*Experimental Aging Research*, 34: 379–391, 2008

ISSN: 0361-073X print/1096-4657 online
DOI: 10.1080/03610730802271898

# DO OLDER ADULTS SHOW LESS CONFIDENCE IN THEIR MONITORING OF LEARNING?

**Michael J. Serra**
**John Dunlosky**

Department of Psychology, Kent State University, Kent, Ohio, USA

**Christopher Hertzog**

School of Psychology, Georgia Institute of Technology, Atlanta, Georgia, USA

*Although aging has a minimal effect on the accuracy of people's judgments of learning (JOLs) at predicting future memory performance, older adults may be less confident in these memory judgments—similar to the age declines often reported with memory self-efficacy. To evaluate this possibility, the authors had younger and older adults make JOLs for paired associates and rate their confidence in the accuracy of each JOL. Age-related declines in confidence in judgments were evident for immediate JOLs but not for delayed JOLs. Implications of these outcomes for theory of JOLs and explaining age-related differences in self-regulated study are discussed.*

About two decades of research on aging and metacognitive monitoring has supported the conclusion that people's monitoring of their learning of new materials remains largely intact late into adulthood.

Received 24 January 2006; accepted 28 March 2007.
Current address Michael J. Serra is Department of Psychology, Columbia University, New York, New York, USA.
This research was supported by a grant from the National Institute on Aging, one of the National Institutes of Health (R37 AG13148). Results from the experiment were presented at the 50th Annual Meeting of the Southeastern Psychological Association in March, 2004. The authors thank Julie Baker and Joshua Kelley for their help in running the participants.
Address correspondence to Michael J. Serra, 1190 Amsterdam Avenue, 406 Schermerhorn Hall, Psychology Department, Columbia University, New York, NY 10027, USA. E-mail: ms3439@columbia.edu

Much of this research has used variations on a relatively standard procedure. Participants study simple materials—e.g., paired associates—for an upcoming test, and during study, they also explicitly judge the likelihood of remembering each item. These *judgments of learning* (JOLs) are influenced by many factors (e.g., semantic relatedness and encoding fluency), and their accuracy at predicting test performance is usually above chance. Most relevant here is that the influence of these factors on JOLs and their accuracy is typically equivalent for older adults and younger adults (for a review, see Hertzog & Hultsch, 2000).

Such age equivalence in JOL accuracy contrasts the age-related differences that have been demonstrated in how people use JOLs to regulate learning (e.g., Dunlosky & Connor, 1997; Miles & Stine-Morrow, 2004; see also Souchay & Isingrini, 2004). Specifically, older adults appear to rely less on their JOLs to differentially allocate self-paced study time to items that require further study, which in turn can undermine their learning. Although these outcomes may in fact be attributable to age deficits in self-regulation per se, an alternative possibility is that subtle age differences involving monitoring are responsible. In particular, even though the accuracy of monitoring judgments is influenced minimally by aging, older adults may be *less confident* in their JOLs and hence may be less likely to use them to guide learning. Our primary goal for the present research was to evaluate this possibility by directly estimating older and younger adults' confidence in their monitoring judgments.

One reason to expect age-related differences in confidence about JOLs is offered by research on memory self-efficacy. Compared to younger adults, older adults often have much less confidence in their ability to learn in general, and they more often complain about forgetting. Such changes in memory self-efficacy are not entirely unwarranted, although individual differences in memory self-efficacy do not entirely reflect actual declines in memory across time (McDonald-Miszczak, Hertzog, & Hultsch, 1995). Instead, these changes also likely reflect (inaccurate) beliefs about personal memory dysfunction. Similarly, although older adults do not demonstrate deficits in the accuracy of their judgments about learning, they may demonstrate less confidence—or efficacy—in those judgments.

In the present study, we tested the hypothesis that aging influences confidence in monitoring in the following manner. Participants were asked to remember paired associates (pairs of unrelated nouns, e.g., daffodil–blood) for a later test. During study, participants made a JOL for each item in which they predicted the likelihood of retrieving the second word of each pair when shown the first on the upcoming test. They made these JOLs on a standard scale, which ranged from 0 (0% chance of later retrieving the item) to 100 (100% chance of later retrieving the item). After making each JOL, participants then made a second-order judgment (SOJ) in which they rated their confidence in the accuracy of the JOL itself. This judgment also ranged from 0 (0% chance the JOL was accurate) to 100 (100% chance the JOL was accurate). As demonstrated in previous research (Dunlosky, Serra, Matvey, & Rawson, 2005), SOJs provide valuable information about confidence in judgments that cannot be inferred from the JOLs themselves. For instance, in their most informative analysis, Dunlosky et al. (2005) plotted mean SOJs as a function of the JOL ratings. Importantly, the curve was asymmetrical: younger adults showed reliably lower SOJs for the lowest JOLs (near 0) than for the highest JOLs (near 100). SOJs for JOL values near the middle of the scale (40 and 50) received the lowest SOJs. Given such asymmetry, it is apparent that SOJs and JOLs are not functionally equivalent and that the former must be collected in order to obtain a measure of confidence in the first-order monitoring judgment (for detailed arguments, see Dunlosky et al., 2005). In the present research, the SOJ-JOL function will also prove most sensitive for assessing whether age-related differences occur in the confidence of JOLs. If older adults are less confident, then SOJs will be lower (at each level of JOL) for older than younger adults.

In the present experiment, participants studied 60 paired associates and made either an immediate JOL or delayed JOL for each one. The former is made immediately after an item is studied, and the latter is made some time after a filled interval occurs (e.g., a minute or so filled by the study of other pairs). The two kinds of JOL were collected because predictive accuracy is typically much greater for delayed than immediate JOLs (Nelson & Dunlosky, 1991). Accordingly, if SOJs themselves are accurate, they should be higher in magnitude for delayed than immediate JOLs. This pattern was reported by Dunlosky et al. (2005) and was analyzed here using a directional planned comparison. More important, key outcomes concerned the overall magnitude of SOJs and the SOJ-JOL functions. In both cases, lower SOJs by older adults would signify diminished confidence in judgment in late adulthood.

## *METHODS*

### *Participants*

Thirty-two older adults (mean age = 66 years, $SD = 6.5$ years) were recruited from Greensboro, North Carolina. Forty-eight younger adults (mean age = 20 years, $SD = 4.0$ years) from an Introductory Psychology class at the University of North Carolina at Greensboro (UNCG) volunteered to participate in the study. In exchange for their participation, the older adults received 20 dollars, and the younger adults received partial credit towards a class requirement. The mean number of years of education was 16 for older adults and 13 for younger adults. Older adults reported taking more medications ($M = 2.5$, $SD = 3.3$) than did younger adults ($M = 0.3$, $SD = 0.6$), $t(75) = 4.44$, $p < .05$. However, on a scale from 1 (excellent) to 4 (poor), both older adults ($M = 1.7$, $SD = 0.6$) and younger adults ($M = 1.6$, $SD = 0.6$) rated themselves as being in quite good health.

## *Design*

A 2 (age: older versus younger) $\times$ 2 (kind of JOL: immediate versus delayed) mixed design was used, with the kind of JOL being manipulated within each participant. Immediate JOLs were made for half of all items studied, and delayed JOLs were made for the remaining half.

## *Materials*

The same set of 66 paired associates used in Dunlosky et al. (2005) was used in this study. Each pair was composed of two unrelated nouns (e.g., ticket–dummy). Frequency of occurrence, concreteness, and meaningfulness values for the words of each pair are presented in Table A1.[1] Apple iMac computers were used to present the materials to the participants and to collect all responses.

[1]Given that word frequency can influence people's metacognitive judgments via familiarity processes (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989), older adults may be at an advantage if newer words—e.g., words that entered the vocabulary in the past decade—were included in the materials, perhaps because older adults would be less familiar with those items and hence could focus more on memorability per se. (We thank an anonymous reviewer for suggesting this possibility.) In contrast to this possibility, the words comprising our paired associates were largely (except for, "planet" and "glass") from a word list published in 1968 (Paivio, Yuille, & Madigan, 1968). Moreover, as shown in Table A1, the words were not highly polarized on the key dimensions—i.e., our list does not have a mix of concrete and abstract words. The mean concreteness rating for 130 of the 132 words used in this study (values were not available for "planet" and "glass") was 6.8 ($SD = .25$) and the mean meaningfulness rating of these words was 6.8 ($SD = .78$). As the words were all chosen to be concrete and meaningful, it is not likely that extreme variability in either of these factors substantively influenced the age-relevant effects examined in the current research. The mean frequency rating was 31.6 per million ($SD = 18.3$; to compute mean values, the value of 50 was used for all words with ratings of "A" or "AA").

**Table A1. Characteristics of the paired-associate items used in the present study: Frequency of occurrence, concreteness, and meaningfulness of the 66 word pairs**

| | Stimulus | | | | Response | | | |
|---|---|---|---|---|---|---|---|---|
| Pair | Word | Freq. | Conc. | Mean | Word | Freq. | Conc. | Mean |
| 1 | Arm | AA | 6.96 | 6.92 | Market | AA | 6.08 | 7.04 |
| 2 | Icebox | 1 | 6.90 | 6.28 | Acrobat | 1 | 6.38 | 5.67 |
| 3 | Banner | 23 | 6.58 | 6.20 | Nun | 9 | 6.76 | 6.60 |
| 4 | Macaroni | 2 | 7.00 | 6.00 | Bar | A | 6.83 | 6.08 |
| 5 | Barrel | 32 | 6.94 | 6.16 | Star | AA | 6.73 | 7.04 |
| 6 | Beast | A | 6.51 | 6.08 | Fabric | 21 | 6.55 | 6.48 |
| 7 | Vest | 21 | 6.73 | 5.96 | Bird | AA | 6.96 | 7.88 |
| 8 | Blister | 6 | 6.67 | 7.13 | Cabin | A | 6.96 | 7.20 |
| 9 | Daffodil | 3 | 7.00 | 6.96 | Blood | AA | 6.82 | 6.56 |
| 10 | Blossom | 43 | 6.62 | 7.60 | Locker | 3 | 6.96 | 6.44 |
| 11 | Rattle | 28 | 6.60 | 6.80 | Board | AA | 6.87 | 6.92 |
| 12 | Lawn | 36 | 6.96 | 7.28 | Book | AA | 6.96 | 7.68 |
| 13 | Piston | 5 | 6.87 | 7.40 | Boulder | A | 6.96 | 5.88 |
| 14 | Pelt | 6 | 6.20 | 6.76 | Brain | A | 6.63 | 6.80 |
| 15 | Bronze | 19 | 6.59 | 6.19 | Whale | 8 | 6.96 | 7.24 |
| 16 | Bullet | 22 | 7.00 | 7.68 | Yacht | 2 | 6.96 | 7.20 |
| 17 | Candy | 32 | 6.56 | 6.39 | Prairie | 25 | 6.80 | 8.16 |
| 18 | Cat | AA | 7.00 | 6.76 | Jury | 26 | 6.17 | 6.88 |
| 19 | Cellar | 32 | 6.83 | 6.79 | Elbow | 26 | 6.94 | 5.16 |
| 20 | Woods | AA | 6.87 | 7.84 | Chin | 27 | 6.96 | 5.28 |
| 21 | Church | AA | 6.59 | 7.52 | Mammal | 6 | 6.31 | 5.80 |
| 22 | Claw | 23 | 6.89 | 6.67 | Salad | 28 | 6.83 | 7.20 |
| 23 | Jail | 22 | 6.69 | 7.38 | Coffee | A | 6.89 | 7.28 |
| 24 | Glacier | 8 | 6.93 | 7.21 | Cord | 30 | 6.93 | 6.00 |
| 25 | Corn | A | 6.90 | 6.96 | Planet | — | — | — |
| 26 | Cotton | AA | 6.90 | 7.13 | Reptile | 8 | 6.65 | 6.52 |
| 27 | Diamond | A | 6.94 | 7.84 | Umbrella | 13 | 7.00 | 6.76 |
| 28 | Monarch | 20 | 6.40 | 6.92 | Doll | 46 | 6.94 | 6.12 |
| 29 | Door | AA | 7.00 | 7.96 | Officer | AA | 6.32 | 5.43 |
| 30 | Slipper | 20 | 6.94 | 6.04 | Dove | 19 | 6.90 | 6.36 |
| 31 | Ticket | A | 7.00 | 7.21 | Dummy | 2 | 6.34 | 4.88 |
| 32 | Sunburn | 3 | 6.52 | 7.72 | Elephant | 35 | 7.00 | 6.88 |
| 33 | Flag | A | 6.94 | 6.54 | Window | AA | 7.00 | 6.76 |
| 34 | Flesh | A | 6.90 | 5.84 | Kettle | 27 | 7.00 | 7.44 |
| 35 | Foam | 21 | 6.73 | 6.44 | Meadow | 47 | 6.69 | 8.00 |
| 36 | Suds | 3 | 6.75 | 7.29 | Fowl | 20 | 6.58 | 7.36 |
| 37 | Potato | A | 7.00 | 7.13 | Frog | 25 | 6.96 | 6.56 |
| 38 | Fur | A | 6.69 | 7.36 | Oats | 23 | 6.90 | 7.16 |
| 39 | Glass | — | — | — | Journal | AA | 6.69 | 5.88 |
| 40 | Volcano | 14 | 6.83 | 7.60 | Hammer | 34 | 6.96 | 6.92 |
| 41 | Lake | AA | 6.90 | 9.22 | Harp | 20 | 6.94 | 6.00 |

(*Continued*)

**Table A1. (Continued)**

| | Stimulus | | | | Response | | | |
|---|---|---|---|---|---|---|---|---|
| Pair | Word | Freq. | Conc. | Mean | Word | Freq. | Conc. | Mean |
| 42 | Hillside | 20 | 6.59 | 6.92 | Revolver | 9 | 6.96 | 7.40 |
| 43 | Poet | A | 6.35 | 5.36 | Home | AA | 6.25 | 6.88 |
| 44 | Hoof | 26 | 6.90 | 6.32 | Slave | A | 6.38 | 6.17 |
| 45 | Hotel | A | 6.80 | 5.96 | Noose | 2 | 6.20 | 6.20 |
| 46 | Harness | 29 | 6.93 | 6.20 | Snake | 28 | 7.00 | 6.88 |
| 47 | Ink | 20 | 6.87 | 6.96 | Lark | 22 | 6.83 | 5.84 |
| 48 | Iron | AA | 6.87 | 6.12 | Leopard | 6 | 7.00 | 6.83 |
| 49 | Juggler | 1 | 6.45 | 5.84 | Mast | 26 | 6.96 | 7.12 |
| 50 | Leaflet | 7 | 6.62 | 4.92 | Tower | A | 6.96 | 6.42 |
| 51 | Seat | AA | 6.79 | 6.08 | Letter | AA | 6.94 | 5.96 |
| 52 | Doctor | AA | 6.62 | 7.32 | Lobster | 7 | 6.96 | 6.84 |
| 53 | Lump | 20 | 6.20 | 5.44 | Wine | A | 6.96 | 7.54 |
| 54 | Bowl | A | 6.90 | 7.24 | Missile | 2 | 6.80 | 6.48 |
| 55 | Mountain | AA | 7.00 | 7.58 | Skin | AA | 6.96 | 5.67 |
| 56 | String | A | 6.90 | 5.29 | Mule | 29 | 6.96 | 6.12 |
| 57 | Oven | 29 | 6.96 | 8.08 | Ship | AA | 6.93 | 7.87 |
| 58 | Toy | 49 | 6.63 | 6.96 | Pencil | 40 | 7.70 | 6.48 |
| 59 | Python | 1 | 6.42 | 5.88 | Building | AA | 6.94 | 5.48 |
| 60 | Spinach | 8 | 6.90 | 7.08 | Typhoon | 0 | 6.20 | 6.28 |
| 61 | Sugar | AA | 6.96 | 7.00 | Prison | A | 6.62 | 7.21 |
| 62 | Corpse | 9 | 6.89 | 6.52 | Forest | AA | 6.69 | 9.12 |
| 63 | Fox | 25 | 7.00 | 7.40 | Pudding | 17 | 6.60 | 7.31 |
| 64 | Butter | AA | 6.96 | 6.91 | Hospital | A | 6.80 | 7.44 |
| 65 | Person | AA | 6.51 | 5.68 | Storm | AA | 6.45 | 7.21 |
| 66 | Fire | AA | 6.66 | 7.36 | Apple | A | 7.00 | 7.67 |

*Note.* "Freq." = Frequency of Occurrence; "Conc." = Concreteness; "Mean." = Meaningfulness. Values are from Paivio, Yuille, & Madigan (1968). The frequency of the occurrence of the words represents the number of times the word appears per 1 million words; "A" and "AA" indicate "relatively high" frequencies of occurrence per million (i.e., $> 50$). Concreteness of the words was measured on a 7-point Likert scale with 1 indicating "highly abstract" and 7 meaning "highly concrete." Meaningfulness of the words was measured by having participants free-associate to the target word; the number is the mean number of words free-associated to the target word. Values were not available for the words "planet" and "glass."

## *Procedure*

The procedure was identical to that of Dunlosky et al. (2005). In particular, participants first read detailed instructions describing the task and the judgments that they would be making for each item. The order of the 66 pairs were randomized for each participant and presented individually for a fixed duration. This presentation rate was 8 s/pair for older adults and was 3 s/pair for the younger adults so as to minimize the typical age-related deficits demonstrated in recall performance. Equating groups on recall is critical for interpreting any age-related effects on judgments as deficits in metamemory and not to deficits in memory per se (for detailed rationale, see Connor, Dunlosky, & Hertzog, 1997). The first six items were used as a primacy buffer and were not entered into the analyses. For the remaining 60 items, half were randomly assigned to the immediate JOL condition and half to the delayed JOL condition. For immediate JOLs, after studying a given pair, participants were presented with the stimulus alone and were asked to judge the likelihood that they would later be able to recall the corresponding response on the criterion test. This JOL was made on a scale from 0 (0% chance of recall) to 100 (100% chance of recall) in increments of 10. For delayed JOLs, a delay occurred (greater than 1 min) between the study and JOL, with this delay being filled with the study and judgment of other pairs (as in Nelson & Dunlosky, 1991). Immediately after making each JOL (regardless of whether it was an immediate JOL or delayed JOL), participants made a SOJ for the JOL. For a SOJ, a participant judged how accurate the JOL he or she just made would be at predicting later recall. SOJs were also made on a scale from 0 (0% chance the JOL was accurate) to 100 (100% chance the JOL was accurate) in increments of 10.

After participants had studied and judged all the items, the order of pairs was randomized anew, and each stimulus was presented individually for paired-associate recall. Participants were encouraged to recall (and then type) a response corresponding to each stimulus, but they were allowed to omit responses.

## RESULTS

We first present secondary—but standard—analyses involving recall performance and JOLs, followed by the focal analyses of SOJs.

### Recall Performance

The proportion of correct recall performance was computed separately for items slated with immediate JOLs and with delayed JOLs. Means across older adults' values were .18 ($SE = .03$) and .19 ($SE = .03$) for items with immediate and delayed JOLs, respectively. For younger adults, means were .26 ($SE = .03$) and .24 ($SE = .03$) for items with immediate and delayed JOLs, respectively. The main effect of kind of JOL (immediate versus delayed), $F(1, 76) = 0.17$, $MSE = .01, p > .05$, and the interaction, $F(1, 76) = 0.24$, $MSE = .01$,

$p > .05$, were not significant. A slight trend occurred toward superior recall by younger adults, although it was not statistically significant, $F(1, 76) = 2.65$, $MSE = .06$, $p > .05$, indicating that our attempt to equate age groups on overall recall performance was mainly successful.

### Judgments of Learning (JOLs)

#### JOL Accuracy

Relative accuracy was operationalized by a gamma correlation between each participant's JOLs and recall performance across items. For immediate and delayed JOLs, respectively, the means across participants' gammas were .31 ($SE = .09$) and .73 ($SE = .09$) for older adults and were .46 ($SE = .06$) and .72 ($SE = .09$) for younger adults. Accuracy was significantly greater for delayed than immediate JOLs, $F(1, 66) = 15.5$, $MSE = .25$, $p < .05$, $ES = .19$, and this delayed JOL effect was equivalent in magnitude for both age groups, $F(1, 66) = 0.80$, $MSE = .21$, $p > .05$.

#### JOL Magnitude

For each participant, we computed the mean JOL value. The mean JOL magnitudes across participants for immediate JOLs and delayed JOLs, respectively, were 38 ($SE = 4.4$) and 25 ($SE = 3.7$) for older adults and were 41 ($SE = 3.2$) and 26 ($SE = 2.5$) for younger adults. The main effect of age and the interaction were not significant, $F(1, 76)$ values $< 1.0$, whereas the main effect of kind of JOL was significant, $F(1, 76) = 52.5$, $MSE = 136.4$, $p < .05$, $ES = .41$, indicating that JOL magnitude was higher for immediate than delayed JOLs.

The present age equivalence in JOL accuracy and JOL magnitude replicates outcomes from previous research (e.g., Connor et al., 1997) and, as important, it sets the stage for our main question: When compared to younger adults, do older adults show as much confidence in their judgments of learning?

### Second-Order Judgments (SOJs)

#### SOJ Magnitude

The mean across SOJs was computed for each participant. The mean SOJs across immediate JOLs and delayed JOLs, respectively, were 53 ($SE = 4.7$) and 57 ($SE = 6.0$) for older adults and were 58 ($SE = 4.0$) and 61 ($SE = 5.2$) for younger adults. Although an age-related trend is evident, the overall effect was not statistically significant, $F(1, 76) = 0.33$, $MSE = 1761.5$, $p = .57$. As expected, SOJ magnitude was greater for delayed than immediate JOLs, $F(1, 76) = 2.90$, $MSE = 163.0$, $p < .10$ (planned comparison), which demonstrates that people have greater confidence in the more accurate delayed JOLs. The interaction was not significant, $F(1, 76) = 0.09$, $MSE = 163.0$, $p > .05$.

### *SOJ Magnitude as a Function of JOLs*

The SOJ magnitude at each level of JOL (0, 10, 20, . . . , 90, 100) provides a fine-grained analysis of confidence in judgments that can reveal effects that would not be evident in the SOJ magnitude reported above (Dunlosky et al., 2005). For each participant, we computed the mean SOJ magnitude (as above) for all pairs that were given the same JOL rating. Mean SOJ magnitude plotted as a function of JOL ratings are presented in Figure 1.

Several noteworthy outcomes are apparent in this figure. First, the form of the SOJ-JOL function is not linear. In general, people's SOJs were greatest for the higher JOLs (80 and above) and decreased with JOLs, with a flattening of the curve around 40. Second, SOJs were higher for delayed than immediate JOLs. Of the 22 comparisons between SOJs for immediate versus delayed JOLs (11 for each age group), SOJs were lower for immediate JOLs in 18 comparisons ($p$ values $< .05$, sign test), greater in 2 ($p$ values $< .05$), and tied in 2 ($p$ values $> .05$). These results indicate that confidence in one's JOLs is not redundant with the JOLs themselves; that is, one cannot infer people's confidence in their judgments directly from a given

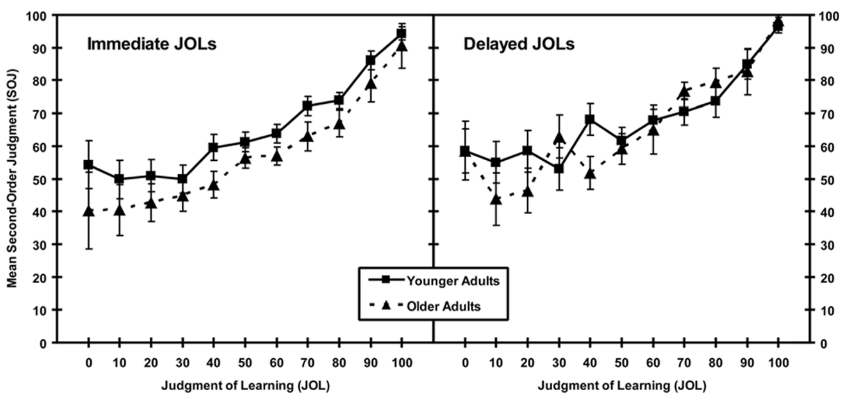


**Figure 1. Mean second order judgment (SOJ) magnitude as a function of judgment-of-learning (JOL) ratings both for immediate JOLs (*left panel*) and delayed JOLs (*right panel*). Bars represent corresponding standard errors of each mean.**

JOL rating. Finally, and most important, were the age contrasts. Inspection of Figure 1 indicates that older adults were consistently less confident in their immediate JOLs ($p < .05$, sign test), whereas age-related differences in confidence for delayed JOLs were minimal ($p > .05$, sign test).

## DISCUSSION

Our main question concerned older adults' efficacy in their ability to accurately judge their learning for individual word pairs. That is, as compared to their younger counterparts, are older adults less confident in their memory monitoring? Evidence from the present research provided an interpretable pattern of age-related effects. In particular, age-related differences were evident in SOJs for immediate JOLs, whereas age equivalence was evident for delayed JOLs. One explanation for these disparate effects is based on the idea that people use a variety of heuristics to make JOLs (Koriat, 1997). These heuristics concern how specific cues influence performance. For instance, Hertzog, Dunlosky, Robinson, and Kidder (2003) demonstrated that the latencies of generating interactive images to encode stimuli were negatively correlated with participants' JOLs, even though the time to generate these images was unrelated to recall performance. In other words, participants in their studies used the speed with which they generated an interactive image to encode an item as part of the basis of the JOL for that item. The fact that image-generation latency was not diagnostic of performance in this case negatively affected the accuracy of their judgments. Several studies have also shown that the fluency with which an item is retrieved on an earlier test is often positively correlated with the JOLs assigned to those items on subsequent study trials (for examples see Benjamin, Bjork, & Schwartz, 1998; Koriat & Ma'yan, 2005; Serra & Dunlosky, 2005). However, as discussed by Serra and Dunlosky (2005), retrieval fluency can be differentially diagnostic of later test performance depending on the type of materials being studied. JOLs based on retrieval fluency, therefore, should be more accurate when retrieval fluency predicts later test performance (Serra & Dunlsoky, 2005) than when it does not (Benjamin et al., 1998). But how does the use of such heuristics to make JOLs explain our age-related inconsistencies (and lack thereof) in SOJs?

The pattern of age differences in SOJs may result from the following. Based on the heuristic approach, highly salient cues may not be readily available for use as a basis for immediate JOLs (Koriat, 1997) for the homogeneous unrelated noun–noun pairs used in the present experiment. In this case, where people are not aware of relevant cues that could inform their JOLs, older adults may be reluctant to endorse extreme confidence in their judgments. By contrast, for the delayed JOLs made in the current experiment, a highly relevant cue was readily available—namely, retrieval fluency. When shown the prompt for a delayed JOL (e.g., the prompt "dog–?" for the pair "dog–spoon"), people attempt to retrieve the response (i.e., "spoon") and use the outcome of this retrieval attempt as a basis for the JOL (Koriat & Ma'yan, 2005; Nelson, Narens, & Dunlosky, 2004; Serra & Dunlosky, 2005; for a two-stage model, see Son & Metcalfe, 2005). It is also evident that delayed JOLs are made similarly by older and younger adults (e.g., Connor et al., 1997). In the case of delayed JOLs, people's confidence in their JOLs may increase because they can readily identify a particular cue—retrieval fluency—that informs their JOLs. The idea offered here is simply that age differences in confidence *about* monitoring may emerge when cues are not readily available as a basis for JOLs yet will diminish once salient cues become available. Although this is a post hoc hypothesis, it offers numerous empirical predictions that can guide future research. In addition to exploring the effects that other cues might have on SOJs, future research should seek to explain how JOLs are made when cues are not readily available and why younger adults may be more confident in these cases than are older adults.

Regardless of the ultimate explanation for these age-related differences (and equivalences) in confidence about monitoring, the pattern of effects has straightforward implications for explaining age differences in self-regulation. When regulation is based on immediate JOLs, older adults' reduced confidence in their monitoring may be partially responsible for the age deficits found in self-regulation (e.g., Miles & Stine-Morrow, 2004). Note, however, that older adults did not have reduced confidence in their delayed JOLs, yet age differences in regulation persist even when self-paced study is regulated by delayed JOLs (e.g., Dunlosky & Connor, 1997). Thus, reduced confidence in monitoring cannot provide a sufficient account for age-related deficits in self-regulation. Other explanations must be further pursued, such as the possibility that age-related differences in metacognitive regulation are yet another manifestation of general deficits with executive functioning that occur in advanced age (e.g., Souchay & Isingrini, 2004).

In summary, although older adults showed less confidence in their immediate JOLs as compared to younger adults, this age difference was relatively small and was not evident in confidence for delayed JOLs. Thus, not only does judgment accuracy appear to remain largely intact as we grow older, so does one's confidence in his or her judgments of learning.

## REFERENCES

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, *28*, 610–632.

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metacognitive index. *Journal of Experimental Psychology: General*, *127*, 55–68.

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

Dunlosky, J. & 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., Serra, M. J., Matvey, G., & Rawson, K. A. (2005). Second-order judgments about judgments of learning. (Special Issue). *Journal of General Psychology*, *132*, 335–346.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 22–34.

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

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

Koriat, A. & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*, 478–492.

McDonald-Miszczak, L., Hertzog, C., & Hultsch, D. F. (1995). Stability and accuracy of metamemory in adulthood and aging: A longitudinal analysis. *Psychology and Aging*, *10*, 553–564.

Miles, J. R. & Stine-Morrow, E. A. L. (2004). Adult age differences in self-regulated learning from reading sentences. *Psychology & Aging*, *19*, 626–636.

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., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods*, *9*, 53–69.

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

Serra, M. J. & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1258–1266.

Son, L. K. & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage model. *Memory & Cognition*, *33*, 1116–1129.

Souchay, C. & Isingrini, M. (2004). Age related differences in metacognitive control: Role of executive functioning. *Brain and Cognition*, *56*, 89–99.