated to a greater extent in the postdiction group than in the no-postdiction group. The beneficial effect of postdictions in the older population lies in the same direction as the young subjects but falls far short of reaching significance, $F(1, 30) = 0.29$; $p > .50$. However, the power to detect an effect of the same magnitude as that evident in the younger adults is only 0.08 here. Whether or not the beneficial
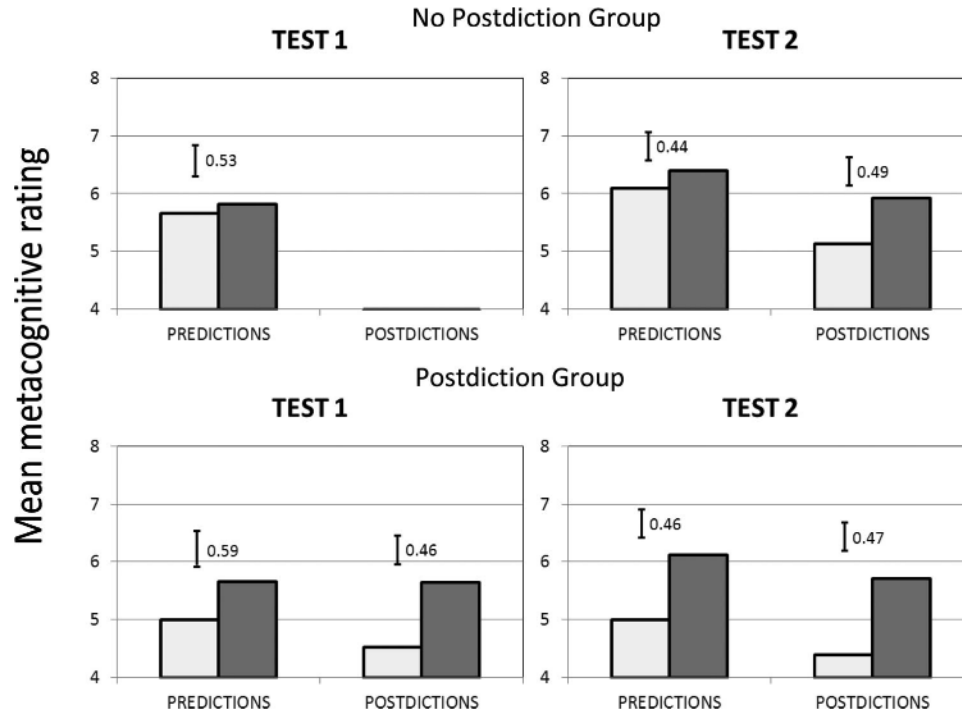
Figure 3. Metacognitive predictions and postdictions by younger subjects on a 1–9 scale. Error bars and values show the width of within subject 95% confidence intervals of the difference between high and low frequency ratings (see Figure 1). Error bars are not placed on the means themselves, however, because they show the variability of the differences between conditions.

effects of postdictions generalize to older adults cannot be determined from these data, though it should be noted that the effect size is smaller in the older adults.

Discussion

The goal of this experiment was to evaluate whether older adults exhibit knowledge updating in a task that is suited to the detection of such metacognitive knowledge. Much like younger adults, older adults changed their metacognitive beliefs about the effects of word frequency on memorability through task experience. Although their predictions for HF and LF items did not differ during the initial study phase, older adults changed their beliefs about the mnemonic influence of word frequency and rated LF items as more memorable than HF items during the prediction phase of the second cycle. The older adults updated their knowledge of word frequency effects to a numerically greater extent than the comparison group of younger subjects, thus revealing that, under these auspicious circumstances, knowledge updating occurs for older adults just as it does for younger adults. Unlike in Benjamin (2003), younger adults did not rate HF words as more memorable than LF words in the first prediction cycle; however, several other studies have also failed to detect this effect (in Guttentag & Carroll, 1998, see Experiment 1; in Wixted, 1992, see Experiment 5), indicating that other uncontrolled factors (such as population differences between the undergraduate subjects who participated in this study and those who participated in Benjamin [2003]) may affect this result. Knowledge updating is revealed by, of course,

the change in predictions across cycles and therefore is not particularly sensitive to the exact pattern evident in the first cycle.

The role that postdictions play in knowledge updating is less clear. For both younger and older adults, the shift across study-test cycles was more dramatic when postdictions were made during the intervening recognition test. This effect is small, however, and only detectable within the younger population when the young group was combined with prior data from young subjects. The effects of postdictions in promoting knowledge updating, and whether this effect differs across age groups, remains a target for future study. Postdictions may draw a learner's attention away from fleeting nondiagnostic cues and focus attention on more stable diagnostic cues (in this case, word frequency). Shifting the type of cues relied on has the potential to increase the accuracy of resulting predictions (e.g., Castel, 2008).

The impressive performance of older adults in this task bears on the competing hypotheses about knowledge updating in the older population reviewed earlier. If failures of knowledge updating reveal an inability to successfully associate a class of items or learning circumstances with an outcome, as proposed by Price et al. (2008), then the current task should have been no less difficult than previous ones in which older adults failed to update knowledge successfully. Our results, thus, speak against that explanation of such effects. Furthermore, in this task, older learners seem to have no trouble tallying successful recognition attempts across item type and modulating predictions for each type of item from
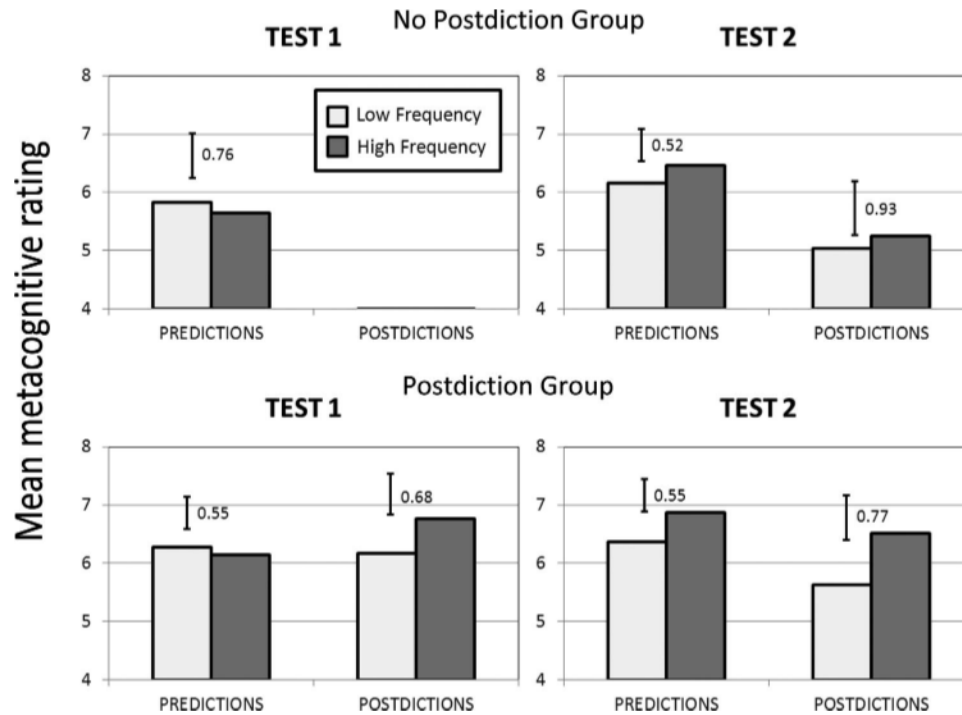
Figure 4. Metacognitive predictions and postdictions by older subjects on a 1–9 scale. Error bars and values show width of within subject 95% confidence intervals of the difference between high- and low-frequency ratings.

overall mnemonic performance (a potential deficiency hypothesized by Matvey et al. [2002]).

The results are, however, consistent with explanations that postulate limitations in working memory or attention as the basis for failures of knowledge updating. Because items of varying word frequency carry with them their stimulus class, thus relieving learners of the burden of remembering how each item was studied,
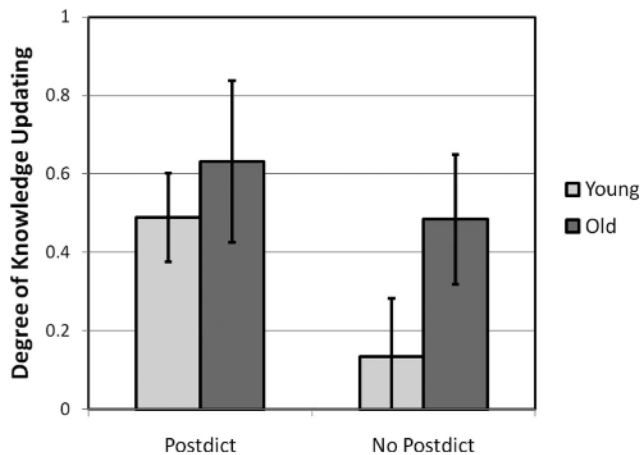

Figure 5. Knowledge updating, as defined by the difference between Cycle 2 and Cycle 1 of the differences between predictions for low-frequency and high-frequency items (LF pred – HF pred)$_{Cycle2}$ – (LF pred – HF pred)$_{Cycle1}$. Higher values indicate greater knowledge updating. Error bars show standard errors of the mean for each condition. Postdict = postdiction.

knowledge updating for such intrinsic stimulus characteristics should be easier than knowledge updating for manipulations of the extrinsic conditions of learning. Such an explanation clearly predicts that knowledge updating should be easier for intrinsic than extrinsic manipulations of memorability, a prediction that is supported by the successful performance here. This theory is corroborated by evidence that suggests that learners struggle to track the prior study conditions for items across long retention intervals, and this inability ultimately prevents accurate knowledge updating even in younger learners (Tullis, Finley, & Benjamin, 2011). However, because we have only tested knowledge updating of word frequency effects, the results reported here may be more specific to word frequency. Working memory load may be reduced specifically in our experiment because word frequency may be processed automatically (Hasher & Zacks, 1979). Further, given older adults' intact language skills and additional experience processing language, assessing the differential memorability of HF and LF items may not require any additional cognitive resources in older adults.

These results provide hope for the metacognitive abilities of older learners who need to adapt their metacognitive beliefs to changing internal, cognitive, and external, environmental conditions. We have shown that older learners can update their metacognitive knowledge about the mnemonic consequences of word frequency in conditions under which working memory may not be overtaxed by the demands of knowledge updating. Greater exploration of the circumstances that allow for accurate knowledge updating is warranted; for example, providing learners information about prior learning conditions at test may help reduce working

memory load and improve knowledge updating in older learners. The presence of this information may benefit younger learners, who can take advantage of it, more than older learners, who may not be able to take advantage of additional information because of limited cognitive resources. However, under auspicious circumstances, both younger and older learners may be able to update metacognitive beliefs such that their knowledge accurately reflects the true influence of memory activities. Accurate metacognitive monitoring should then lead to improved control over learning and ultimately benefit mnemonic performance.

References

Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology, 96,* 124–129. doi: 10.1037/h0033475

Begg, I., Duft, S., LaLonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory & Language, 28,* 610–632. doi:10.1016/0749-596X(89)90016-8

Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition, 31,* 297–305. doi:10.3758/BF03194388

Bieman-Copland, S., & Charness, N. (1994). Memory knowledge and memory monitoring in adulthood. *Psychology and Aging, 9,* 287–302. doi:10.1037/0882-7974.9.2.287

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging, 3,* 249–257. doi:10.1037/0882-7974.3.3.249

Carroll, J. B., Davies, P., & Richman, B. (1973). *Word frequency book.* New York, NY: American Heritage.

Castel, A. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition, 36,* 429–437. doi:10.3758/MC.36.2.429

Diaz, M., & Benjamin, A. S. (2011). The effects of proactive interference (PI) and release from PI on judgments of learning. *Memory & Cognition, 39,* 196–203. doi:10.3758/s13421-010-0010-y

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. doi:10.1037/0882-7974.15.3.462

Finley, J. R., & Benjamin, A. S. (2011). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. Manuscript submitted for publication.

Finley, J. R., Tullis, J. G., & Benjamin, A. S. (2009). Metacognitive control of learning and remembering. In M. S. Khine & I. M. Saleh (Eds.), *New science of learning: Cognition, computers and collaboration in education.* New York, NY: Springer Science & Business Media.

Glanzer, M., & Bowles, N. (1976). Analysis of the word frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning & Memory, 2,* 21–31. doi:10.1037/0278-7393.2.1.21

Gorman, A. N. (1961). Recognition memory for names as a function of abstractness and frequency. *Journal of Experimental Psychology, 61,* 23–29. doi:10.1037/h0040561

Guttentag, R., & Carroll, D. (1998). Memorability judgments for high- and low-frequency words. *Memory & Cognition, 26,* 951–958. doi:10.3758/BF03201175

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388. doi:10.1037/0096-3445.108.3.356

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* (pp. 417–466). Mahwah, NJ: Erlbaum.

Hertzog, C., Price, J., Burpee, A., Frentzel, W. J., Feldstein, S., & Dunlosky, J. (2009). Why do people show minimal knowledge updating with task experience: Inferential deficit or experimental artifact? *Quarterly Journal of Experimental Psychology, 62,* 155–173.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469–486. doi:10.1037/a0017341

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 549–570. doi:10.1037/0096-3445.126.4.349

Kornell, N., & Bjork, R. A. (2008). Optimizing self-regulated study: The benefits and costs of dropping flashcards. *Memory, 16,* 125–136. doi:10.1080/09658210701763899

Loftus, G. R. & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1,* 476–490.

Matvey, G., Dunlosky, J., Shaw, R. J., Parks, C., & Hertzog, C. (2002). Age-related equivalence and deficit in knowledge updating of cue effectiveness. *Psychology and Aging, 17,* 589–597. doi:10.1037/0882-7974.17.4.589

McGillivray, S., & Castel, A. D. (2011). Betting on memory leads to metacognitive improvement by younger and older adults. *Psychology and Aging, 26,* 137–142.

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15,* 174–179. doi:10.3758/PBR.15.1.174

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1170–1187. doi:10.1037/0278-7393.26.5.1170

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 1–45). New York, NY: Academic Press.

Price, J., Hertzog, C., & Dunlosky, J. (2008). Age-related differences in strategy knowledge updating: Blocked testing produces greater improvements in metacognitive accuracy for younger than older adults. *Aging, Neuropsychology, and Cognition, 15,* 601–626. doi:10.1080/13825580801956225

Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2,* 179–183. doi:10.1111/j.1467-9280.1991.tb00127.x

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95,* 66–73.

Theide, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items and self-paced study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 4,* 1024–1037. doi:10.1037/0278-7393.25.4.1024

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language, 64,* 109–118. doi:10.1016/j.jml.2010.11.002

Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2011). *Metacognition of the testing effect: When do learners predict the benefits of retrieval?* Manuscript submitted for publication.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 681–690.