Betting on Memory Leads to

Metacognitive Improvement by Younger and Older Adults

Shannon McGillivray & Alan D. Castel
University of California, Los Angeles

_In Press, Psychology and Aging_

Author Note

Shannon McGillivray and Alan D. Castel, Department of Psychology, University of California, Los Angeles.

We would like to sincerely thank Matthew Rhodes and David McCabe for their insightful comments and feedback. In addition we would like to thank past and present members of the UCLA Aging Lab, including Michael Friedman, Teal Eich, and Kathy Gerlach for their support and enthusiasm. Portions of this research were presented at the 13th Biennial Cognitive Aging Conference in Atlanta, Georgia, April 2010.

Please address all correspondence to Shannon McGillivray, University of California, Los Angeles, Department of Psychology, 1285 Franz Hall Box 951563, Los Angeles, CA 90095-1563, email: smcgillivray@ucla.edu.
Abstract

The present study examined how younger and older adults choose to selectively remember important information. Participants studied words paired with point values, and "bet" on whether they could later recall each word. If they bet on and recalled the word, they received the points, but if they failed to recall it, they lost those points. Participants (especially older adults) initially bet on more words than they later recalled, but greatly improved with task experience. The incorporation of rewards and penalties associated with metacognitive predictions, and multiple study-test trials, revealed that both younger and older adults can learn to maximize performance.

*Keywords:* aging, metamemory, memory, value, task experience

Word count: 3,809

Lines of text: 265
Individuals are often presented with more information than they can remember, and thus need to choose what information to focus on. If some information is presented as being more important, people will likely choose to selectively encode that information (e.g., Ariel, Dunlosky, & Bailey, 2009; Castel, 2008). However, the most important information to remember is often also associated with the largest negative consequences if forgotten. This potential trade-off between the importance of information and consequences if memory fails may be most important to those with declining memory function, such as older adults. The present study examines the degree to which informational importance influences what and how much older and younger adults try to remember. The current study also examines how consequences associated with the accuracy of memory predictions can change, and potentially improve, with task experience.

There is conflicting evidence as to whether metacognitive monitoring and control are negatively impacted in aging (Hertzog & Hultsch, 2000). Metacognition (or more specifically, metamemory) includes beliefs and knowledge about one’s memory ability and task demands (Dunlosky & Metcalfe, 2009), and these beliefs, in turn, can influence performance expectations, effort exerted during a memory task, and even one’s actual memory performance (Dixon, Rust, Feltmate, & Kwong See, 2007; Lachman, 2006; Lachman & Andreoletti, 2006). Metamemory tasks frequently involve asking participants to make judgments about what they think they will be able to remember, referred to as judgments of learning (JOLs). While some studies utilizing JOLs have found that older adults’ calibration between JOLs and actual memory performance does not differ from younger adults (Connor, Dunlosky, & Hertzog, 1997; Hertzog, Dunlosky, Powell-Moman, & Kidder, 2002; Hertzog & Hultsch, 2000; Souchay & Isingrini, 2004), older adults have, at times, been found to display larger patterns of overconfidence in their memory
abilities compared with younger adults, predicting they remember more than they actually are able to (Bruce, Coyne, & Botwinick, 1982; Bunnell, Baken, & Richards-Ward, 1999).

Fewer studies have examined the impact of task experience on metacognitive monitoring and control in older adults. Knowledge updating with task experience is typically assessed by presenting participants with the same set of information twice and examining the degree of improvement in predictions and strategy usage. Dunlosky and Herzog (2000) found that the absolute accuracy of global predictions was not greatly improved for either younger or older adults across two study-test trials; however, correlations between predictions and performance did increase with task experience for both age groups. When the benefits of task experience on knowledge updating are more apparent, it has been found that older adults’ ability to accurately update metacognitive predictions are impaired relative to younger adults (Matvey et al., 2002; Price, Hertzog, & Dunlosky, 2008). However, in general, these studies have all found that older and younger adults’ tend to lower their predictions with task experience, a finding consistent with the underconfidence-with-practice effect (Koriat, Sheffer, & Ma'ayan, 2002).

What is currently lacking in the existing literature is an investigation of older adults’ metacognitive accuracy when multiple (i.e., more than two), study-test trials are utilized (for an exception see Rast & Zimprich, 2009). Task experience and feedback may be particularly important for older adults (e.g., Jacoby, Wahlheim, Rhodes, Daniels, & Rogers, 2010) in order to learn to calibrate predictions with actual performance. On-line monitoring needed for accurate predictions may tax attentional and working memory systems that can become compromised in old age (Bieman-Copland & Charness, 1994; Craik, 2002; Craik & Byrd, 1982; Hasher & Zacks, 1988), and older individuals may require more time and experience to adopt appropriate
strategies and reach levels of performance on par with younger adults (e.g., Rogers, Hertzog, & Fisk, 2000; Touron, Hoyer, & Cerella, 2004).

The present study examines how participants strategically choose to focus on important information and how participants learn to accurately predict memory for the selected information. To examine this, we used a value-directed remembering task (Castel, Benjamin, Craik, & Watkins, 2002), in which participants are presented with lists of words paired with varying point values, and are told the point value indicates how much the word is worth. Words are shown one at a time, followed by an immediate free-recall task. Studies utilizing this task have found that while older adults recalled fewer items compared with younger adults, no age differences in memory were present for the most “valuable” words (e.g., Castel et al., 2002; Castel, Balota, & McCabe, 2009). Importantly, this pattern of selectivity emerged only with multiple study-test trials, suggesting that task experience are a crucial element in the development of successful task strategies (Castel et al., 2009).

In the current task, the value-directed remembering paradigm was modified such that participants, on an item by item basis, had to “bet on” which items they would be able to remember, a form of metacognitive monitoring and control. Thus, for a given item, if a participant “bet” on it, they would receive whatever points were associated with that item if they were able later to recall it, but would lose those points if they failed to recall it. Participants were told the goal was to maximize their score. Thus, there were rewards associated with accurately predicting which items would be recalled, and penalties if one failed to do so. Each list contained 12 different words, and each word was paired with a point value (1-10, 15, and 20). The inclusion of 15 and 20 points were used in order to assess the impact of extreme incentive/loss potential (e.g., Loftus & Wickens, 1970). Furthermore, individuals were told their overall point
score after each list, and engaged in six study-test trials. Thus, this adapted and novel paradigm allows for the investigation into the impact of item importance, and extensive task experience and feedback, on metacognitive judgments and accuracy.

The introduction of negative consequences when one provides inaccurate metamemory judgments during encoding is a departure from classic memory and metacognition paradigms. It introduces an aspect of risk which could influence which items, as well as how many items, one will “bet” on, and could potentially create a more stressful situation which might impact performance. Older adults who believe that their memory abilities are declining or hold negative stereotypes regarding cognitive aging (Levy-Cushman & Abeles, 1998; Levy & Langer, 1994) may be less likely to engage in effective strategies (Hertzog and Hultsch, 2000), and thus may be hesitant to take risks and “gamble” on their ability to remember high reward information. However, the incorporation of incentives, while potentially increasing anxiety, could also enhance participants’ vigilance and motivation to accurately calibrate their bets to their actual performance abilities. Motivation, incentives and accountability have been shown to increase performance on various cognitive tasks (e.g., Germain & Hess, 2007; Hess, Germain, Swaim, & Osowksi, 2009; Touron, Swaim, & Hertzog, 2007) and older adults may particularly benefit from these added incentives (Hess, Rosenberg, & Waters, 2001). Also, by allowing participants to choose which and how many items they bet on, this could increase the sense of control over the situation, and lower anxiety resulting from the threat of negative consequences.

We predict that older adults will learn to selectively bet on high value items despite potential negative consequences associated with failing to recall those items, as some evidence suggests that older adults may be more sensitive to gains and less sensitive to losses (Denburg, Tranel, & Bechara, 2005; Fein, McGillivray, & Finn, 2007; Samanez-Larkin et al., 2007). It is
also likely that older adults will not remember as many words as younger adults on this task, which could result in overall lower “scores” (point totals), and that both younger and older adults may display overconfidence on the first few trials (Rast & Zimprich, 2009). However, we predict that with sufficient task experience, both younger and older adults will become more strategic and differences in calibration between “bets” and actual performance will be greatly reduced, but older adults may take longer to achieve performance levels comparable to younger adults.

Methods

Participants

Participants were 26 older adults (20 females, 6 males; average age = 77.9 years old) and 26 younger adults (22 females, 4 males; average age = 20.0 years old). Older adults were living in the Los Angeles area, and recruited through community flyer postings as well as through the UCLA Cognition and Aging Laboratory Participant Pool. The older adults had good self-reported health ratings ($M = 8.6$ on a scale of 1-10 with 1 indicating extremely poor health and 10 indicating excellent health), and had 16.9 years of education. Older participants were paid $10 an hour for their time and reimbursed for parking expenses. Younger adults were all University of California, Los Angeles undergraduates and received course credit for their participation.

Materials

Seventy-two common nouns were used as stimuli. The log mean hyperspace analog to language (or HAL, a model of semantics which derives representations for words from analysis of text, Burgess & Lund, 1997) average frequency of the words was 8.8 (range = 7.2-10.1), as obtained from the elexicon.wustl.edu Web site (Balota et al., 2007). All of the words were four or five letters in length (e.g., lion, radio, train). The words were randomly assigned without replacement into one of 6 different lists and each list contained 12 words. Within each list, words
were randomly paired with a point value (1-10, 15, and 20). Each point value was only used once within a list and order of the point values within and across lists was varied such that, for example, the 5-point word appeared in a different position on every list. All stimuli were displayed on a computer via a Microsoft PowerPoint presentation.

**Procedure**

Participants were told that they would be presented with six different lists of words, and that each list contained 12 words. They were informed that each word was paired with a number, and that this number indicated how much the word was “worth.” They were told that the values ranged from 1-20. For each word, they were asked to decide if they wanted to “bet” on it. Thus, if the participant said “yes” when the word was presented and they were later able to remember that word on an immediate free recall test, they would receive the points associated with it. However, if they failed to recall the word that they initially bet on, then they would lose those points. If the participant said “no,” points were not gained or lost regardless of whether the word was recalled. Participants were told the goal was to try to get as many points as possible, and were encouraged to try to maximize gains and to minimize any losses.

Participants were shown the word-number pairs one at a time, each for 5 seconds. As each word was presented they had to indicate whether or not they wanted to “bet” on the word by saying “yes” or “no” aloud. Whether the participant chose to bet on the word or not, each word was displayed for 5 seconds. After all 12 words were presented they were given a 20s free recall test in which they had to verbally recall as many words as they could from the list (they did not need to recall the point values). Their responses were recorded by an experimenter. Immediately following the recall period, participants were informed of their score for the list, but were not given feedback about specific items. Scores were calculated by summing the points associated
with the words participants bet on and successfully recalled, and then subtracting the number of points associated with the words that were bet on but not recalled. The next list began immediately after the scores were calculated and the feedback was provided (approximately 15-20 seconds later). The procedure was repeated until all six lists had been completed.

**Results**

In order to examine the influence of value on betting, the average proportion of items bet on by point value (collapsed across all lists) was computed for both older and younger adults (see Figure 1a). Analyses were conducted collapsing across point values in order to maintain power (values 1-4, 5-8, and 9-20 were grouped and created low, medium, and high point value categories), and a 2 (Age Group) x 3 (Point Value - low, medium, high value) Analysis of Variance (ANOVA) was conducted. Although older adults bet on a lower proportion of items than younger adults, $F(1, 50) = 5.47, MSE = .03, p = .02$, both younger and older adults bet on items with higher point values, $F(2, 100) = 151.42, MSE = .03, p < .001$. Furthermore, a marginal interaction was obtained, $F(2, 100) = 2.58, MSE = .03, p = .08$. Post-hoc tests revealed that both younger and older adults bet on high point value words equally as often, $t(50) = .55, p = .59$. However, compared with older adults, younger adults bet on slightly more items with medium, $t(50) = 1.77, p = .08$, and low point values $t(50) = 2.37, p = .02$.

Figure 1b displays the average proportion of items recalled (across all lists) by point value for both younger and older adults. A 2 (Age Group) x 3 (Point Value - low, medium, and high value) ANOVA revealed older adults recalled a lower proportion of items than younger adults, $F(1, 50) = 17.03, MSE = .02, p < .001$, and recall performance was sensitive to value as recall improved for all participants as point values increased, $F(2, 100) = 147.82, MSE = .02, p < .001$. Age Group did not interact with Point Value, $p = .61$. Overall, these results suggest that
both younger and older adults were sensitive to item importance in terms of bets and recall performance, and were engaging in calculated risks in order to maximize performance.

In order to assess how metacognitive judgments and accuracy changed with task experience, the mean number of words bet on and recalled were examined as a function of list. Figure 2a displays the mean number of words bet on and reveals that, initially, older and younger adults bet on a similar number of items but that older adults (and to some extent, younger adults), bet on fewer items with task experience. In order to maintain appropriate power in the analyses, lists were combined such that performance in the beginning, (Lists 1-2), middle (Lists 3-4) and end (Lists 5-6) of the task could be examined. A 2 (Age Group) x 3 (List - beginning, middle, end) ANOVA showed that older adults bet on fewer words than younger adults, $F(1, 50) = 5.40$, $MSE = 4.42$, $p = .02$, and the number of items bet on decreased with task experience, $F(2, 100) = 32.82$, $MSE = .83$, $p < .001$. Furthermore, an interaction, $F(2, 100) = 13.72$, $MSE = .83$, $p < .001$, was observed. In the beginning, older adults bet on a similar number of words as younger adults, $t(50) = 0.37$, $p = .71$, but bet on fewer words in the middle $t(50) = 2.37$, $p = .02$, and end $t(50) = 4.88$, $p < .001$, of the task compared with younger adults.

Figure 2a also shows overall lower recall performance by older adults compared with younger adults, and a pattern of memory improvement across lists. A 2 (Age Group) x 3 (List - beginning, middle, end) ANOVA revealed that older adults recalled fewer words than younger adults, $F(1, 50) = 17.02$, $MSE = 2.56$, $p < .001$. There was a main effect of List, with the number of words recalled increasing on later trials, $F(2, 100) = 11.89$, $MSE = .90$, $p = .001$. There was no interaction between Age Group and List, $p = .53$. Given the results obtained for the items bet on and recalled, it appears that with task experience both younger and older adults were learning to adjust and calibrate their bets to their actual recall performance.
In order to directly examine calibration, the difference between the number of items bet on versus actually recalled was computed for younger and older adults. Ideally, this measure would be zero if number of items bet on equaled the number of items recalled. As shown in Figure 2b, older and younger adults appear to have comparable calibration, with both groups improving with task experience. A 2 (Age Group) x 3 (List - beginning, middle, end) ANOVA confirmed that both age groups demonstrated comparable calibration across lists, \( p = .38 \).

Furthermore, younger and older adults showed substantial improvement with task experience, \( F(2, 100) = 42.68, MSE = 1.42, p < .001 \). A significant interaction was also obtained, \( F(2, 100) = 4.58, MSE = 1.42, p = .01 \). Post hoc tests showed younger adults were initially marginally better calibrated compared with older adults, \( t(50) = 1.82, p = .08 \), no age-related differences in the middle of the task, \( t(50) = .51, p = .62 \), and at the end older adults were marginally better calibrated than younger adults, \( t(50) = 1.73, p = .09 \). Additionally, the ability to calibrate bets to performance increased substantially from the beginning to middle, and from the middle to end of the experiment for both older (all \( p’s \leq .001 \)) and younger adults (all \( p’s \leq .02 \)).

Within this paradigm, participants were told to maximize their score and were given their score at the end of every list. The average scores (a measure of overall performance) for both younger and older adults on all six lists are displayed in Figure 2c. A 2 (Age Group) x 3 (List - beginning, middle, end) ANOVA revealed that, overall, older adults had lower scores than younger adults, \( F(1, 50) = 5.56, MSE = 1008.44, p = .02 \), and that scores increased on the later trials, \( F(2, 100) = 29.25, MSE = 395.99, p < .001 \). Although Age Group x List interaction was not significant, \( F(2, 100) = 2.01, p = .14 \), planned comparisons showed that younger adults had higher scores than older adults on lists 1-2, \( t(50) = 2.32, p = .02 \) and lists 3-4 \( t(50) = 1.96, p = \).
.06. However, younger and older adults had comparable scores on lists 4-6, \( t(50) = .78, p = .44 \),
despite the fact that older adults were able to recall fewer items on these lists.

**Discussion**

The current study examined the role of value in a novel metamemory task that
incorporated consequences associated with the accuracy of metamemory predictions over
multiple study-test trials. The results suggest that both older and younger adults show a strong
improvement in calibration across trials, and also display sensitivity to value. On initial lists,
younger and older adults showed a form of a “metacognitive failure”, betting on more items than
they could recall, and this discrepancy between predictions and performance was larger for older
adults (Figure 2a). However, with experience and feedback, metacognitive calibration greatly
improved. This pattern of initial overconfidence is consistent with previous findings (Rast &
Zimprich, 2009), and suggests that metacognitive awareness and accuracy is something that may
need to be learned with feedback and task experience (Jacoby et al., 2010). Individuals may not
be aware of how much they are capable of recalling in a new situation, but can calibrate and
update expectations to accurately reflect memory performance, a skill that is especially important
for older adults.

This study extends and supports the general finding that metacognitive control processes
may remain relatively intact with advancing age (e.g., Connor et al., 1997; Dunlosky & Hertzog,
1997). For example, older adults appear to exhibit a similar pattern of metacognitive control as
younger adults on later lists. On initial lists, older (and younger) adults initially overestimate
memory capacity but became better calibrated with task experience. In fact, on later lists, the
correlation between bets and recall was well over .85 for both younger and older adults.
Although previous studies using JOLs have found that older adults’ knowledge updating improves to some extent with task experience (e.g., Matvey et al., 2002), most have not found as large of an improvement in calibration as in the current study. However, our measure of metacognitive prediction (i.e., bets), differs from standard JOLs. Betting involves a consequential yes/no decision, as opposed to a more passive assignment of a JOL. In addition, the present task used a relatively small number of unique items on each list, and participants may be better able to monitor capacity and interference under these conditions. It may be the case that smaller deviations in calibration, and thus inaccuracies, are more easily detected when more continuous judgments are used, or when larger sets of information are present. Critically, the current task introduced negative consequences if predictions were inaccurate, and provided feedback about performance (i.e., scores). The addition of these elements likely increased motivation, and could partially explain the relatively high levels of calibration. Furthermore, our task implemented six study-test cycles with feedback regarding scores, whereas others have used only two, with no feedback. Given the results and the pattern of calibration displayed in Figure 2b, it appears that it may take more than two study-test cycles for older adults to achieve accuracy levels on par with younger adults.

While older adults recalled fewer words than younger adults on every trial, other indicators of performance did not show substantial age-related decrements. Specifically, on the later trials older adults’ scores were comparable to younger adults’, even though they recalled fewer words. Given scores were directly related to how much information participants were able to recall \((r = .70, p < .001)\), this finding is somewhat unexpected. It suggests that older adults were implementing strategies that actually led to the marginally better calibration on the later list, in order to achieve goal-relevant outcomes (high point totals). These findings are in
accordance with the notion of selective optimization and compensation, which posits that some older adults can optimize performance by selectively allocating resources and attention within goal-relevant situations (Baltes & Baltes, 1990). Although selecting high reward/high risk options can lead to disadvantageous decisions, in the current study individuals were able to moderate the level of risk by calibrating the number of items bet on to the number of items recalled. Both younger and older adults were able to achieve this balance between risk and reward with task experience, and the introduction of negative consequences likely served to enhance motivation to accurately monitor and update performance expectations.

In conclusion, the present findings suggest that older and younger adults can learn to use value to inform decisions regarding what and how much information to try to remember in order to maximize memory performance. Using a novel metamemory task that incorporated both rewards and penalties associated with monitoring and control, and multiple study-test trials, it was shown that both younger and older adults could learn to accurately maximize performance. Although memory may decline with advancing age, the ability to learn about one’s memory capabilities, and strategically utilize memory resources in order to maximize performance may remain relatively intact with healthy aging.
References


Figure Captions

*Figure 1.* Average proportion of items bet on (top panel A) and recalled (bottom panel B) as a function of point value by older and younger adults. Error bars represent standard error of the mean.

*Figure 2.* Panel A: Average number of items bet on and recalled by younger and older adults; Panel B: Average calibration (number of items bet on minus number of items recalled) for younger and older adults. (Note: ideal calibration score = 0); Panel C: Average score for younger and older adults. Errors bars represent standard error of the mean.