Psychology and Aging
2002, Vol. 17, No. 2, 209–225

0882-7974/02/$5.00 DOI: 10.1037//0882-7974.17.2.209

Aging and Monitoring Associative Learning: Is Monitoring Accuracy Spared or Impaired?

Christopher Hertzog, Daniel P. Kidder, and
Amy Powell-Moman
Georgia Institute of Technology

John Dunlosky
University of North Carolina at Greensboro

Mixed lists of associatively related and unrelated paired associates were used to study monitoring of associative learning. Older and younger adults produced above-chance levels of relative accuracy, as measured by intraindividual correlations (γ) of judgments of learning (JOLs) with item recall. JOLs were strongly influenced by relatedness, and this effect was greater for older adults. Relative accuracy was higher for unrelated than for related pairs. Correlations of JOLs with item recall for a randomly yoked learner indicated that access to one's own encoding experiences increased relative accuracy. Both age groups manifested a contrast effect (lower JOLs for unrelated items when mixed with related items). Aging appears to spare monitoring of encoding, even though it adversely affects associative learning.

Studies of metacognition have focused on the degree to which individuals can accurately assess the state of their cognitive systems and whether they use such assessments to guide learning (Hacker, Dunlosky, & Graesser, 1998; Metcalfe & Shimamura, 1994; Reder, 1996). Metacognitive models of cognition typically conceptualize the cognitive system as involving separate monitoring and control functions (T. O. Nelson & Narens, 1990). Monitoring provides feedback to control systems about the status of processing and processing outcomes, enabling dynamic self-regulation of multiple aspects of learning (e.g., Thiede & Dunlosky, 1999). Poor utilization of monitoring to control learning can lead to lower levels of learning and long-term retention in older adults (Cavanaugh, 1989; Dunlosky & Connor, 1997; Murphy, Schmitt, Caruso, & Sanders, 1987).

In the present research, we evaluated age differences in how and how accurately individuals monitor encoding processes and outcomes when studying a list of paired associates. To investigate monitoring of encoding, researchers typically have individuals make a judgment of learning (JOL) immediately after studying each item. A JOL is a prediction about the likelihood of future successful retrieval, often scaled as a percentage of confidence that a studied item will later be recalled. In most current perspectives on metacognition, JOLs are treated as constructed judgments that are based on fallible access to cues about item retention at the time judgments are made (Schwartz, Benjamin, & Bjork, 1997). JOLs can also be influenced by a host of factors other than the contents of memory, including relevant task knowledge and beliefs about self-as-rememberer (Hertzog, Dixon, & Hultsch, 1990; Keleman, 2000).

Two techniques are prevalent for measuring the accuracy of JOLs: *relative accuracy* and *absolute accuracy*. For the former, each participant's JOLs are correlated with that individual's subsequent item recall by using Goodman–Kruskal gamma (γ) correlations (T. O. Nelson, 1984). These ordinal correlations measure the relative accuracy of people's judgments, ignoring discrepancies in magnitude or scaling of the JOLs with respect to recall outcomes. By contrast, discrepancies in scaling are central to absolute accuracy, which involves comparing the overall likelihood of predicted recall for a set of items (in this case, mean JOL for a participant) with actual level of recall (overall percentage correct). JOLs are inferential in nature and are based on access to indirect cues about item acquisition rather than direct access to memory traces (Dunlosky & Matvey, 2001; Koriat, 1997). Either type of JOL accuracy will suffer if individuals either base JOLs on cues that are not diagnostic of subsequent retrieval or fail to utilize cues that are diagnostic of subsequent retrieval (e.g., Benjamin, Bjork, & Schwartz, 1998).

Early research on metacognition and aging was based, in part, on the hypothesis that aging would impair metacognitive monitoring, which could in turn be an important cause of age differences in cognitive task performance (e.g., Perlmutter, 1978). It is well-known that older adults are normatively affected by losses in cognitive resources, including attention, working memory, and central executive functions (Craik & Jennings, 1992; Salthouse, 1991; Zacks, Hasher, & Li, 2000). To the extent that monitoring demands central resources (Ryan, Petty, & Wenzlaff, 1982), older adults may have difficulty monitoring learning when task performance is cognitively demanding (Bieman-Copland & Charness, 1994).

However, the literature on age differences in JOL accuracy indicates minimal effects of aging on the accuracy of metacogni-

---

Christopher Hertzog, Daniel P. Kidder, and Amy Powell-Moman, School of Psychology, Georgia Institute of Technology; John Dunlosky, Department of Psychology, University of North Carolina at Greensboro.

This research was supported by National Institute on Aging Grant R37 AG13148. We thank Laurie Wilson, Kirsten Zdravkovich, Kelly Colson, and Kristy Bowen for their assistance in recruiting and testing participants for the study and Jennifer Hertzog for assistance in preparation of this article.

Correspondence concerning this article should be addressed to Christopher Hertzog, School of Psychology, 274 5th Street, Georgia Institute of Technology, Atlanta, Georgia 30332–0170. E-mail: christopher.hertzog@psych.gatech.edu

tive monitoring (e.g., Dunlosky & Hertzog, 2000; Lovelace & Marsh, 1985; Shaw & Craik, 1989). For instance, Rabinowitz, Craik, Ackerman, and Hinchley (1982) had younger and older adults study paired associates that varied in the associative relationship of the paired stimuli (e.g., a *GRASS–COW* item is high in associative relatedness, whereas a *GRASS–AIRPLANE* item is low in relatedness). Both younger and older adults' JOLs were greater for related than for unrelated items. Moreover, both age groups showed similar absolute accuracy of their JOLs. Such findings suggest that monitoring, per se, is relatively spared by aging (Hertzog & Hultsch, 2000). Nevertheless, it may be possible to produce age differences in monitoring accuracy under a number of conditions that have yet to be fully explored empirically.

The literature on aging and JOLs has also been influenced by some methodological problems that may restrict the validity of generalizations about age effects (Hertzog & Hultsch, 2000). One problem is that many studies have relied on absolute accuracy as a measure of monitoring accuracy. Connor, Dunlosky, and Hertzog (1997), following recommendations in the experimental literature (e.g., T. O. Nelson, 1984), used relative accuracy (γ correlations) to assess monitoring of learning. Participants' JOLs were given as a percentage of confidence in the likelihood of subsequent recall (e.g., "I am 40% confident I will remember this item at test"). Connor et al.'s older adults and younger adults produced equivalent γs across three experiments, indicating no age-related loss in relative JOL accuracy (see also Dunlosky & Hertzog, 2000). By contrast, age differences in absolute accuracy of JOLs varied across experiments. In one experiment, older adults displayed significantly smaller differences between mean recall and mean JOLs. Connor et al.'s findings were therefore inconsistent with the argument that age differences in absolute accuracy reflect age-related deficits in metacognitive monitoring.

Nevertheless, a number of factors could in principle act to produce age differences in relative accuracy of JOLs, including experimental manipulations known to affect quality of encoding, such as divided attention (e.g., Anderson, Craik, & Naveh-Benjamin, 1998). Age-related effects in relative accuracy of JOLs could also arise when *multiple* diagnostic sources of information are available as cues for JOLs. The idea is simply that when individuals must monitor and combine multiple cues to optimize JOL accuracy, resource limitations during encoding may be more likely to reduce cue accessibility or to deter optimal cue utilization when making JOLs.

Koriat's (1997) cue-utilization framework provides a useful basis for thinking about the multiple cues potentially available for making JOLs. In Koriat's framework, the term *cue* refers to the type of information accessed by monitoring processes (not to experimenter-provided cues in a recall test). Koriat distinguished between intrinsic, extrinsic, and mnemonic cues. *Intrinsic* cues are stimulus attributes that influence JOLs independently of study activity, such as associative relatedness (as in Rabinowitz et al., 1982). *Extrinsic* cues involve characteristics of task and study activities, such as presentation rate, mediational strategies used during study (e.g., use of interactive imagery or generating sentences to form new associations; see Richardson, 1998), and test format. JOLs are substantially influenced by intrinsic cues but are often relatively unaffected by extrinsic cues (Lovelace, 1990). *Mnemonic* cues reflect aspects of the memory system that are consequences of encoding behavior and psychological states during and after study, which may pertain to a person's idiosyncratic encoding of individual items. In this case, one might be monitoring how fluently and effectively one had implemented a strategy, judging the likely consequences of this idiosyncratic processing for subsequent item memory. Evidence that JOLs are influenced by mnemonic cues above and beyond the influence of intrinsic or extrinsic cues is critical for metacognitive theories of self-regulation. Such evidence would verify that JOLs are influenced by on-line processing, as opposed to being based solely on declarative knowledge about self and memory.

Based on the cue-utilization framework, a key question becomes the following: Are older adults' JOLs as sensitive to different types of cues as younger adults' JOLs? As mentioned previously, Rabinowitz et al. (1982) found that both older and younger adults were sensitive to the intrinsic cue of item relatedness. Rabinowitz et al. also found that JOLs for both younger and older adults were not influenced by the extrinsic cue of instructed mediational strategies (*interactive imagery* instructions vs. *generic "learn"* instructions). However, individuals instructed to learn items are also likely to be using interactive imagery or sentence generation (Dunlosky & Hertzog, 1998a), so the lack of difference between conditions may not reflect insensitivity to extrinsic cues. Rabinowitz et al. did not explicitly evaluate whether reliance on relatedness as a cue diminished the influence of other cues. In particular, older adults may not have attended to differences in encoding processes for items within relatedness classes (e.g., for unrelated items), which could decrease JOL accuracy for these items. This hypothetical outcome represents an *overshadowing* effect (Price & Yates, 1993), in which one cue (relatedness) overshadows the use of other cues (e.g., kind of processing). In the present research, we investigated whether item relatedness produces overshadowing effects and whether they are more extensive for older than for younger adults.

The only relevant available data were reported by Connor et al. (1997), who manipulated item relatedness and reported JOL accuracy (a) for all items, in aggregate, as well as (b) separately for related and unrelated items. Connor et al. found significant γ correlations for the entire list and also for related and unrelated items. They also reported age invariance in γ for each of these measures. However, aspects of their method may have diminished any age-related effects. Connor et al. obtained both immediate and delayed JOLs in a within-subject design. Delayed JOLs are highly accurate because they are based in large part on retrieval of recently learned information held in long-term memory (T. O. Nelson & Dunlosky, 1991). As in the earlier research by T. O. Nelson and Dunlosky, Connor et al. randomly mixed delayed JOLs trials with immediate JOL trials to ensure that order of presentation was not confounded with type of JOL. Thus, the accuracy of immediate JOLs could have been enhanced by interspersing them with highly accurate monitoring of retrieval-based delayed JOLs. Participants may have learned about effectiveness of different mediational strategies and other aspects of encoding from the feedback provided by delayed JOLs' retrieval attempts. Moreover, to help obtain intermediate levels of recall for older adults, two thirds of the mixed list in Connor et al., Experiment 3, consisted of related items. This aspect of the design could have influenced JOL accuracy by making the list, in aggregate, easier to learn or by providing a more salient contrast between frequent, easier related items and less frequent, more-difficult unrelated items. In the present research, we studied immediate JOLs exclusively, and constructed lists with equal numbers of items from both classes.

Another goal was to isolate the source(s) of relative JOL accuracy for older and younger adults. A core issue here is whether the source of accuracy is from an individual's evaluation of observable characteristics of the study situation, from an individual's monitoring of his or her own idiosyncratic aspects of study, or both. To better understand this issue, consider an answer to the following question: Is JOL accuracy based on an ability to evaluate intrinsic and extrinsic cues rather than on access to the status of information recently learned? Apparently it is not, at least in some cases. Several early metacognitive studies demonstrated the phenomenon of *privileged access* by contrasting the correlation of an individual's own JOLs and recall outcomes with the correlation of another individual's JOLs when predicting outcomes of a different, randomly yoked individual. One's JOLs correlate more highly to one's own item recall outcomes than they do with another's recall outcomes (King, Zechmeister, & Shaughnessy, 1980; Lovelace, 1984; Rabinowitz et al., 1982), suggesting that individuals have privileged access to some diagnostic aspects of their own encoding experiences when making JOLs. Koriat (1997) also showed that privileged access effects increase across multiple study-test trials. He hypothesized that as learning proceeds, JOLs are increasingly influenced by idiosyncratic aspects of each person's learning (mnemonic cues) and decreasingly influenced by intrinsic cues. Thus, the relative accuracy of metacognitive judgments reflects more than mere sensitivity to manifestly observable item characteristics such as concreteness or relatedness. However, older adults may easily attend to intrinsic cues like relatedness that are manifest in the actual stimuli presented at study, but may suffer an age-related impairment in the ability to monitor internal states associated with mnemonic cues (such as variations in effectiveness of implementing a mediational strategy). If so, their JOLs could be dominated by intrinsic cues to a degree that would suppress privileged access to idiosyncratic aspects of encoding, leading to age differences in relative accuracy.

Overview of Experiments

An overarching goal of this research was to provide a more refined assessment of the hypothesis of age invariance in the accuracy of memory monitoring, building off the work of Connor et al. (1997). Although our focal measure of accuracy across experiments is relative JOL accuracy, we also present results that demonstrate the problems in interpreting age-related effects on absolute accuracy. More important, we manipulated relatedness in all three experiments to examine influences of item characteristics on JOL accuracy and JOL magnitude.

A major question is whether this intrinsic cue will overshadow variations in encoding behavior and other idiosyncratic aspects of encoding that could also influence JOLs. Overshadowing of this type would result in poor relative accuracy within relatedness classes, such as the subset of items that are unrelated. Koriat (1997) constructed paired-associate-item lists that were heterogeneous in "normative difficulty," with the principal source of that difficulty being the degree of relatedness between the stimulus and response words of paired-associate items. As is well-known, recall is greater for related items (see Kausler, 1994). Koriat's JOLs were, as expected, also influenced by relatedness (cf. Carroll, Nelson, & Kirwan, 1997; Dunlosky & Matvey, 2001; Rabinowitz et al., 1982), and hence γ of JOLs with item recall was relatively substantial (approximately .65 across several experiments). In contrast, JOL accuracy for more homogeneous lists of unrelated pairs is often much lower. For example, the median relative accuracy for immediate JOLs across seven conditions in work by T. O. Nelson and Dunlosky (1991; Dunlosky & Nelson, 1992) was approximately .20. This difference in outcomes is consistent with the hypothesis that much of the relative accuracy in Koriat's experiments is governed by sensitivity of JOLs to associative relatedness.

The question of interest for this study is whether older adults' JOLs are less influenced by cues other than relatedness when they are presented with a list that blocks on high and low levels of relatedness. Assume that older adults' JOLs are dominated by relatedness and do not reliably tap any other cues that vary within a homogeneous subset of items, whereas younger adults' JOLs are influenced by relatedness and variations in encoding for items within a subset. In this case, relative JOL accuracy computed across all items may show approximate age invariance (because both age groups are utilizing the highly diagnostic cue of relatedness), whereas age-related deficits will occur in relative JOL accuracy computed across items within a class (i.e., for unrelated items). We pursued this issue in all three experiments. As demonstrated across Experiments 1 and 2, the kind of rating scale used for JOLs moderates age-related effects on relative accuracy, with apparent overshadowing effects for older adults occurring when a discrete rating scale is used. In Experiment 3, a new experimental test of overshadowing by associative relatedness is introduced, and possible age-related deficits resulting from overshadowing effects are considered in detail.

Finally, in each experiment, we sought to discover the source(s) of relative JOL accuracy for both younger and older adults. This goal was accomplished by comparing the accuracy of an individual's JOLs with the accuracy of a different individual's JOLs when predicting another's recall performance. Outcomes were also relevant to possible age differences in how item relatedness is used as a basis for JOLs. If older adults use item relatedness as a basis for JOLs to the exclusion of all other cues (perhaps because of resource deficits), then an older individual's JOLs may be no more predictive of his or her recall performance than are the JOLs of another individual. That is, older adults may show little, if any, privileged access to idiosyncratic, mnemonic cues when making JOLs.

Experiment 1

Koriat (1997, Experiment 2) found that relative accuracy of JOLs increased over multiple study-test trials. He attributed this increase to accessibility of mnemonic cues about item encoding. In other words, after repeated opportunities to study the pairs, individuals gain more information about how their encoding of an item relates to the degree of learning for that item. A potential problem for Koriat's interpretation is that his data were collected on multiple study-test trials, in which a recall test was made at each trial. A critical distinction can be made between learning about item recall probability through monitoring cues during encoding (when the JOLs are made) versus monitoring performance at test. Performance monitoring is typically very accurate, whether measured by global postdictions (Connor et al., 1997; Devolder, Brigham, &

Pressley, 1990; Hertzog, Saylor, Fleece, & Dixon, 1994) or by item-level retrospective confidence judgments taken immediately after recall of each item (e.g., Dunlosky & Hertzog, 2000).

This issue is relevant to age differences in monitoring learning with and without test experience. Typically, older adults have been found to be equivalent to younger adults in improvements in the relative accuracy of global performance predictions across multiple trials (e.g., Hertzog et al., 1990, 1994) and to demonstrate excellent accuracy of global postdictions given immediately after recall task experience (Devolder et al., 1990; Hertzog et al., 1994). These improvements occur despite the absence of any kind of experimenter-provided performance feedback, suggesting good performance monitoring by older adults. It is unknown, however, whether older adults can improve the accuracy of their monitoring of learning without benefiting directly from performance experience.

The major goal of Experiment 1 was to evaluate whether varying amount of exposure to the word list would influence the relative accuracy of the JOLs and whether this effect would be due to improvements in accuracy of judgments within associative relatedness conditions, perhaps as a function of increased influence of mnemonic cues on JOLs. In Experiment 1, we randomly assigned participants to a standard single-study trial or a two-study trial (i.e., a prestudy phase prior to the standard study-test sequence; see Mazzoni & Nelson, 1995). In this prestudy condition, no recall test was given between the first and second study trials, and no JOLs were collected during prestudy. The item list included equal number of related and unrelated paired associates. This design makes it possible to test several hypotheses. First, if mnemonic cues have greater influence on JOLs after repeated study exposure (Koriat, 1997), then prestudy will increase relative accuracy within the relatedness conditions (i.e., within sets of related or unrelated items) in which the intrinsic cue of relatedness cannot benefit JOL accuracy. Second, if older adults are more reliant on the intrinsic cue of relatedness (relative to mnemonic cues), they should demonstrate lower accuracy of JOLs within relatedness conditions. Any initial age differences in relative accuracy within relatedness conditions could reflect either a monitoring deficit for mnemonic cues or a greater weighting of relatedness as a cue by older adults. If an initial age difference is caused by differential reliance on the intrinsic cue of relatedness instead of an age-related deficit in the ability to monitor mnemonic cues, then prestudy should increase the mnemonic cue influence for older adults and reduce age-related effects on relative accuracy.

Third, we hypothesized a disassociation of prestudy effects on relative and absolute JOL accuracy. Connor et al. (1997) conjectured that absolute accuracy reflects the disparity between actual recall and a subjective estimate anchored by an intuitively plausible guess, such as 50% correct. Anchoring of the judgments near the midpoint of the performance range may not be influenced by additional study time (Koriat, 1997). However, recall should increase with additional study time. Therefore, prestudy was hypothesized to reduce absolute accuracy at the same time it benefited relative accuracy.

Method

Participants

We tested 110 young and 76 old adults in Experiment 1. Extreme marginal values of recall or JOLs can distort measures of metacognitive accuracy. Therefore, those participants who recalled less than 5% correct or above 95% correct of the related or unrelated items were excluded from the analysis. Participants whose JOL responses fell into two or fewer of the six JOL categories were also excluded. Fifty-eight younger adults and 26 older adults were excluded. The high exclusion rate was primarily due to ceiling effects in the prestudy condition, especially for related items. After exclusion, there were 52 younger adults (24 in the prestudy condition and 28 in the no-prestudy condition) and 50 older adults (24 in the prestudy condition and 26 in the no-prestudy condition). There were no significant differences in demographic variables, including cognitive tests (see proceeding paragraphs) between participants who were included or excluded in the analyses.

The younger adults (age: $M = 20.4$ years, $SD = 2.6$) were undergraduates at Georgia Institute of Technology who received extra credit in a psychology course for their participation. The older participants (age: $M = 68.8$ years, $SD = 6.4$) were community-dwelling adults, recruited from a list of interested alumni of Georgia Institute of Technology and from advertisements placed in local magazines and newspapers. The older adults received $15 for their participation.

The older adults were relatively well educated ($M = 14.9$ years, $SD = 2.9$), rated themselves as being in good health ($M = 1.7$, $SD = 0.7$; on a scale of 1 [*excellent*] to 4 [*poor*]), perceived themselves as having good health relative to same-aged peers ($M = 1.5$, $SD = 0.8$; on a scale of 1 [*better*] to 3 [*worse*]), and were taking, on average, 1.5 medications ($SD = 1.4$).

Participants were given a modified version of the Educational Testing Service Advanced Vocabulary Test—V-4 (Ekstrom, French, Harman, & Dermen, 1976; see Hertzog, 1989); the Computation Span Test (Salthouse & Babcock, 1991), a measure of working memory span; and the Letter Comparison and Pattern Comparison Tests (Salthouse, 1996), two measures of perceptual speed. Older adults performed significantly better than younger adults on the Advanced Vocabulary test (old: $M = 23.4$, $SD = 6.5$; young: $M = 15.5$, $SD = 4.5$), $t(96) = 7.30$, $p < .001$. Younger adults performed better on the Computation Span task (young: $M = 6.1$, $SD = 1.6$; old: $M = 4.4$, $SD = 2.5$), $t(93) = 4.01$, $p < .001$; Letter Comparison (young: $M = 26.3$, $SD = 5.0$; old: $M = 17.8$, $SD = 3.4$), $t(100) = 10.04$, $p < .001$; and Pattern Comparison (young: $M = 44.6$, $SD = 6.7$; old: $M = 30.6$, $SD = 5.4$), $t(100) = 11.54$, $p < .001$. Thus, the participants showed typical patterns of age differences in cognition (Kausler, 1994; Salthouse, 1991).

Apparatus and Materials

The paired-associate learning task was programed in HyperCard and run on Macintosh computers. Paired-associate items were divided equally into pairs that were either high or low in degree of intrinsic association (e.g., *SUGAR–COFFEE* vs. *SALT–MAYOR*). The 66 paired associates used in the task were taken from two sources. Unrelated pairs were drawn from Paivio, Yuille, and Madigan (1968) word norms; associatively related pairs were a subset of the items used by Connor et al. (1997). The stimuli are provided in the Appendix.

Design and Procedure

The experimental design was a 2 × 2 × 2 factorial with age (young vs. old) and prestudy (no prestudy vs. prestudy) as the between-subject variables and relatedness (high vs. low) as the within-subject variable. Participants were tested in groups of 1 to 3 persons in Christopher Hertzog's laboratory at the Georgia Institute of Technology, with sessions lasting between 60 and 90 min. After giving informed consent, participants completed a brief demographic questionnaire and then performed the computerized experimental task. Following the experimental task, they completed a debriefing questionnaire about their experience and were then administered four psychometric tests in the following order: Letter Comparisons, Pattern Comparisons, the Advanced Vocabulary test, and the Computation Span task.

The paired-associate task involved presentation of each word pair, centrally located on the computer screen, for a period of 10 s. The JOL screen immediately followed the presentation of each word pair. The stimulus word of each pair was presented (e.g., *DOG–???*), along with the following instructions: "How confident are you that in about ten minutes from now you will be able to recall the second word of the item when prompted with the first word?" They were given six discrete response options: 0% (*definitely won't recall*); 20% (*sure*); 40%, 60%, 80%, and 100% (*definitely will recall*). Participants made a JOL by typing the chosen percentage, and the program accepted only the responses of 0, 20, 40, 60, 80, or 100. After the JOL had been entered, the next word pair was immediately displayed for study.

Six of the pairs were designated as practice items, and the remaining 60 pairs were experimental items. For both practice and experimental items, half the paired associates were related and half were unrelated. The order of item presentation during practice or during the actual experiment was randomized. At test, the items were presented in a different randomized order. Individuals indicated their response by typing the response word when prompted. Presentation of the items at test was participant paced. To minimize the role of misspelling, a response was correct if the first three letters of the typed word matched the response word.

The task began after individuals sat in front of the computer, where they read, under their own pacing, a set of task instructions. They then practiced studying word pairs and making JOLs for six items, followed by a practice recall test. After practice, they were told that the main experiment was about to begin, and that they would be studying 60 word pairs just like those seen during practice. They were then asked to predict their future recall performance of these 60 items by entering the percentage of items they would recall. Participants typed in an integer within the range of 0% to 100%. After the participants had finished studying all 60 items, they were again asked to make a global prediction, using the same instructions as before study. Following the recall test, participants were asked to provide a postdiction (i.e., a numeric estimate) of the percentage of all items they had successfully recalled. We do not report outcomes involving the global predictions and postdictions in this article.

The prestudy condition provided participants with an additional study trial. Following instructions and practice trials, the items were presented for 10 s. No JOLs were collected during prestudy. The order of item presentation was randomized separately for prestudy and for the regular study-test trial. Individuals in the prestudy condition also made one additional global prediction prior to prestudy.

In this study we used several quantitative indices of various aspects of accuracy of the JOLs as predictions of future paired-associate recall. In all cases, these indices were calculated separately for each individual participant (or yoked pairs of participants) and were then analyzed as dependent variables. Relative accuracy of JOLs was operationalized as an intraindividual γ correlation between JOLs and subsequent recall performance. The γ possesses the desirable qualities of being formally independent from levels of recall and having a straightforward probabilistic interpretation (T. O. Nelson, 1984). Specifically, it can be interpreted as the conditional probability that recall of Item A is higher than Item B, given that Item A was given a higher JOL than Item B.

Two indices of absolute accuracy were measured: *simple* and *absolute* differences between participants' mean JOL and mean recall in an experimental condition. The simple difference between individuals' mean JOL and mean recall measures average overconfidence and underconfidence. The second index assesses absolute discrepancy between mean JOL and recall, irrespective of whether the JOLs overestimate or underestimate recall. It is also the index of absolute accuracy recommended by some scientists (e.g., Devolder et al., 1990). In a number of studies, these two indices have led to similar conclusions (e.g., Connor et al., 1997; Hertzog et al., 1994) regarding age differences in absolute accuracy of metacognitive judgments.

The privileged access analyses were conducted as follows. For a given group, a participant was selected and yoked with another participant chosen randomly from the set of participants who had not already been yoked. This procedure continued until all participants had been yoked, or in the case of an odd number of participants in the initial group, until only one unyoked participant remained. For each dyad, the participants' records were matched so that the JOLs and recalls for the set of items from one participant were linked to the JOL and recall for the corresponding items of the yoked participant. Let us arbitrarily label the first member of a yoked pair *Person A* and the second member of the pair *Person B.* Two γ correlations were then computed for all items in each experimental condition: one between Person A's JOLs and Person B's item recalls; the other between Person A's item recalls and Person B's JOLs. These two correlations were then averaged over the sets of pairs, and a pooled standard error of estimate was computed by using the standard errors for the two half-sets. To estimate consistency in JOLs across items, a third γ correlation between JOLs for Person A and Person B in a yoked pair was also computed.

Results

Recall

Table 1 reports the means and standard deviations for paired-associate recall.[1] As expected, younger adults correctly recalled more response words. This age effect and the significant main effect for relatedness ($p < .001$) were qualified by an expected Age × Relatedness interaction, $F(1, 98) = 20.91$, $MSE = 4.93$, $p < .001$. The overall age differences were smaller for highly related pairs (differences of .08 for related stimuli [.82 vs. .74] compared with .21 for unrelated stimuli [.57 vs. .36]). Scaled in standard deviation units, using Cohen's (1988) d, the age differences were approximately $d = 0.6$ for related word pairs and $d = 1.0$ for unrelated pairs. Expressed as sensitivity to relatedness, the proportion recall difference between related and unrelated pairs was .25 and .38 for younger and older adults, respectively.

As expected, prestudy increased paired-associate recall, $F(1, 98) = 7.02$, $MSE = 0.31$, $p < .01$, despite exclusion of individuals with near ceiling performance. The marginal mean difference in proportion of correct recall between the prestudy and control conditions was .08, $d = 0.6$. No other effects were statistically reliable.

JOLs

Table 2 reports the mean and standard deviations of JOLs, given as a percentage of confidence (ranging from 0 to 100) in the four conditions. In contrast to the recall data, mean JOLs were unaffected by prestudy (unweighted marginal means of 51% confidence for both conditions). Both age groups demonstrated sensitivity of JOLs to relatedness, and a significant interaction showed that this effect was greater for older adults, $F(1, 98) = 10.01$, $MSE = 0.06$, $p < .01$. The marginal mean difference between

---

[1] As noted in the *Method* sections, all analyses in the reported experiments are based on a reduced sample meeting inclusion criteria for recall and JOL distributions. Exclusion of participants had little effect on means for any variable, including recall.

Table 1
Paired-Associate Recall by Age Group and Experimental Condition for Experiments 1 and 2

Age group and condition	Associative relatedness					
	Mean recall			Standard deviation		
	Unrelated	Related	Overall	Unrelated	Related	Overall
			Experiment 1			
Young	.57	.82	.70	.21	.12	.15
Prestudy	.61	.83		.23	.12	
No prestudy	.54	.81		.19	.12	
Old	.36	.74	.55	.19	.15	.16
Prestudy	.43	.79		.21	.11	
No prestudy	.30	.70		.15	.16	
			Experiment 2			
Young	.50	.81	.65	.22	.12	.16
Old	.26	.65	.46	.17	.15	.14

related and unrelated pairs was 39% for older adults and 30% for younger adults.

JOL Accuracy

Absolute accuracy. Each person's mean JOL was scaled as a percentage of confidence (as in Table 2), and difference scores (mean percentage of JOL minus mean percentage of recall) were computed for each condition to get measures of over- and underconfidence. Older adults' mean JOLs were closer to their actual levels of mean recall (differences of −2% for related and −3% for unrelated pairs) than was the case for younger adults (differences of −16% for related and −21% for unrelated pairs). The main effect of age was significant, $F(1, 98) = 20.18$, $MSE = 1.34$, $p < .001$. On average, younger adults underestimated their recall for both types of items, whereas older adults were relatively accurate. In fact, the differences in recall for related and unrelated pairs were quite similar in magnitude to the differences in JOLs. Thus, the age difference could be characterized as greater underconfidence on the part of younger adults. The Age × Relatedness interaction was not reliable ($F < 1$).

The hypothesis of reduced absolute accuracy after additional study was confirmed. On average, absolute accuracy was lower in the prestudy condition (−15%) relative to the control condition (−7%), $F(1, 98) = 4.95$, $MSE = 0.33$, $p < .05$. No other effects were statistically reliable.

Absolute difference measures also showed a main effect for age, $F(1, 98) = 19.84$, $MSE = 0.03$, $p < .001$, with older adults having superior accuracy (mean absolute deviation = 15%) relative to younger adults (M = 25%). The effect was qualified by an Age × Prestudy interaction, $F(1, 98) = 4.10$, $MSE = 0.03$, $p < .05$. Age differences in absolute accuracy were greater after prestudy (M_{diff} = 14%) than before it (M_{diff} = 5%). Absolute differences

Table 2
Judgments of Learning Ratings by Age Group and Experimental Condition for Experiments 1 and 2

Age group and condition	Associative relatedness					
	Mean rating			Standard deviation		
	Unrelated	Related	Overall	Unrelated	Related	Overall
			Experiment 1			
Young	36	66	51	17	15	14
Prestudy	36	61	49	18	16	16
No prestudy	35	70	52	15	14	12
Old	33	72	52	18	16	15
Prestudy	31	77	54	17	13	13
No prestudy	35	67	50	20	16	16
			Experiment 2			
Young	41	69	55	15	15	13
Old	36	70	53	18	17	15

Table 3
Relative JOL Accuracy (γ Correlations) by Age Group and Experimental Condition for Experiments 1 and 2

Age group and condition	Mean accuracy: Unrelated	Mean accuracy: Related	Mean accuracy: Overall	Standard deviation: Unrelated	Standard deviation: Related	Standard deviation: Overall
Experiment 1						
Young	.38	.29	.47	.32	.41	.22
Prestudy	.45	.40	.51	.31	.31	.21
No prestudy	.33	.19	.43	.32	.46	.23
Old	.31	.08	.51	.34	.50	.20
Prestudy	.27	.20	.54	.35	.51	.16
No prestudy	.35	−.03	.48	.33	.48	.23
Experiment 2						
Young	.20	.08	.37	.27	.35	.20
Old	.23	.14	.46	.35	.24	.24

Associative relatedness

Note. JOL = judgment of learning.

were also reliably worse for unrelated pairs (M = 23%) compared with related pairs (M = 17%), $F(1, 98) = 8.10$, $MSE = 0.01$, $p < .01$. No associated interaction effects were reliable.

Thus, older adults had better absolute JOL accuracy for both indices, and were less likely than younger adults to have absolute accuracy impaired by the prestudy trial.

Relative accuracy. One might conclude from the age differences in absolute accuracy that older adults are better at monitoring learning during study than younger adults. However, the most appropriate statistic for assessing monitoring accuracy is the γ correlation between an individual's JOLs and recall, reported in Table 3.[2]

The aggregate γ, ignoring the relatedness manipulation, differed reliably from zero, $p < .001$, for younger and older adults in both prestudy conditions. Age was not reliably related to the aggregate γs (marginal means of .47 and .51 for younger and older adults, respectively; $F < 1$). However, when disaggregated by relatedness, an interesting pattern in relative accuracy emerged. Relatedness had a significant effect on relative accuracy, $F(1, 98) = 8.09$, $MSE = 0.50$, $p < .01$, with accuracy being greater for unrelated items than for related items (marginal mean γ of .35 vs. .19, respectively). The relatedness effect tended to be smaller with prestudy, but the associated interaction was not statistically reliable, $F(1, 98) = 3.13$, $MSE = 0.50$, $p > .05$. When γ was computed separately for the related and unrelated conditions, however, there was an overall age main effect, $F(1, 98) = 7.28$, $MSE = 1.05$, $p < .01$. Younger adults had, on average, higher within-condition relative accuracy (M = .34 vs. .20 for older adults). Age did not interact with the other two experimental variables.

Thus, in contrast to absolute accuracy, older adults were, on average, worse in relative accuracy. However, prestudy did increase relative accuracy, as hypothesized (Table 3). For the aggregate data (ignoring relatedness), prestudy increased γ by .07 on average, although this effect just missed significance, $F(1, 98) = 3.06$, $p > .05$. When data were disaggregated into related and unrelated items, the effect of prestudy on relative accuracy was reliable, $F(1, 98) = 5.07$, $MSE = 0.73$, $p < .05$. The marginal mean γ after prestudy was .33, relative to a mean γ of .21 without prestudy. The pattern was observed for both age groups, and suggested that both younger and older adults benefited from mnemonic cues whose influence on the JOLs was enhanced by additional study.

Privileged access. To examine whether the aggregate γ and within-relatedness γ correlations reflected the impact of monitoring idiosyncratic aspects of encoding, we randomly yoked pairs of individuals within the same age groups (see *Method* sections for details). To the extent that relative accuracy is based only on normative information about item difficulty, then the yoked γ should be of equal magnitude to the standard γ already reported.

Table 4 presents the mean γ and the mean yoked γ for JOLs and recall. Three rows of data are reported for each Age × Prestudy cell: one aggregating over all items (ignoring relatedness), one for unrelated items, and a third for related items. A given row reports three γ correlations: the γ for a person's JOLs with that same person's recall, the γ of a person's JOLs with another's yoked recall, and the mean γ between the JOLs for the two yoked persons. The difference between the first two γ correlations measures privileged access, whereas the latter γ assesses the degree to which JOLs are entrained by item characteristics, especially the intrinsic cue of associative relatedness.

For the aggregate data in the no-prestudy condition, the yoked γ was equivalent to the actual γ for older adults, $t(24) = 0.10$, $p > .10$. The actual γ was greater than yoked γ for the younger adults, although the difference was not reliable, $t(24) = 1.50$, $p > .05$. However, prestudy increased the discrepancy between yoked and

---

[2] Exclusion of participants because of poor recall and JOL distributions reduced variance and standard errors for γ, increasing statistical power.

Table 4

Privileged Access Effects: γ Correlations for JOL–Recall (Self), Yoked JOL–Recall, and Yoked JOL–JOL Pairings by Age Group and Experimental Condition for Experiments 1 and 2

Age group and condition	JOL–Recall	Yoked JOL–Recall	Yoked JOL–JOL
Experiment 1			
Young			
No prestudy			
Overall	.43	.29	.44
Unrelated	.34	.01	.22
Related	.17	−.08	−.10
Prestudy			
Overall	.52	.26	.29
Unrelated	.44	.17	.19
Related	.41	−.06	.01
Old			
No prestudy			
Overall	.46	.45	.50
Unrelated	.30	.18	.31
Related	.00	−.05	.01
Prestudy			
Overall	.54	.47	.64
Unrelated	.27	.10	.26
Related	.20	.04	.07
Experiment 2			
Young			
Overall	.36	.25	.25
Unrelated	.20	−.01	.01
Related	.07	.02	.03
Old			
Overall	.48	.43	.45
Unrelated	.27	.17	.22
Related	.11	.00	.03

Note. JOL = judgment of learning.

standard γ for younger adults, resulting in a significant difference, $t(22) = 2.54$, $p < .05$. It also resulted in a slight elevation of standard γ, relative to yoked γ, for older adults, although there was still no significant privileged access effect, $t(22) = 1.15$, $p > .10$. The JOL–JOL γ correlations suggest an explanation of why older adults' actual γ and yoked γ were similar. After prestudy, older adults' JOLs were more highly correlated with JOLs of randomly yoked pairs than were younger adults' JOLs, $t(48) = 4.36$, $p < .01$. This pattern is further evidence for the greater sensitivity of older adults' JOLs to associative relatedness. Moreover, these data suggest that the relatively high aggregate γ correlations for older adults are predominantly a function of the influence of relatedness on their JOLs.

Within relatedness conditions, younger adults showed significant privileged access effects for unrelated items without prestudy, $t(24) = 2.68$, $p < .05$, and with prestudy, $t(22) = 1.77$, $p < .05$ (one-tailed). Moreover, after prestudy younger adults showed a privileged access effect for related items as well, $t(22) = 3.16$, $p < .01$. Although both age groups showed trends for larger differences between actual γ and yoked γ after prestudy, in no case did older adults demonstrate a statistically robust privileged access effect. After prestudy, older adults' difference between actual γ and yoked γ for unrelated items was still not reliable, $t(24) = 1.24$, $p > .10$. Thus, prestudy increased privileged access effects in the sample data for both age groups, but this effect was statistically robust only for younger adults.

Discussion

Experiment 1 replicated previous work showing that associative relatedness has a potent influence on JOLs, and that use of relatedness as a cue results in substantial relative accuracy for both age groups. The γ correlations for the aggregated data, ignoring relatedness, were approximately .5 for both age groups, well above chance. Both younger and older adults' JOLs were sensitive to associative relatedness. Indeed, older adults appeared to be more sensitive to the relatedness manipulation, at the cost of relative accuracy of JOLs within relatedness conditions. The yoking analyses showed greater consistency between randomly paired older adults' JOLs than younger adults' JOLs aggregated across the entire list. This heightened sensitivity appeared to elevate the relative accuracy of the JOLs in the aggregate, producing equivalent relative accuracy for older and younger adults. When separated into relatedness conditions, however, the younger adults had higher relative accuracy. Moreover, younger adults showed evidence of robust privileged access, whereas older adults did not. Collectively, these outcomes suggest that older adults' JOLs were more strongly influenced by normative item relatedness and less by idiosyncratic aspects of encoding different word pairs. They also suggest a possible overshadowing of mnemonic cues by associative relatedness for older adults' JOLs.

Certainly, such outcomes are consistent with the hypothesis that older adults' resource limitations act to limit their ability to attend to multiple sources of information when making JOLs (e.g., Bieman-Copland & Charness, 1994). However, other aspects of the data suggest that older adults are not generically constrained by any such limitations. Additional study exposure with the list increased older adults' relative accuracy within relatedness conditions and increased the difference between actual and yoked relative accuracy, although older adults still did not show a robust privileged access effect. These trends suggest that the differential reliance on relatedness information by older adults when making JOLs in a single study-test trial may not reflect a generic age deficit in metacognitive monitoring ability. An alternative hypothesis is that the differential sensitivity to the relatedness manipulation by older adults is a judgmental phenomenon and, as such, involves an implicit or explicit discounting of cues other than associative relatedness that can be altered by instructions, incentives, or other manipulations.

The results of Experiment 1 were consistent with many earlier studies in showing that younger adults may be underconfident, whereas older adults are overconfident of their paired-associate learning abilities (e.g., Bruce, Coyne, & Botwinick, 1982). However, older adults' absolute JOL accuracy was superior to younger adults' accuracy. Consistent with Connor et al. (1997), the age group with mean performance nearest the midpoint of the performance range (in this case, older adults) had the best absolute accuracy. Moreover, absolute accuracy degraded further for younger adults as prestudy moved their recall levels closer to ceiling, without affecting their JOLs. Younger adults were even more underconfident in the prestudy condition. Thus, Experiment 1 reinforced concerns about interpreting age differences in absolute accuracy.

Experiment 2

Experiment 1 suggested that older adults were not as effective in utilizing cues other than relatedness to produce accurate JOLs. Although this pattern could indicate a resource limitation for older adults, Experiment 2 was designed to evaluate the alternative hypothesis that the age difference was an artifact of combining a salient intrinsic cue (relatedness) with a limited number of JOL response options. That is, in Experiment 2 we evaluated whether age differences in relative accuracy within relatedness conditions reflected an age difference in the ability to monitor learning or, alternatively, an unintended consequence of the use of a discrete JOL rating scale when simultaneously manipulating item relatedness. In Experiment 2 a continuous JOL scale was used. Participants entered a number from 0 to 100 to indicate their relative confidence in the likelihood of future recall.

Why would a continuous JOL rating scale possibly affect age differences in judgment accuracy? This hypothesis is based on the assumption that relatedness is a highly salient, intrinsic cue that is easily and rapidly accessed during study and has a potent impact on anchoring JOL ratings. We assume, consistent with the robust effects of relatedness on JOLs in Experiment 1 (and in other studies), that related items will normatively be regarded as having a high probability of recall, whereas unrelated items would normatively be regarded as having a low probability of recall. JOLs require conversion of this subjective ordinal difference to a scaling of subjective confidence. If raters associate the midpoint of the rating scale (50% confidence) with subjective neutrality (Connor et al., 1997), then related items would typically be assigned greater than 50% confidence, whereas unrelated items would be assigned less than 50% confidence. This partition of the six-level discrete JOL scale leaves few degrees of freedom for ratings to vary within the relatedness conditions (say, 0%, 20%, or 40% for an unrelated item; 60%, 80%, or 100% for a related item). Combined with an aversion to use the extremes of the rating scale (i.e., 0 and 100% ratings; see Dunlosky & Nelson, 1994), individuals could be often functionally limited to using two rating scale values to express variations in subjective confidence because of cues other than relatedness (e.g., JOL ratings of 60% and 80% for related items). Such scaling behavior could suppress discrete JOL variability and within-condition relative accuracy. Older adults may have been less effective at discriminating between other cues that influence recall of related and unrelated items. However, the scaling hypothesis suggests they may have had equivalent cue discrimination but deficient translation of these cues into graded JOL ratings, given the limited range of discrete JOL response options. If the age differences in the first experiment were an artifact of the discrete JOL rating scale, then the switch to a continuous JOL scale would eliminate the age differences in relative accuracy within levels of relatedness.

Use of the continuous rating scale could also affect the age differences in absolute accuracy. The continuous JOL scale is probably more appropriate for assessing absolute accuracy than is a discrete, ordinal scale. The underconfidence of younger adults might reflect the limited opportunity to select rating scale values above 60% confidence (i.e., only the 80% and 100% options were available). Given that younger adults' mean item recall was about 70% in the single study-test trial condition of Experiment 1, their apparent underconfidence could be an artifact of the discrete JOL rating scale.

Method

Participants

Fifty-two young adults and 43 old adults, recruited using the same techniques already described, initially participated. Fourteen younger and 9 older adults were excluded because of ceiling or floor recall, and 1 younger and 3 older adults were excluded because of insufficient variability in JOLs within relatedness conditions. The resulting sample consisted of 37 younger adults (age: $M = 19.8$ years, $SD = 1.6$) and 31 older adults (age $M = 69.5$ years, $SD = 5.2$). Included and excluded participants did not differ on demographic and cognitive variables.

The older adults were relatively well educated ($M = 15.6$ years, $SD = 2.3$), rated themselves as being in good health ($M = 1.8$, $SD = 0.7$), perceived themselves as having good health relative to same-aged peers ($M = 1.3$, $SD = 0.7$), and were taking, on average, 2.4 medications ($SD = 2.3$). Older adults performed significantly better than the younger adults on the advanced vocabulary test (old: $M = 23.7$, $SD = 8.5$; young: $M = 15.8$, $SD = 4.4$), $t(60) = 4.70$, $p < .001$. Younger adults performed better on the computation span task (young: $M = 6.0$, $SD = 1.8$; old: $M = 4.0$, $SD = 2.5$), $t(65) = 3.72$, $p < .001$; letter comparison (young: $M = 24.5$, $SD = 4.7$; old: $M = 16.8$, $SD = 3.9$), $t(64) = 7.18$, $p < .001$; and pattern comparison (young: $M = 41.2$, $SD = 6.6$; old: $M = 30.3$, $SD = 5.3$), $t(66) = 7.37$, $p < .001$.

Design and Procedure

The design was identical to the no-prestudy condition of Experiment 1, but with the use of continuous rather than discrete JOLs. When prompted after study to make a judgment, participants entered a number between 0 and 100 to indicate their percentage of confidence that they would recall the item approximately 30 min later.

Results

Recall

Mean recall was similar to Experiment 1 (see Table 1), producing a similar pattern of significant effects. There were robust main effects of age and relatedness ($p < .001$), qualified by a significant Age × Relatedness interaction, $F(1, 66) = 4.23$, $MSE = 0.08$, $p < .05$. However, the interaction effect was somewhat weaker in Experiment 2. The relatedness effect for older adults was .39, compared with an effect of .31 for younger adults. Effect sizes for age were actually similar for the two levels of relatedness ($d = 1.19$ and $d = 1.23$ for related and unrelated items, respectively).

JOLs

As reported in Table 2, the two age groups did not differ in mean JOLs (marginal mean difference = 55% for younger adults and 53% for older adults; $F < 1$), despite robust differences in subsequent recall. Mean JOLs again were sensitive to associative relatedness (marginal mean differences of 28% for younger adults and

34% for older adults). As had been the case for discrete JOLs, the sensitivity of continuous JOLs tended to be greater for older adults. However, there was no evidence of an Age × Relatedness interaction in JOLs ($F < 1$).

JOL Accuracy

Absolute accuracy. Relative to the comparable conditions for younger adults in Experiment 1, continuous JOLs appeared to produce somewhat less underestimation of performance by younger adults in Experiment 2 (see Table 2). Younger adults' mean JOLs underestimated paired-associate recall (marginal mean difference of -10%). The null hypothesis of perfect absolute accuracy (difference $= 0\%$) was rejected, $t(36) = 3.57$, $p < .001$. Older adults slightly overestimated performance (marginal mean difference of $+8\%$), $t(30) = 2.23$, $p < .05$. The age difference in absolute accuracy was statistically reliable, $F(1, 66) = 15.34$, $MSE = 1.00$, $p < .001$. Thus, the general pattern of overconfidence for older adults and underconfidence for younger adults was not affected by using continuous JOLs. Unlike Experiment 1, younger adults were not more inaccurate, in a global sense. In fact, when absolute differences between JOLs and recall were computed, older and younger adults had similar absolute accuracy ($M = 15\%$ vs. $M = 16\%$, respectively; $F < 1$).

Relative accuracy. Older adults tended to have somewhat higher aggregate γ correlations (Table 3), although this difference was not reliable, $t(66) = 1.69$, $p > .10$. Unlike Experiment 1, however, age differences in relative accuracy were negligible within the relatedness conditions (marginal mean γs of .14 and .19 for younger and older adults, $F < 1$, $d = 0.15$). Indeed, in contrast to the first experiment, mean γ was actually higher for older adults than for younger adults. The trend for a relatedness effect on relative accuracy was still evident in both groups, but was not statistically reliable, $F(1, 65) = 2.98$, $MSE = 0.30$, $p < .10$. Unrelated items produced a mean γ of .22, whereas related items yielded a mean γ of .11 ($d = 0.20$). For both age groups, γ correlations for unrelated items were reliably greater than zero ($p < .01$).

Privileged access. As can be seen in Table 4, both age groups showed evidence of privileged access. Older adults showed a significantly larger yoked γ in aggregate, $t(28) = 1.80$, $p < .05$ (one-tailed), and this effect was again attributable in large part to greater consistency in the JOLs between randomly paired older respondents, $t(28) = 3.45$, $p < .01$. Thus, older adults' JOLs were more entrained by associative relatedness. Unlike the first experiment, however, older adults in Experiment 2 showed evidence of privileged access within the two relatedness conditions. Yoked γ was not significantly different from zero for either related or unrelated items for both age groups.

Discussion

The results of the present experiment suggest that the age differences in relative accuracy within relatedness conditions observed in Experiment 1 were a function of combining the relatedness manipulation with the use of discrete JOLs. To the extent that individuals' attend to the relatively accessible cue of associative relatedness, they may tend to emphasize an ordinal (high vs. low) distinction in the likelihood of future recall between related and unrelated items. Such an initial parsing of the rating scale leaves few degrees of freedom to vary JOLs according to other cues. The present results suggest that older adults can more effectively use a continuous JOL rating scale to reflect other cues (i.e., without sacrificing relative accuracy) than a discrete JOL rating scale.

The switch to the continuous rating scale eliminated the Age × Relatedness interaction for JOLs found in Experiment 1. However, the continuous rating scale did not eliminate the dominant contribution of relatedness to older adults' relative JOL accuracy. Older adults' yoked γs, in aggregate, were nearly as large as the standard γ correlations obtained by using their own JOLs. On the other hand, the continuous JOL rating scale also resulted in improved absolute JOL accuracy for younger adults. Older and younger adults had roughly equivalent levels of absolute accuracy, with younger adults showing substantial underconfidence in JOLs and older adults showing overconfidence. Thus, age differences in absolute accuracy that had favored older adults in Experiment 1 were eliminated by the use of a continuous JOL rating scale.

Experiment 3

Experiment 3 was designed to provide a more definitive evaluation of three phenomena identified in Experiments 1 and 2: (a) greater sensitivity of older adults' JOLs to associative relatedness (found in Experiment 1, but not in Experiment 2); (b) higher relative accuracy of JOLs for unrelated items; and (c) overshadowing of mnemonic cues by associative relatedness (suggested in outcomes for older adults in Experiment 1). We also wanted to examine the possibility of a fourth phenomenon—a contrast effect for JOLs. In this context, a contrast effect would be obtained if unrelated items receive lower JOLs when paired with related items than when paired with other unrelated items.

A problem in memory research is that experimental effects that depend on manipulating item properties within a list are always potentially confounded with item variation on other unintended stimulus characteristics that may covary with the properties of interest. The interpretation of such effects is therefore critically dependent on the assumption that the item differences occur only because of the intended manipulation. This problem is specifically relevant to the present study. The difference in relative accuracy for unrelated versus related items could be attributed to other variables. In particular, related and unrelated items differed in the stimulus words used to prompt JOLs and to prompt recall at test. Likewise, the apparent overshadowing of mnemonic cues by an intrinsic cue for older adults in Experiment 1 could be an artifact of the specific set of items used.

With this problem in mind, Experiment 3 was designed to provide tighter control for such differences. We constructed two paired-associate lists that better isolated the relatedness effect and created a better opportunity to observe overshadowing. One list consisted entirely of unrelated items (the unrelated list). The second list (the mixed list) was divided into two halves. The first half-set of items was identical to half of the items from the unrelated list. We refer to this set of items as *Set 1.* The second half-set used identical stimulus words to the second half of the unrelated list. However, it paired these stimuli with related rather than unrelated response words. We refer to this matched set of items as *Set 2.* Participants in the experiment were randomly assigned to a single study-test trial with one of the two lists.

The chief advantage of this design is that participants receiving either the unrelated list or the mixed list experience exactly the same prompts for JOLs and recall—the stimulus words are identical between the two lists. Hence, a between-subjects comparison of related and unrelated-item JOLs and JOL accuracy can be made when the stimulus words are identical for participants who receive either related or unrelated items. Another important feature is that both groups of participants receive an identical set of unrelated items (Set 1). The design, therefore, affords more definitive tests of the relatedness effects. If relative accuracy is truly lower for related items, then γ will differ not only between related and unrelated items of the mixed list, but also between different lists of Set 2 that share the same stimulus words but are different in whether they contain either related or unrelated response words. Similarly, if older adults weight relatedness more than younger adults when making JOLs, then they will show a larger difference between unrelated and related items for the Set 2 paired-associate items, even though all individuals are prompted with identical stimuli.

The design also affords a more definitive test of overshadowing. To infer that overshadowing by relatedness occurs, we must show that (a) unrelated items in Set 1 show above-chance relative accuracy when not paired with related items but that (b) the same Set 1 items have lower relative accuracy of JOLs when paired with other unrelated items. If associative relatedness overshadows other mnemonic cues when making JOLs, then relative accuracy should be greater for Set 1 in the unrelated list than for Set 1 in the mixed list.

The design also makes it possible to test for a contrast effect that would due to associative relatedness. Contrasting unrelated items to related items could lower JOLs for the unrelated items, relative to the JOLs that would have been obtained had these items been paired with other unrelated items. That is, when explicitly contrasted with related items, unrelated items may produce lower confidence in the likelihood of item recall (see Carroll & Nelson, 1993, for a similar effect on feeling-of-knowing judgments). A contrast effect would be inferred if mean JOLs are higher for Set 1 in the unrelated list than for Set 1 in the mixed list. Such contrast effects would indirectly reflect the salience of relatedness as a cue. Given our earlier results on disassociations of absolute and relative accuracy, one can anticipate obtaining a contrast effect for mean JOLs without it being accompanied by an overshadowing effect on relative accuracy for those same items.

Given that older adults' JOLs seem to be strongly influenced by relatedness, older adults may show both a greater contrast effect on unrelated JOLs and also a stronger overshadowing effect. Such an outcome would help attribute any observed age difference in relative accuracy to differential access to multiple dimensions of cues for making JOLs. Conversely, the effects in Experiment 1 may have been an artifact of combining discrete JOLs with a relatedness manipulation, as suggested by Experiment 2. If so, then one might expect to observe age-related invariance in relative accuracy for JOLs in the unrelated list and equivalent associative relatedness effects and overshadowing effects for the two age groups. By analogy, one might also expect to observe an equivalent contrast effect that is due to associative relatedness.

Method

Participants

There were 134 young and 88 old adults tested in Experiment 3. Fifteen young and 39 old adults were excluded from the analysis by using the same criteria in Experiments 1 and 2. The younger adults were primarily excluded because of extreme marginal values of the JOLs, and the excluded older adults because of floor effects on recall performance. There were 119 younger (62 in the mixed condition and 57 in the unrelated condition) and 49 older adults (20 in the mixed condition and 29 in the unrelated condition) who performed in the analysis after exclusion. Excluded and included participants did not differ on demographic and cognitive variables.

The younger adults (age: $M = 19.2$ years, $SD = 2.2$) were either undergraduate students from Georgia Institute of Technology or the University of North Carolina at Greensboro who received extra credit in a psychology course for their participation. The older adults (age: $M = 70.2$ years, $SD = 6.0$) were community-dwelling adults from the Atlanta metro area. They were either recruited through a locator service using random digit telephone dialing or had previously participated in an unrelated study at Georgia institute of Technology. The older adults received $20 for their participation.

The older adults perceived themselves as being in good health ($M = 1.7$, $SD = 0.6$), being in good health relative to same-aged peers ($M = 1.3$, $SD = 0.5$), and, on average, were taking 2.2 medications ($SD = 1.7$). They were also relatively well educated ($M = 14.9$, $SD = 2.4$).

As in the first two experiments, younger adults performed significantly lower than did the older adults on the advanced vocabulary test (young: $M = 15.1$, $SD = 3.7$; old: $M = 21.3$, $SD = 7.7$), $t(164) = -7.00$, $p < .01$. The younger adults performed better on letter comparison (young: $M = 25.9$, $SD = 4.8$; old: $M = 17.8$, $SD = 5.2$), $t(162) = 9.50$, $p < .01$, as well as on the pattern comparison test (young: $M = 43.5$, $SD = 6.8$; old: $M = 31.4$, $SD = 6.1$ than did the older adults), $t(162) = 10.40$, $p < .01$.

Apparatus and Materials

The apparatus was identical to Experiments 1 and 2. The two new paired-associate lists for Experiment 3 involved 60 items, split into half-lists. Items were drawn from a total set of 95 items, including 5 unrelated pairs used for practice. A subset of the words was drawn from the word lists used in Experiments 1 and 2, whereas the other words were drawn from the D. L. Nelson, McEvoy, and Schreiber (1998) Free Association Norms. The 60 experimental items were divided into two sets. Set 1 consisted of 30 unrelated items (e.g., *BARREL–STAR*) and was included in both the unrelated and mixed lists. Set 2 in each list comprised either unrelated paired associates in the unrelated list or related items in the mixed list. Set 2 items were constructed by taking 30 stimulus words and pairing them with related response words (e.g., *FLUTE–BAND*) or unrelated response words (e.g., *FLUTE–BEER*). Hence, the lists were constructed to consist of an identical half-list of unrelated items (Set 1), and then a half-list of related or unrelated items (Set 2) with identical stimulus words. The lists were constructed in this fashion so that individuals in both the mixed and unrelated conditions would receive exactly the same stimulus cues for JOLs and for the recall test. The two paired-associate lists are provided in the Appendix.

Design and Procedure

The design of the experiment was a 2 × 2 factorial design with age (young vs. old) and list type (unrelated vs. mixed) as between-subjects variables. Set was a within-subject variable that carried the relatedness effect in the mixed-list condition. Participants were randomly assigned to either the unrelated or the mixed condition. They were tested in groups of 1 to 4 persons. The experimental sessions lasted from 60 to 90 min. After the participants gave their informed consent, they completed a brief set of questionnaires, followed by the computerized paired-associate task. As in the other experiments, psychometric tests were administered after completion of the memory task.

Each paired-associate item was presented in the center of the visual display for 4 s for the younger adults and 8 s for the older adults. Different presentation rates were used to minimize age differences in level of recall, and to ensure above-floor performance in the unrelated condition for older adults. Age comparisons of absolute accuracy would be most interpretable if the age groups had equivalent levels of recall performance (Dunlosky & Hertzog, 2000), although this is difficult to achieve in practice. Procedures for obtaining continuous JOLs were identical to Experiment 2. The presentation order of the practice items and the experimental items was randomized for each participant. The items were presented in a different randomized order at test.

Results

Table 5 reports the key outcomes for mean JOLs and relative accuracy. Because Experiment 3 was motivated by four central hypotheses, we focus on the relevant experimental effects here. Before reporting key results, we note that the different presentation rates did not equate older and younger adults on recall. Younger adults had higher levels of recall for unrelated ($M = 0.47$) and related items ($M = 0.79$), compared with older adults ($M = 0.33$ and 0.67, respectively), $F(1, 164) = 18.02$, $MSE = 0.062$, $p < .001$.

Table 5
Paired-Associate Recall, Judgments of Learning, and Relative Accuracy (γ) for Young and Old Adults in the Unrelated and Mixed-Lists Conditions, Separately for Set 1 (Unrelated) and Set 2 (Unrelated or Related) Items, and Aggregated Over the Entire List

Condition	Recall		JOL		γ	
	M	*SD*	*M*	*SD*	*M*	*SD*
Set 1						
Unrelated						
Young	.45	.19	42	17	.28	.28
Old	.37	.19	38	21	.29	.29
Mixed						
Young	.47	.19	35	17	.28	.32
Old	.31	.22	33	19	.36	.32
Set 2						
Unrelated						
Young	.48	.22	46	16	.35	.27
Old	.32	.19	39	21	.27	.38
Mixed						
Young	.79	.13	68	14	.10	.34
Old	.67	.18	75	14	.05	.33
Overall data						
Unrelated						
Young	.47	.20	44	16	.32	.20
Old	.35	.19	38	21	.29	.26
Mixed						
Young	.63	.14	52	14	.45	.18
Old	.49	.23	54	15	.49	.23

Note. Unrelated young, $n = 57$, unrelated old, $n = 29$; mixed young, $n = 62$, mixed old, $n = 20$. JOL = judgment of learning.

First, we hypothesized a contrast effect on unrelated items' JOLs when paired with related items. Mean JOLs were indeed lower for Set 1 unrelated items in the mixed list than for the same items when included in an unrelated list, $F(1, 164) = 4.11$, $MSE = 322.40$, $p < .05$, although the effect size was small (unweighted marginal means of 40% vs. 34%, $d = 0.3$). There were no significant age differences in the magnitude of JOLs for Set 1 unrelated items, $F(1, 164) = 1.08$, $MSE = 322.40$, $p > .25$. More critically, the Age × List interaction was negligible, $F(1, 164) = 0.16$, $MSE = 322.40$, $p > .25$, suggesting little age difference in the magnitude of the contrast effect. The contrast effect actually impaired younger adults' absolute accuracy for unrelated items, increasing the underestimation from –3% to –12%. Indeed, the focused statistical comparison produced an unrelated versus mixed-list effect on younger adults' absolute accuracy, $t(118) = 2.11$, $MSE = 497.30$, $p < .05$. The contrast effect had no influence on older adults' absolute accuracy, which was highly accurate in both conditions (differences of +1% and +2% in the unrelated and mixed conditions, respectively). Note that, in this case, the influence of the contrast effect on absolute accuracy was greater for younger adults, despite the fact that their mean level of unrelated item recall (47%) was closer to the midpoint of the scale (50%) than older adults' recall (33%).

Second, we also hypothesized that older adults would show greater sensitivity of JOLs to the relatedness manipulation. The mean JOLs showed that older adults were more sensitive to the relatedness manipulation than were younger adults. In the mixed-list condition, JOLs were higher for related items than for unrelated items, $F(1, 80) = 420.44$, $MSE = 104.00$, $p < .001$. More critically, the relatedness effect was stronger for older adults, $F(1, 80) = 6.09$, $MSE = 104.00$, $p < .05$. The older adults' JOLs produced a 42% mean difference in recall confidence between related and unrelated items; the corresponding difference for younger adults was 33%. An even more compelling demonstration of the effect was obtained by comparing unrelated versus mixed-condition mean JOLs for Set 2 items. In this case, the stimulus words presented for related and unrelated items' JOLs were identical. There was again a large relatedness effect, $F(1, 164) = 118.76$, $MSE = 252.40$, $p < .001$, and an Age × Relatedness interaction, $F(1, 164) = 6.91$, $MSE = 252.40$, $p < .01$. Older adults showed a 36% difference in JOLs between related and unrelated items, compared with younger adults' 22% difference. Thus, Experiment 3 definitively demonstrates that older adults' JOLs are more sensitive to the relatedness manipulation.

Third, there was no evidence of an overshadowing effect in Experiment 3. The γ correlations for Set 1 items did not differ between conditions ($F < 1$). Nor were there any age differences in relative accuracy ($F < 1$). Indeed, the γs reported in Table 5 were slightly higher for older adults.

Fourth, the γ correlations in Table 5 yielded differences in relative accuracy for unrelated versus related items. The marginal mean γ correlation was .31 across all three unrelated conditions (Sets 1 and 2 in the unrelated list; Set 1 in the mixed list). In all cases, the null hypothesis that the population $\gamma = 0$ could be rejected ($p < .001$). On the other hand, γ was much lower for unrelated items (Set 2, mixed list), marginal mean $\gamma = .08$. The null hypothesis that JOLs are uncorrelated with recall for related items could not be rejected, $F(1, 80) = 2.87$ $MSE = 0.12$, $p > .05$. The difference in relative accuracy for unrelated versus related items was robust. Significant differences were detected in both a within-subject test for the mixed list, $F(1, 80) = 21.03$, $MSE = 0.09$, $p < .001$, and a between-subjects test of γ for Set 2 items, $F(1, 167) = 17.90$, $MSE = 0.09$, $p < .001$. In neither case did the Age × Relatedness interaction approach significance ($p >$ .25). Hence, we can conclude that both younger and older adults do attend to aspects of encoding unrelated paired associates that are diagnostic of later recall when they make JOLs. On the other hand, people's judgments about encoding related items do not discriminate effectively between related items' differential probability of recall.

Finally, Table 6 reports the privileged access effect associated with the yoking procedure. The yoked γ of JOLs with others' recall were small for the unrelated items of Set 1 and for the unrelated condition, Set 2, for both young and older adults, ranging from −.01 to .17 (young: $M = 0.07$, old: $M = 0.09$). Furthermore, in five out of six cases, the yoked γ was significantly smaller than the actual γ ($p < .05$). Thus, for unrelated items, both age groups demonstrated privileged access. As in previous experiments, there was no significant privileged access effect for related items for either younger, $t(60) = 1.34$, $p > .10$, or older adults, $t(18) = 0.60$, $p > .10$. However, when the mixed-list condition was treated in the aggregate, pooling unrelated and related items, then yoked γ was greater than chance for both age groups and conspicuously higher for the older adults. Indeed, the overall yoked γ of .41 for older adults in the mixed condition was not significantly different from their actual γ of .49, $t(18) = 1.18$, $p > .10$. This difference was reliable in the younger adults, $t(60) = 3.87$, $p < .01$. Moreover, yoked JOLs were more highly correlated for older adults ($\gamma = .56$, $SE = 0.05$) than for younger adults ($\gamma = .40$, $SE = 0.02$), $t(39) = 2.95$, $p < .01$, suggesting greater entrainment of older adults' judgments by associative relatedness. This pattern replicates earlier outcomes from Experiment 1, suggesting the importance of relatedness for JOLs and the greater influence of this cue for older adults' monitoring.

One final aspect of Experiment 3 is of note: Absolute accuracy was better for younger adults than for older adults, $F(1, 191) = 7.61$, $p < .01$, $MSE = 0.03$ ($M = 18\%$ absolute deviation for younger adults and $M = 23\%$ absolute deviation for older adults). No other effects were reliable. However, the two indices of absolute accuracy diverged. Younger adults demonstrated poorer absolute accuracy of simple difference scores overall, $F(1, 192) = 11.60$, $MSE = 0.106$, $p < .001$ ($M = -12\%$ for younger adults and $M = 4\%$ for older adults).

Discussion

Experiment 3 provided persuasive evidence that older adults' JOLs are indeed more sensitive to associative relatedness than are younger adults' JOLs. Older adults' mean JOLs showed a greater relatedness effect, and older adults' relative accuracy in the mixed list, pooling over all items, was not appreciably greater than their associated yoked γ. At the same time, this greater sensitivity to relatedness did not produce a greater contrast effect or an overshadowing effect for older adults. Instead, both age groups showed a contrast effect (lowered JOLs for unrelated items when paired with related items), and neither age group showed an overshadowing effect (reduced relative accuracy for unrelated items when paired with related items).

The failure to find an overshadowing effect is compelling, given that the contrast effect demonstrated that JOLs for unrelated items were indeed influenced by mixing them with related items. Despite this influence, relative accuracy was unaffected by the mixed-list manipulation. Thus, we can conclude that the apparent overshadowing effect in Experiment 1 emerges only when a discrete JOL rating scale is combined with the relatedness manipulation. Likewise, older adults do not show reduced relative accuracy within the unrelated items set when a continuous JOL rating is used, despite their greater sensitivity to the relatedness manipulation.

Thus, Experiment 3 does not support the hypothesis of an age-related deficit in monitoring learning. Instead, older adults produce equivalent relative accuracy within the unrelated-item sets, and the relatedness manipulation does not distort this pattern, despite its other effects. Furthermore, Experiment 3 provides evidence that individuals in both age groups can and do monitor multiple cues when making JOLs, at least for unrelated items (when γ correlations were greater than chance, even in the mixed-list condition). Older adults did show poorer absolute JOL accuracy in this experiment, when absolute accuracy is defined as absolute deviations of mean JOLs from mean recall. However, younger adults demonstrated robust underconfidence that was more inaccurate than the older adults' trend toward mild overcon-

Table 6

Privileged Access Effects for Experiment 3 With Actual Recall–JOL γ, Yoked JOL–Recall γ, and Yoked JOL–JOL γ Correlations for Young and Old Adults in the Unrelated and Mixed-List Conditions

	Actual γ						Yoked γ						Yoked JOL–JOL γ					
	Overall		Set 1		Set 2		Overall		Set 1		Set 2		Overall		Set 1		Set 2	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Unrelated																		
Young	.33	.21	.28	.28	.35	.28	.09	.19	.06	.24	.12	.30	.09	.11	.08	.17	.10	.13
Old	.30	.26	.30	.28	.28	.38	.07	.21	.15	.33	.00	.29	.10	.15	.13	.20	.10	.21
Mixed																		
Young	.45	.18	.28	.32	.10	.33	.33	.18	.04	.30	.01	.40	.40	.14	.06	.18	.13	.19
Old	.49	.23	.36	.32	.05	.32	.41	.23	.17	.32	−.01	.38	.56	.15	.14	.25	.17	.22

Note. JOL = judgment of learning.

fidence. The variability in patterns of age differences in absolute accuracy between indices would not seem consistent with the hypothesis of a molar age deficit in judgment accuracy.

General Discussion

Influences on JOLs and JOL Accuracy

The experiments reported here support the cue-utilization approach to JOLs discussed by Koriat (1997) and others (e.g., Schwartz et al., 1997). The experimental manipulation of associative relatedness resulted in effects on JOLs and JOL accuracy consistent with the argument that individuals can attend to different types of cues regarding degree of item learning. In mixed paired-associate lists (i.e., those including both related and unrelated items), relatedness is a cue that has a major impact on JOL ratings. This intrinsic cue has a substantial impact on the judgments, and the sensitivity of JOLs to relatedness acts to increase relative accuracy because relatedness is strongly related to actual probability of recall.

Is sensitivity to intrinsic cues a sufficient account of relative JOL accuracy? Consistent with Koriat (1997), our evidence shows that it is not. Relative accuracy was significantly greater than chance for unrelated items, and γ for both unrelated and related items increased after prestudy. Furthermore, the privileged access effects, although confirming that relatedness is an important contributor to aggregate accuracy, demonstrate that individuals have access to aspects of their own encoding that predict subsequent recall independently of normative item characteristics.

A cue-utilization framework holds promise for advancing our understanding of sources of variance that influence metacognitive judgments. Consider the finding that relative JOL accuracy was higher for unrelated items than for related items (see also Connor et al., 1997, Experiment 3). What is the explanation of this phenomenon? It appears that a source of information used by individuals to make JOLs is more diagnostic of recall for unrelated items than for related items. It is well-known that production of sentential or imaginal mediators during study substantially increases paired-associate recall for unrelated concrete nouns, relative to items for which no mediator is produced (e.g., Bower, 1970; Dunlosky & Hertzog, 1998a; Richardson, 1998). Mediator production has less of an effect on recall of related items (Dunlosky & Hertzog, 1998a). It may be the case that JOLs are based on the subjective success in forming an effective mediator between word pairs, at least in conditions (as in these experiments) in which participants are informed about mediational strategies.

This pattern of results is potentially relevant to the issue of cue utilization when making JOLs. It could be the case that the effectiveness of the generated mediator has a larger impact on probability of recall for unrelated pairs. For example, recall of related items can benefit from associative priming after presentation of the stimulus words during test (McEvoy, Holley, & Nelson, 1995), whereas such priming would have little impact on recall of unrelated items. Hence, differences in mediator quality should have greater impact for unrelated items. To the extent that individuals access cues regarding mediator quality regardless of item type, then the differential diagnosticity of these cues could account for the differences in relative accuracy between related and unrelated items. The hypothesis generated from a cue-utilization perspective is that mediator quality influences JOLs, and that subjective quality of mediators at encoding will generate differences in relative accuracy of JOLs for related and unrelated items. This hypothesis is testable, given a theory of mediator effectiveness and measurement of both mediator properties and subjective ratings of mediator quality.

Age Differences in Monitoring Learning

The present research provides further evidence that no major age-related impairments occur for monitoring accuracy. Older adults actually had superior absolute JOL accuracy in most instances and had equivalent relative JOL accuracy compared with younger adults in Experiments 2 and 3. They also showed a benefit of prestudy on relative accuracy in Experiment 1, as did younger adults.

Had we conducted only Experiment 1, we probably would have concluded that there was an age deficit in monitoring that is due to older adults' overreliance on relatedness as a cue for making JOLs. In Experiment 1, older adults displayed lower relative accuracy for discrete JOLs within the relatedness conditions, and did not manifest a robust privileged access effect. However, the age differences in relative accuracy disappeared when a continuous JOL was used in Experiments 2 and 3. Thus, it appears that older adults' greater emphasis on associative relatedness in making JOLs is disadvantageous when a discrete JOL rating scale with limited response options is used. However, this effect does not represent an age deficit in monitoring multiple cues. The fact that older adults showed equivalent relative accuracy for unrelated items in Experiment 3, even in the mixed-list condition (i.e., no overshadowing effect), persuasively argues against a deficit in attending to multiple cues. Older adults can and do attend to cues other than the salient cue of relatedness when making judgments about unrelated items.

The age differences in confidence effects, when observed, favored older adults in all three experiments. Younger adults underestimated future recall performance in their JOLs, especially for related items. This result was actually expected, given our earlier work (Connor et al., 1997) and the inclusion of related items in the word lists. Younger adults' recall was excellent and, for all experiments mixing related and unrelated words, was well above 50%. We interpret the age difference in absolute accuracy as support for the argument that midpoint anchoring of JOLs can act to constrain absolute accuracy as a function of recall level, without affecting relative accuracy. However, other patterns in the data show that midpoint anchoring alone cannot account for absolute accuracy effects, and a more comprehensive theory of processes that influence JOLs is needed to fully explain variable absolute accuracy patterns across experiments and studies. Nevertheless, this study (see also Connor et al., 1997) indicates that inferences of age differences in prediction accuracy prevalent in the early gerontological literature can be discounted. Typical findings of overconfidence by older adults are at this point more parsimoniously explained by reference to age differences in level of memory performance, irrespective of age. Overestimation by older adults can often be attributed to mean levels of performance well below 50% for the older group, combined with predictions or mean JOLs much closer to 50%, on average (Connor et al., 1997).

The effect was most obvious in Experiment 1, in which prestudy elevated the level of recall without affecting mean JOLs for either age group. In these experimental conditions, younger adults performed well above midpoint, and in some cases relatively close to ceiling. As a result, their JOLs and global predictions substantially underestimated performance. An interesting question is whether the JOLs can be experimentally manipulated to better track the high levels of recall in such conditions. Additional unpublished data from our experiments with younger adults suggest that discrete JOLs are surprisingly resistant to upward movement toward high levels of recall. In one follow-up experiment, we provided participants with younger adults' mean recall for related and unrelated items from Experiment 1 prior to asking them to study and make JOLs. Despite the information that mean recall was above 80% for related items, mean JOLs for related items, using the discrete 6-point rating scale, did not appreciably increase from the mean JOL of 70% reported in Experiments 1 and 2 (see Table 2).

Experiment 3 showed that the contrast effect degraded absolute accuracy for younger adults, but not for older adults. When aging studies have found greater magnitudes of overestimation by older adults, they have typically argued for a monitoring deficit (e.g., Bruce et al., 1982). Using this logic, one might claim that Experiments 1 and 3 provide evidence for an experience deficit in younger adults. Of course, it seems more plausible that the problem is with use of absolute accuracy as a measure of metacognitive monitoring. Note also that in one case, Experiment 3, younger adults had better absolute deviation scores than older adults, opposing the consistent pattern in over- and underconfidence. We argue that the results of this study, and of Connor et al.'s (1997), indicate that age differences in absolute accuracy are most likely an artifact of the scaling of JOLs and experimental design (Wallsten, 1996). In any event, a definitive test of age differences in absolute accuracy requires control on levels of recall, equating the two groups in some manner (e.g., by providing additional study time to older adults; see Dunlosky & Hertzog, 2000). Without adequate control for group differences in level of recall, unambiguous inferences regarding age differences in absolute accuracy do not appear possible.

This study provides additional evidence that, despite some age differences in aspects of monitoring learning (i.e., a greater sensitivity to associative relatedness), older adults are essentially unimpaired in their ability to monitor learning. Thus, monitoring learning is spared, even when learning itself is impaired. Further progress in our understanding of the boundary conditions under which older adults might show apparent or even real monitoring deficits is possible, provided that principles of the cue-utilization framework articulated by Koriat (1997) and others are applied and extended to new experiments. It may well be the case that when monitoring requires inferences about relative effects of multiple experimental manipulations, or when judgments must be made quickly or under cognitive load, age differences in monitoring accuracy will emerge. The results of these experiments suggest that one should be conservative about attributing any observed age differences in relative accuracy (such as those we observed in Experiment 1) to a fundamental age deficit in monitoring or resource utilization during monitoring. We regard the present study as showing that a multiple-cue utilization study could further examine how individuals do or do not attend differentially to multiple cues that vary in their diagnosticity for actual learning and performance. Whether failure to account for relevant, available cues reflects a resource limitation or a mere difference in how cues are weighted and utilized in making the JOL will remain an issue that can be resolved, ultimately, only by having an accurate model of how different individuals make JOLs.

For now, we continue to be intrigued by the possibility that older adults may be fully capable of effective monitoring of learning, but may not always spontaneously utilize monitoring effectively as a control process to optimize learning (Dunlosky & Connor, 1997; Murphy et al., 1987). If true, then interventions that target utilization of monitoring through procedures such as self-testing hold promise for enhancing learning by older adults (without necessarily having any effect on the age-related causes of learning deficits, per se; see Dunlosky & Hertzog, 1998b).

References

Anderson, N. D., Craik, F. I. M., & Naveh-Benjamin, M. (1998). The attentional demands of encoding and retrieval in younger and older adults: I. Evidence from divided attention costs. *Psychology and Aging, 13*, 405–423.

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.

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

Bower, G. H. (1970). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, *9*, 529–533.

Bruce, P. R., Coyne, A. C., & Botwinick, J. (1982). Age differences in metamemory. *Journal of Gerontology*, *37*, 354–357.

Carroll, M., & Nelson, T. O. (1993). Effects of overlearning on the feeling of knowing are more detectable in within-subjects than in between-subjects designs. *American Journal of Psychology*, *106*, 227–235.

Carroll, M., Nelson, T. O., & Kirwan, A. (1997). Tradeoff of semantic relatedness and degree of overlearning: Differential effects on metamemory and long-term retention. *Acta Psychologica*, *95*, 239–253.

Cavanaugh, J. C. (1989). The importance of awareness in memory aging. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 416–436). New York: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Craik, F. I. M., & Jennings, J. (1992). Human memory. In F. I. M. Craik & T. A. Salthouse, (Eds.), *The handbook of aging and cognition* (pp. 51–110). Hillsdale, NJ: Erlbaum.

Devolder, P. A., Brigham, M. C., & Pressley, M. (1990). Memory performance awareness in younger and older adults. *Psychology and Aging*, *5*, 291–303.

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. (1998a). Aging and deficits in associative memory: What is the role of strategy use? *Psychology and Aging*, *13*, 597–607.

Dunlosky, J., & Hertzog, C. (1998b). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–275). Mahwah, NJ: Erlbaum.

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

Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic–extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1180–1191.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 374–380.

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545–565.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (1998). *Metacognition in educational theory and practice*. Mahwah, NJ: Erlbaum.

Hertzog, C. (1989). The influence of cognitive slowing on age differences in intelligence. *Developmental Psychology*, *25*, 636–651.

Hertzog, C., Dixon, R. A., & Hultsch, D. F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, *5*, 215–227.

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and aging. In T. A. Salthouse & F. I. M. Craik (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 417–466). Mahwah, NJ: Erlbaum.

Hertzog, C., Saylor, L. L., Fleece, A. M., & Dixon, R. A. (1994). Metamemory and aging: Relations between predicted, actual, and perceived memory task performance. *Aging and Cognition*, *1*, 203–237.

Kausler, D. H. (1994). *Learning and memory in normal aging*. New York: Academic Press.

Keleman, W. L. (2000). Metamemory cues and monitoring accuracy: Judging what you know and what you will know. *Journal of Educational Psychology*, *92*, 800–810.

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *American Journal of Psychology*, *93*, 329–343.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cued-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349–370.

Lovelace, E. A. (1984). Metamemory: Monitoring future recall ability during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 756–766.

Lovelace, E. A. (1990). Aging and metacognitions concerning memory function. In E. A. Lovelace (Ed.), *Aging and cognition: Mental processes, self awareness, and interventions* (pp. 157–188). Amsterdam: North-Holland.

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

Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1263–1274.

McEvoy, C. L., Holley, P. E., & Nelson, D. L. (1995). Age effects in cued recall: Sources from implicit and explicit memory. *Psychology and Aging*, *10*, 314–324.

Metcalfe, J., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press.

Murphy, M. D., Schmitt, E. 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, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved August 1999 from http://www.usf.edu/FreeAssociation.

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., & 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., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–141). New York: Academic Press.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monographs*, *76*(Suppl. 1, Pt. 2), 1–25.

Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology*, 14, 330–345.

Price, P. C., & Yates, J. F. (1993). Judgmental overshadowing: Further evidence of cue interaction in contingency judgment. *Memory & Cognition*, *21*, 561–572.

Rabinowitz, J. C., Craik, F. I. M., Ackerman, B. P., & Hinchley, J. L. (1982). Aging and metamemory: The roles of relatedness and imagery. *Journal of Gerontology*, *37*, 688–695.

Reder, L. M. (Ed.). (1996). *Implicit memory and metacognition*. Hillsdale, NJ: Erlbaum.

Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, *5*, 597–614.

Ryan, M. P., Petty, C. R., & Wenzlaff, R. M. (1982). Motivated remembering efforts during tip-of-tongue states. *Acta Psychologica*, *51*, 137–147.

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

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*, 763–776.

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science*, *6*, 132–137.

Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging*, *4*, 131–135.

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.

Wallsten, T. S. (1996). An analysis of judgment research analyses. *Organizational Behavior and Human Decision Processes*, *65*, 220–226.

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

Appendix

Stimulus Materials

Paired Associate List: Experiments 1 and 2

Unrelated pairs		Related pairs	
BONE	WINE	CHAIR	ARM
ANIMAL	LIBRARY	PAINT	ARTIST
THROAT	PRINCE	FLUTE	BAND
BABY	FOREST	PANTS	BELT
MESSAGE	SOLDIER	SKIN	BODY
MILK	CLERK	PAPER	BOOK
UNCLE	COAL	LEAF	BRANCH
COLUMN	KNIFE	CAB	BUS
TEXT	MAID	BICYCLE	CAR
TOOL	COAST	FENCE	CHAIN
ROAD	DANCER	HEART	CHEST
FLOWER	HORN	CAKE	PARTY
WHEEL	SENATOR	PIPE	CIGARETTE
WIRE	FORT	TIE	COAT
ROOM	BIRD	SUGAR	COFFEE
CROWD	RICE	PIG	COW
TONGUE	PICTURE	SIDEWALK	CONCRETE
HANDLE	CORN	LAMP	DESK
CAMERA	BRAIN	WALL	DOOR
TARGET	SEED	CRAYON	DRAWING
COTTON	SNAKE	SNOW	RAIN
DIRT	QUEEN	NOSE	EAR
COMPOSER	WOOD	MILL	FACTORY
DRESS	MOVIE	MIRROR	FACE
CLOTH	ATOM	OYSTER	FISH
PRISON	TISSUE	WEED	GARDEN
BAKER	WAGON	PLASTIC	GLASS
SALT	MAYOR	BEAR	HONEY
BEER	GRASS	SUIT	JACKET
TRAFFIC	BEACH	STOVE	KITCHEN

Paired Associate List: Experiment 3

Unrelated pairs		Shared stimulus	Related response	Unrelated response
CAT	MARKET	CHAIR	ARM	MAMMAL
POTATO	FROG	PAINT	ARTIST	INCH
VOLCANO	DOLL	FLUTE	BAND	BEER
GLACIER	SEAT	LEATHER	BELT	WIRE
BARREL	STAR	SKIN	BODY	COACH
IVY	BIRD	PAPER	BOOK	MONARCH
BLOSSOM	HOOF	LEAF	BRANCH	CROWD
DOCTOR	BULLET	BICYCLE	CAR	RICE
BOWL	JURY	FENCE	CHAIN	MOVIE
FOAM	MEADOW	HEART	CHEST	TOOL
MANTLE	ELEPHANT	CAKE	PARTY	TARGET
FLAG	LEOPARD	PIPE	CIGARETTE	KNIFE
LUMP	MULE	TIE	COAT	UNCLE
STRING	IRON	SUGAR	COFFEE	COLUMN
HOTEL	UMBRELLA	PIG	COW	MAYOR
DIAMOND	CELLAR	SIDEWALK	CONCRETE	ATOM
LOCKER	NUN	LAMP	DESK	SOLDIER
SLIPPER	YACHT	WALL	DOOR	MESSAGE
HARNESS	HARP	CRAYON	DRAWING	HORN
SNAKE	KETTLE	SNOW	RAIN	CLERK
HOME	POET	MIRROR	FACE	COTTON
TOY	HILLSIDE	MILL	FACTORY	SALT
PISTON	BOULDER	STAMP	LETTER	COAST
OATS	MAST	OYSTER	FISH	WHEEL
PLANET	BOARD	WEED	GARDEN	BRAIN
ICEBOX	ACROBAT	PLASTIC	GLASS	THROAT
INK	TOWER	BEAR	HONEY	LIBRARY
JOURNAL	MACARONI	STOVE	KITCHEN	TRAFFIC
SPINACH	BLOOD	PELT	FUR	TEXT
LAWN	BANNER	FORT	WOODS	BABY

Received August 24, 2000
Revision received May 29, 2001
Accepted June 1, 2001 ■