



Published in final edited form as:

*Psychol Aging*. 2012 December ; 27(4): 1053–1065. doi:10.1037/a0027686.

## Age-Related Reduction of the Confidence-Accuracy Relationship in Episodic Memory: Effects of Recollection Quality and Retrieval Monitoring

Jessica T. Wong, Stefanie J. Cramer, and David A. Gallo

University of Chicago

### Abstract

We investigated age-related reductions in episodic metamemory accuracy. Participants studied pictures and words in different colors, and then took forced-choice recollection tests. These tests required recollection of the earlier presentation color, holding familiarity of the response options constant. Metamemory accuracy was assessed for each participant by comparing recollection test accuracy to corresponding confidence judgments. We found that recollection test accuracy was greater in younger than older adults, and also for pictures than font color. Metamemory accuracy tracked each of these recollection differences, as well as individual differences in recollection test accuracy within each age group, suggesting that recollection ability affects metamemory accuracy. Critically, the age-related impairment in metamemory accuracy persisted even when the groups were matched on recollection test accuracy, suggesting that metamemory declines were not entirely due to differences in recollection frequency or quantity, but that differences in recollection quality and/or monitoring also played a role. We also found that age-related impairments in recollection and metamemory accuracy were equivalent for pictures and font colors. This result contrasted with previous false recognition findings, which predicted that older adults would be differentially impaired when monitoring memory for less distinctive memories. These and other results suggest that age-related reductions in metamemory accuracy are not entirely attributable to false recognition effects, but also depend heavily on deficient recollection and/or monitoring of specific details associated with studied stimuli.

### Keywords

metamemory; aging; recollection; monitoring; distinctiveness

---

Older adults frequently report problems with their memory, potentially demonstrating accurate metamemory or self-awareness of age-related cognitive decline (Hultsch, Hertzog, & Dixon, 1987; Jonker, Geerlings, & Schmand, 2000; Ryan, 1992) However, although metamemory should reflect actual memory ability, metamemory accuracy in aging also might be affected by other factors, such as the ability to accurately engage metacognitive monitoring processes when making metamemory decisions (e.g. Flavell, 1971; Koriati & Goldsmith, 1996; see Dunlosky & Connor, 1997; Kuhlman & Touron, 2011). Understanding

---

Corresponding author: Jessica T. Wong Department of Psychology University of Chicago Chicago, IL 60637 jtwong@uchicago.edu  
Tel: 773-702-4376 Fax: 773-702-0886.

**Publisher's Disclaimer:** The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at [www.apa.org/pubs/journals/pag](http://www.apa.org/pubs/journals/pag)

how these different factors may affect metamemory accuracy with aging is a topic with significant implications. For example, older adults may need to accurately assess their own memory abilities in order to seek help for age-related cognitive decline.

Metamemory ability is often studied using episodic memory tasks, such as item-by-item confidence judgments on recognition memory tests. There is considerable evidence that retrospective confidence judgments made at retrieval can be less accurate in older adults than in younger adults (Chua, Schacter, & Sperling, 2009; Dodson & Krueger, 2006; Gallo, Foster, & Johnson, 2009; Gopie, Craik, & Hasher, 2010; Jacoby, Wahlheim, Rhodes, Daniels, & Rogers, 2010; Karpel, Hoyer, & Tolia, 2001; Kelley & Sahakyan, 2003; Norman & Schacter, 1997; for related evidence from the “feeling of knowing” task, see Souchay, Isingrini, & Espagnet, 2000; Thomas, Bulevich, & Dubois, 2011). In contrast, tasks that require participants to prospectively rate the expected memorability of study items (i.e., judgments of learning) have generally found age-related sparing (Connor, Dunlosky, & Hertzog, 1997; Dunlosky, Baker, Rawson, & Hertzog, 2006; see Hertzog, 2002), as have metamemory judgments on general knowledge or semantic memory tasks (e.g., Dahl, Allwood, & Hagberg, 2009; Lachman, Lachman, & Thronesbery, 1979; Pliske & Mutter, 1996). These findings suggest that age-related metamemory problems may be specific to retrospective assessments of episodic memory retrieval, as measured by confidence judgments.

Here we consider two interrelated factors that might affect the age-related reduction in confidence judgment accuracy in episodic memory. The first factor is age-related reductions in the ability to recollect studied information (e.g., Balota, Dolan, & Duchek, 2000; Nilsson, 2003). According to dual process theories of memory, aging can impair the recollection of specific features associated with previously encountered stimuli, but aging is less likely to impair familiarity, or a more general feeling that stimuli were previously encountered in the absence specific recollections (Jacoby & Rhodes, 2006; Jennings & Jacoby, 1993; Parkin & Walter, 1992; Prull, Dawes, Martin, Rosenberg, & Light, 2006). In the context of a recognition memory test, an age-related recollection impairment could reduce the number of studied items that are retrieved (i.e., recollection quantity) as well as the number of unique features retrieved for each studied item (i.e., recollection quality, see Scimeca, McDonough, & Gallo, 2011). This recollection impairment would restrict the range of detailed features that could be accurately retrieved for studied items, features that otherwise would be used to differentiate correct responses (target recognition) from incorrect responses (lure recognition) when making confidence judgments.

The second factor that may contribute to age-related reductions in confidence judgment accuracy is impaired metacognitive monitoring processes. This concept partly draws from the source monitoring framework, which distinguishes between memory retrieval, on the one hand, and monitoring processes operating on retrieved information to make memory attributions or decisions, on the other (Johnson, Hashtroudi, & Lindsay, 1993; see also Nelson & Narens, 1990; Pansky, Goldsmith, Koriati, & Perlman-Avni, 2009). Although recollection and monitoring impairments can be tightly linked with each other (cf. Kelley & Sahakyan, 2003), metacognitive monitoring deficits could cause less diagnostic information to corrupt confidence judgments even if accurate recollections were otherwise available. Such information might include the potentially misleading effects of familiarity (Jacoby, Bishara, Hessels & Toth, 2005) or the recollection of noncritical or irrelevant information (Brewer, Marsh, Clark-Foos & Meeks, 2010). More generally, monitoring might be impaired by inappropriate anxiety about one’s memory abilities, which could unnecessarily truncate the range of confidence judgments (i.e., over-confidence for people that worry too little, and under-confidence for those that worry too much), as well as declines in working memory (e.g. Daniels, Toth, & Jacoby, 2006; Souchay & Isingrini, 2004), which could

reduce one's ability to compare a recollected item to other items in memory when making confidence judgments. All of these possibilities could increase the likelihood that less accurate information would be used to differentiate correct responses (target recognition) from incorrect responses (lure recognition) when making confidence judgments.

## Misrecollection and False Recognition

Several studies by Dodson and colleagues provided additional insight into the factors that might cause age-related metamemory impairments (Dodson & Krueger, 2006; Dodson, Bawa & Krueger, 2007a; Dodson, Bawa & Slotnick, 2007b). The novel feature of these studies was to match younger and older adults on overall accuracy on a source memory test, such as the ability to recollect the voice that had earlier spoken a particular sentence, by introducing a longer retention interval in younger adults. This matching is theoretically important, because otherwise differences in confidence judgment accuracy across groups could be driven by differences in guessing rates (we elaborate on this idea more in the Discussion section). Even under these matched source accuracy conditions, Dodson and colleagues found that source memory errors were made with greater confidence in older adults than younger adults. They argued that their results supported a misrecollection hypothesis, whereby aging elevates the creation of high-confidence false recollections (e.g., Dodson et al., 2007a; see also Gallo et al., 2009; Norman & Schacter, 1997).

The false recollection account ultimately appeals to age differences in recollection quality, but it is important to note that these differences could be the result of multiple deficits in older adults. False recollection could result from the impoverished encoding and/or retrieval of coherent representations for studied items (i.e., binding deficits, Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000), as well as a greater likelihood of constructing illusory recollections from these fragmented recollections at retrieval (i.e., monitoring deficits, see Gallo & Roediger, 2003; Lampinen, Meier, Arnal, & Leding, 2005; Lyle & Johnson, 2006). The fact that the age groups were matched on their ability to recollect source information in the aforementioned studies is consistent with a monitoring deficit. However, it also is possible that aging spares monitoring abilities, and instead confidence judgments were simply more sensitive than source memory judgments to age-related differences in recollection quality caused by other processes, such as a binding deficit that automatically creates false recollections. This interpretation does not preclude monitoring differences, but it also does not require them (for a relevant signal detection model, see Dodson et al., 2007b).

One way to further investigate age differences in confidence judgment accuracy is to independently manipulate the quality of the to-be-recollected details. In the false recognition literature, it is well documented that both younger and older adults are less susceptible to false recognition errors when they are tested on stimuli that elicit qualitatively richer or more distinctive recollections (e.g., pictures versus words, see Dodson & Schacter, 2001; Gallo, Cotel, Moore, & Schacter, 2007; Schacter, Israel, & Racine, 1999). For example, Gallo et al. (2007) found that both age groups were more prone to false recognition on a test that required the recollection of font color, relative to a test that required the recollection of pictures (which had more unique visual details across items). Following Schacter et al. (1999) and Dodson and Schacter (2001), they argued that participants expected more distinctive or higher-quality recollections when retrieval was oriented towards pictures than font color. Because lures were unlikely to elicit distinctive recollections, participants were better able to reject them via a diagnostic retrieval monitoring process. Gallo et al. (2007) also found that age-related increases in false recognition were greatest on the font color test, suggesting that older adults were more prone to false recollection and/or familiarity-based guessing when monitoring memory for less distinctive recollections.

The finding that older adults were more impaired when monitoring memory for font color than for pictures raises the question as to whether similar mechanisms may be involved in making confidence judgments. False recognition and confidence judgments are both thought to tap similar monitoring processes, such as comparing retrieved information to one's expectations of what one should retrieve. In fact, the distinctiveness heuristic has been associated with a change in one's response criteria (e.g., Schacter et al., 1999), and confidence judgments often are assumed to reflect different response criteria in signal detection theories (e.g., Rotello & Macmillan, 2008). By extension, increasing recollection quality might minimize age-related differences in the confidence-accuracy relationship, much like it minimizes age-related differences in false recognition.

## The Current Study

To test these ideas in the current study we developed a cued-recollection task that allowed us to compare recollection and confidence judgment accuracy for different kinds of stimuli. Participants studied stimuli that were high in recollection quality (i.e., object pictures presented in full color or as line drawings) or low in recollection quality (i.e. words presented in red or blue font color), and then took two-alternative forced-choice (2AFC) recollection tests followed by confidence judgments. All test words were presented in a neutral font, and each word in the test pair was studied once, so that the targets and the lures were equally familiar on average. Participants had to choose the test word that was previously presented as a colored picture (picture test) or in blue font (font test). Thus, this test required participants to recollect specific details in order to make an accurate decision (i.e., picture color on the picture test and font color on the font test), so that each trial would either result in the successful recollection of the criterial information (leading to an accurate response) or not (leading to responses based on guesses or nondiagnostic information, with chance performance at 50%). These testing conditions should have minimized age-related differences in the use of familiarity (e.g., Kelley & Sahakyan, 2003) or in the setting of a yes/no response criterion (e.g., Pansky et al., 2009) when making the recollection memory judgment, either of which might complicate interpretations of confidence judgment accuracy. By removing individual differences in yes/no response bias and by precluding familiarity as an accurate basis for the 2AFC judgment, this task provides a straightforward assessment of the relationship between recollection accuracy and subsequent confidence judgments in younger and older adults.

The current experiment had two specific aims. The first aim was to compare the confidence-accuracy relationship in younger and older adults using a task that explicitly required the recollection of specific information. We predicted that older adults would have impaired recollection on this kind of task relative to younger adults, as well as less accurate confidence judgments. These metamemory deficits might be driven by reduced confidence to correct responses, due to the forgetting of specific details associated with studied items, as well as elevated confidence to incorrect responses, due to the false recollection of specific details attributed to lures. We also predicted that these age-related impairments would persist even when recollection accuracy was artificially matched between the two age groups (via a divided attention manipulation), thereby controlling for group differences in recollection quantity and corresponding guessing rates. These findings would conceptually replicate the results of Dodson and Krueger (2006) and Dodson et al. (2007a, 2007b), implicating a role for age-related differences in recollection quality and/or monitoring.

The second aim was to investigate the role of recollection quality using a stimulus manipulation that does not depend on artificially matching the age groups on accuracy. We predicted that both age groups would show lower confidence judgment accuracy for font colors than for pictures, analogous to the distinctiveness effects observed in false recognition

tasks. The relatively impoverished recollections for font colors should reduce confidence for correctly recognized targets, and also might enhance confidence to inaccurately recognized lures to the extent that these errors are more likely to be based on false recollections. Critically, if the processes that drive age-related false recognition effects also influence confidence judgments in the same way, then older adults should be more impaired in confidence judgments for less distinctive stimuli. Confidence judgments for font colors should be more susceptible to false recollection and/or familiarity than pictures, especially in older adults. In contrast, processes other than false recollection and familiarity may reduce confidence judgment accuracy in older adults, such as a general reduction in the ability to recollect and monitor specific details associated with any kind of studied stimuli. In this case, the age-related impairment in confidence judgment accuracy might not interact with the manipulation of recollection quality, especially if older adults are equally impaired in their ability to recollect specific details from these different kinds of stimuli.

## Method

### Participants

There were 56 older adults and 112 younger adults in the study. The older adult participants (aged 65-90 years,  $M = 77.72$ ,  $SD = 0.98$ ) lived independently and reported no physical or mental problems that impaired daily functioning. They were recruited from the Chicago metropolitan area and were paid for participation. They scored high on the Mini Mental State Exam (Folstein, Folstein, & McHugh, 1997;  $M = 28.62$ ,  $SD = 1.25$ ), and scored low on the Geriatric Depression Scale (Brink, Yesavage, Lum, Heersema, & Adey, 1982;  $M = 3.36$ ,  $SD = .48$ ). Fifty-six younger adults (aged 18-25 years,  $M = 19.73$ ,  $SD = 1.3$ ) were tested under full attention and an additional 56 younger adults (aged 18-22 years,  $M = 19.69$ ,  $SD = 0.21$ ) were tested under divided attention. All the younger adults were recruited from the University of Chicago Psychology Department participant pool and received course credit or payment.

### Materials and Design

The pool of stimuli consisted of 192 pictures of common objects and their corresponding verbal labels. The pictures depicted simple, everyday items as either black and white line drawings or as colored pictures. The black and white line drawings were taken from Snodgrass and Vanderwart (1980) and supplemented by Szekely et al. (2005) as well as various Internet sources. These drawings were modified so that all the objects appeared on a white background. The colored pictures were taken from public domain Internet sites and cropped to display each object on a neutral background. All images were formatted to be about the same size. The verbal labels were one-word text stimuli presented in either red or blue font. All of the stimuli were unique and did not have a large degree of conceptual overlap (e.g. “guitar” was included in the stimuli set, but “banjo” was not).

The task consisted of two study blocks (i.e. picture block, font color block) and two test blocks. The block order was counterbalanced between participants. The picture block contained 96 randomized pictures (48 black and white line drawings, 48 colored pictures) and the font color block contained 96 randomized words (48 in red font, 48 in blue font). All the stimuli were counterbalanced so that each one was shown in each of the four different study conditions. The order of the study blocks was counterbalanced across participants, and the order of the test blocks followed that pattern (e.g. if shown the font color study block first and the picture test second, then they completed the font test first and the picture test second). The younger adults completed both study blocks first, followed by both test blocks, whereas the task was made easier for older adults by having them complete one study and test cycle first, followed by the second study and test cycle. Pilot testing indicated that this

procedural difference would increase the likelihood that both age groups would perform in an intermediate range on the recollection tests (i.e., avoiding ceiling effects in younger adults or floor effects in older adults), which is important for calculating metamemory measures. Note that these differences in task difficulty might have affected the absolute level of confidence judgments across groups (e.g., overall greater confidence on easier tasks), but they should not have affected metamemory measures assessing the extent that each individual's confidence judgments tracked accuracy differences across trials (i.e., within-task variability). Using an easier version of the task in older adults also reduced group differences in guessing rates, which is important for comparing metamemory measures.

## Procedure

All stimuli were presented on the computer and the experimenter entered the responses for the older adults. On each study trial, the name of an item appeared in black text in the center of the computer screen. To ensure deep processing of the label, participants had to decide whether or not the item was pleasant (i.e. yes or no judgment; self-paced). After making their pleasantness judgment, they either saw a picture of that item or the same label again in colored font, depending on the block. Each picture was presented for 500 ms and each colored word was presented for 3000 ms. The duration of the words was longer than that of the pictures in order to help improve memory for font color.

To better match younger and older adults on recollection, one group of younger adults studied the to-be-remembered stimuli under divided attention. While viewing the labels and stimuli, participants repeated aloud random digits spoken every two seconds (not locked to stimulus onset). Each black label was presented for 1500 ms and they did not make pleasantness judgments. Errors on the divided attention task were very rare ( $M < 1$  per block).

Memory was tested using a self-paced 2AFC format. There were two types of tests: the picture test and the font test. On the picture test, participants were presented with 48 pairs of black words. One word corresponded to a previously studied colored picture and the other corresponded to a previously studied line drawing. Participants had to decide which one of the two labels was studied as a colored picture. After making each memory decision, participants rated their confidence in their answer according to the following scale: 50 (chance), 60, 70, 80, 90, and 100 (certain). It was explained to participants that responses based on guessing would lead to chance performance (50%), so they should choose 50% confidence if they were completely guessing. On the font test, participants saw 48 pairs of black labels, one label corresponding to a red word and the other corresponding to a blue word. Participants decided which label was studied in blue font and then rated their confidence.

Three confidence-accuracy measures were calculated: the Goodman-Kruskal gamma correlation (Goodman & Kruskal, 1954) to measure confidence resolution, a calibration error score, and a confidence-accuracy discrimination score. We found a large degree of agreement between all three measures. The gamma correlation evaluates relative metamemory accuracy in terms of one's ability to differentiate accurate and inaccurate 2AFC test trials with the confidence scale. Since gamma is based on ordering differences between items, it is recommended for ordinal-scaled variables (Gonzalez & Nelson, 1996; Nelson, 1996; Nelson, 1984). Larger gamma scores denote better metamemory accuracy. The calibration error score calculates the absolute difference between actual accuracy and predicted accuracy (i.e. confidence), as a function of the frequency of 2AFC trials that receive the particular confidence rating (e.g. Dodson et al., 2007a). Perfect calibration occurs when there is a match between actual and predicted accuracy (e.g., the items with a

70% confidence rating have an average accuracy of 70%), and smaller calibration error scores denote better metamemory. Finally, the discrimination score subtracts the average confidence rating for incorrect 2AFC trials from that for correct trials, under the assumption that greater metamemory accuracy should be reflected in higher confidence judgments for correct test responses compared to incorrect responses. We also separately analyzed the confidence associated with correct and incorrect responses as part of the discrimination analysis.

## Results

The results are presented in six sections. The first two sections present analyses of recollection accuracy obtained on the 2AFC judgment, as well as the overall distribution of confidence judgments. The next three sections present analyses of our metamemory measures of confidence judgment accuracy (resolution, calibration, and discrimination). For each of these metamemory measures we report three different analyses. The first analysis compared performance between the two age groups under the full attention conditions ( $n = 56$  per group), on which age-related impairments in 2AFC recollection accuracy were observed. The second analysis compared groups that were artificially matched on 2AFC recollection accuracy on both tests by excluding some participants ( $n = 36$  per group), as described in the next section. We also directly compared the picture test for all of the older adults and all of the younger adults (divided attention), on which 2AFC accuracy was successfully matched without excluding participants. To anticipate, we found similar patterns of results across our metamemory measures with all three of these analyses, and all three of our measures of metamemory accuracy were correlated with each other.<sup>1</sup> In the final results section we report correlations between recollection and metamemory accuracy across individuals. Unless otherwise specified, all analyses used the traditional  $p < .05$  significance level.

### Recollection Accuracy

Proportions of correct 2AFC trials for the two full attention groups are displayed in Figure 1 (left panel). A 2 (age group: younger, older)  $\times$  2 (test: font, picture) ANOVA on these data revealed an effect of age group,  $F(1, 110) = 56.683$ ,  $MSE = .02$ ,  $p < .001$ ,  $\eta^2_p = .340$ , as younger adults had superior recollection accuracy to older adults. There also was an effect of test,  $F(1, 110) = 191.013$ ,  $MSE = .01$ ,  $p < .001$ ,  $\eta^2_p = .635$ , and no interaction, indicating that recollection accuracy was greater on the picture test than the font test in both age groups. This enhanced recollection accuracy for pictures over font color is consistent with Gallo et al. (2007), although we found the same age-related decline on each test in the current experiment, likely because both tests were designed to require the recollection of fine-grained detail (i.e., the color of either a word or a picture). As expected, older adults had difficulty on the font test (0.57), but nevertheless performed significantly above chance (0.50),  $t(55) = 4.375$ ,  $p < .001$ . A similar ANOVA comparing the older adults to the younger adult divided attention group revealed an age group  $\times$  test interaction,  $F(1, 110) = 10.6$ ,  $MSE = .01$ ,  $p < .01$ ,  $\eta^2_p = .09$ . For the picture test, younger adults under divided attention had comparable memory as the older adults (0.76 versus 0.74,  $t(78) < 1$ ), whereas age-related recollection declines persisted on the font test (0.68 for younger adults versus 0.57 for older adults,  $t(110) = 4.33$ ,  $SEM = .03$ ,  $p < .001$ ). Dividing attention attenuated the picture superiority effect in younger adults, potentially because pictures had more details to encode than font color.

<sup>1</sup>Across all participants ( $n = 168$ ) and collapsing across tests, the three measures of metamemory accuracy were significantly related to each other (resolution and discrimination,  $r = +.76$ , resolution and calibration,  $r = -.49$ , discrimination and calibration  $r = -.50$ , all  $p$ 's  $< .001$ ). These relationships remained significant when each group was analyzed separately (all  $p$ 's  $< .01$ ).

In order to match recollection accuracy on both the picture test and the font test for subsequent metamemory analyses, we also selectively analyzed a subset of older adults and a subset of younger adults in the divided attention condition ( $n = 36$  per group).<sup>2</sup> As can be seen in Figure 1 (right panel), we were able to match accuracy across the age groups on both the font test and the picture test, and only the effect of test persisted,  $F(1, 70) = 70.81$ ,  $MSE = .007$ ,  $p < .001$ ,  $\eta^2_p = .50$ . This matching procedure avoided potential complications in metamemory analyses owing to accuracy differences or extreme accuracy scores (ceiling or floor). It is important to note that excluding participants in this way disrupted the random assignment process, and as such may have introduced unintended subject-selection artifacts (i.e. removing poor performing older adults and well performing younger adults). Nevertheless, analysis of age-related effects on metamemory performance in the matched groups yielded similar results to the full groups. This outcome is theoretically important, because it suggests that the group differences observed when analyzing the full groups could not be attributed entirely to group differences in recollection accuracy.

### Confidence Distributions

Table 1 shows the mean proportion of total test trials that were assigned into each of three confidence judgment bins (low 50/60, medium 70/80, and high 90/100). Because these proportions are inter-dependent, we focus on the high-confidence bin in our analysis. On the font test, older adults made fewer high-confidence responses (.15) than did the younger adults (.36 full, .32 divided, both  $p$ 's  $< .001$ ). This difference likely reflects group differences in overall recollection ability (younger  $>$  older), and when font test accuracy was compared across the matched groups, the difference between younger and older adults was reduced and no longer significant (.25. vs. .17,  $p = .1$ ). On the picture test, younger adults in the full attention group were more likely to use the high confidence than the other two groups (.72 vs. .50 and .46, both  $p$ 's  $< .001$ ), again tracking differences in recollection ability across the groups. When we compared the younger divided attention group to the older adults on the picture test, the distribution of confidence use was similar for both the unmatched and matched groups, all  $p$ 's were not significant. The similarity of these confidence judgment distributions across the groups, especially when matched on accuracy, indicate that any corresponding differences in metamemory accuracy are not likely due to fundamental differences in the understanding or use of confidence judgments across the groups.

### Confidence Resolution

The average gamma correlations for the full attention groups are presented in Figure 2 (left panel).<sup>3</sup> A 2 (age group: younger, older)  $\times$  2 (test: font, picture) ANOVA on the full attention data revealed an effect of age group,  $F(1, 100) = 15.92$ ,  $MSE = .12$ ,  $p < .001$ ,  $\eta^2_p = .14$ , as older adults had lower gamma correlations than the younger adults. There also was an effect of test,  $F(1, 100) = 48.44$ ,  $MSE = .09$ ,  $p < .001$ ,  $\eta^2_p = .33$ , and no interaction ( $F(1, 100) = .009$ ,  $p > .05$ ), as both groups had higher gamma correlations for the picture test than the font test. Consistent with our predictions, the confidence-accuracy relationship was reduced in older adults relative to younger adults, and also for font color relative to pictures. However, the effect of age group did not interact with recollection quality, which is inconsistent with the idea that lower quality recollections would enhance the age-related susceptibility to false recollection when making confidence judgments.

<sup>2</sup>For this analysis, we selectively analyzed participants with font test accuracy between 54% and 83% (younger adults) and 54% and 85% (older adults), and picture test accuracy between 60% and 94% (younger adults) and 56% and 94% (older adults).

<sup>3</sup>Gamma scores for pictures were not computed for one older adult and eight younger adults in the full attention condition because there was no variability in either accuracy (i.e. all answers were correct) or confidence ratings (i.e. same confidence rating for all trials). In addition, gamma for font color was not computed for one younger adult under full attention.



Analysis of the groups that were matched on 2AFC recollection accuracy on each test (Figure 2, right panel) revealed the same pattern as the unmatched groups, with an effect of age group,  $F(1, 70) = 4.18$ ,  $MSE = .11$ ,  $p < .05$ ,  $\eta^2_p = .06$ , and test,  $F(1, 70) = 23.14$ ,  $MSE = .08$ ,  $p < .001$ ,  $\eta^2_p = .25$ , and no interaction,  $F(1, 70) = .931$ ,  $p > .05$ . A direct comparison of the gamma correlation on the picture test for all of the older adults (.51) and all of the younger adults in the divided attention condition (.63) also showed this age group difference,  $t(109) = 2.54$ ,  $SEM = .05$ ,  $p < .05$ , again demonstrating reduced metamemory accuracy in older adults when 2AFC accuracy was matched. These findings are consistent with the results of Dodson and Krueger (2006) and Dodson et al. (2007a). As in the full dataset, though, the lack of an interaction between age and recollection quality is inconsistent with the idea that lower quality recollections would enhance the age-related susceptibility to false recollection.

### Confidence Calibration

The average calibration error scores for the full attention groups are presented in Figure 3 (left panel). A 2 (age group: younger, older) x 2 (test: font, picture) ANOVA on the full attention data revealed an effect of age group,  $F(1, 110) = 7.54$ ,  $MSE = .01$ ,  $p < .01$ ,  $\eta^2_p = .06$ , as older adults had greater calibration error scores than the younger adults. There also was an effect of test,  $F(1, 110) = 17.837$ ,  $MSE = .003$ ,  $p < .001$ ,  $\eta^2_p = .14$ , as calibration error scores were lower on the picture test than the font test, and an age group x test interaction,  $F(1, 110) = 4.317$ ,  $MSE = .003$ ,  $p < .05$ ,  $\eta^2_p = .04$ , as the effect of age group was greater for the picture test. This interaction was not expected, but it again was inconsistent with the idea that lower quality recollections would enhance the age-related susceptibility to false recollection. Similar patterns were found in the matched groups (Figure 3, right panel), but none of the effects reached significance. However, a direct comparison of calibration error scores on the picture test for all of the older adults (.14) and all of the younger adults (.11, divided attention) revealed a marginal difference,  $t(110) = 1.75$ ,  $SEM = .01$ ,  $p = .08$ , again suggesting reduced metamemory in older adults when 2AFC accuracy was matched.

### Confidence Discrimination

Average confidence judgments for correct and incorrect 2AFC responses for the full attention groups are presented in Figure 4.<sup>4</sup> The first point to take is that the average confidence judgment was significantly greater than 50% (guessing) even for incorrect responses (all  $p$ 's  $< .001$ ), suggesting that participants had based many of their incorrect decisions on the retrieval of inaccurate or false information. To analyze these responses we calculated discrimination scores, or the difference in confidence between correct and incorrect responses (correct minus incorrect). A 2 (age group: younger, older) x 2 (test: font, picture) ANOVA on the full attention discrimination scores revealed an effect of age group,  $F(1, 101) = 41.28$ ,  $MSE = 95.77$ ,  $p < .001$ ,  $\eta^2_p = .29$ , as older adults had lower confidence discrimination scores than younger adults. There also was an effect of test,  $F(1, 101) = 76.89$ ,  $MSE = 47.49$ ,  $p < .001$ ,  $\eta^2_p = .43$ , and no interaction, as both groups had higher confidence discrimination scores for the picture test than the font test. Analysis of the groups that were matched on 2AFC recollection accuracy on each test (Figure 5) revealed the same pattern, with an effect of age,  $F(1, 70) = 22.85$ ,  $MSE = 63.50$ ,  $p < .001$ ,  $\eta^2_p = .25$ , and test,  $F(1, 70) = 48.11$ ,  $MSE = 38.00$ ,  $p < .001$ ,  $\eta^2_p = .41$ , and no interaction. A direct comparison of confidence discrimination scores on the picture test for all of the older adults (9.99) and all of the younger adults (16.89, divided attention) also was significant,  $t(109) = 4.82$ ,  $SEM = 1.43$ ,  $p < .001$ , again demonstrating reduced metamemory accuracy in older

<sup>4</sup>Discrimination scores on the picture test were not computed for one older adult and eight younger adults in the full attention condition because they had no errors.

adults when 2AFC accuracy was matched. Overall, these effects were consistent with those obtained with the other metamemory measures.

We also separately analyzed the average confidence judgments for correct and incorrect responses, using a 2 (age group: younger, older) x 2 (test: font, picture) x 2 (response: correct, incorrect) ANOVA. When comparing the two full attention groups, there was an effect of age,  $F(1, 101) = 9.05$ ,  $MSE = 315.15$ ,  $p < .01$ ,  $\eta^2_p = .08$ , test,  $F(1, 101) = 101.12$ ,  $MSE = 104.31$ ,  $p < .001$ ,  $\eta^2_p = .50$ , and response,  $F(1, 101) = 252.96$ ,  $MSE = 47.89$ ,  $p < .001$ ,  $\eta^2_p = .72$ , and these were qualified by two interactions. First, there was an age x response interaction,  $F(1, 101) = 41.28$ ,  $MSE = 47.89$ ,  $p < .001$ ,  $\eta^2_p = .29$ , as older adults were less confident than younger adults for correct decisions on each test (both  $p$ 's  $< .001$ ), with no age differences in confidence to incorrect responses. These findings are inconsistent with the false recollection hypothesis, which predicts greater high-confidence errors in older adults. Second, there was a test x response interaction,  $F(1, 101) = 76.89$ ,  $MSE = 23.74$ ,  $p < .001$ ,  $\eta^2_p = .43$ , as both groups showed greater confidence on the picture test than on the font color test, and these effects were larger for correct than incorrect responses (albeit significant for both, all  $p$ 's  $< .05$ ). The effect on correct responses is consistent with the idea that picture recollections were more distinctive than font color recollections, leading to greater confidence judgments. The finding that incorrect responses were associated with greater confidence on the picture test than on the font test is inconsistent with the idea that false recollection would be greater on the font test than on the picture test.

When comparing the groups that were matched on recollection accuracy, there again were effects of test,  $F(1, 70) = 57.03$ ,  $MSE = 96.83$ ,  $p < .001$ ,  $\eta^2_p = .45$ , and response,  $F(1, 70) = 261.47$ ,  $MSE = 31.75$ ,  $p < .001$ ,  $\eta^2_p = .79$ , as well as the group x response,  $F(1, 70) = 22.85$ ,  $MSE = 31.75$ ,  $p < .001$ ,  $\eta^2_p = .25$ , and test x response interactions,  $F(1, 70) = 48.11$ ,  $MSE = 19.00$ ,  $p < .001$ ,  $\eta^2_p = .41$ . There was no group difference in confidence for correct responses on the picture test (both means = .83), although on the font test there was a trend for lower confidence to correct responses in older adults (.68) than younger adults (.73, divided attention),  $t(70) = 1.92$ ,  $SEM = 2.57$ ,  $p = .06$ . For incorrect responses, older adults gave significantly higher confidence on the picture test (.71) than younger adults (.66, divided attention),  $t(70) = 2.09$ ,  $SEM = 2.58$ ,  $p < .01$ , and this effect also was significant when all of the participants from these groups were included ( $t(109) = 2.26$ ,  $SEM = 2.04$ ,  $p < .05$ ). This finding is consistent with the false recollection hypothesis, but there was no age difference in confidence for incorrect responses on the font test (.64 and .63, respectively,  $t < 1$ ), which is inconsistent with that hypothesis, as well as the idea that age-related false recollection differences should be greatest when recollection quality is low. As with the full dataset, the test x response interaction indicated that confidence was greater on the picture test than on the font test, particularly for correct responses, although the effect was significant for both correct and incorrect responses (all  $p$ 's  $< .05$ ).

As an alternative to the preceding analysis, we also separated the proportion of correct and incorrect test trials that were assigned to each of three different confidence bins (see Table 2). Analysis of the high-confidence responses (i.e., the 90/100 bin) yielded similar results and conclusions to those reported above, and for simplicity we summarize the key findings here. On the font test, older adults made fewer high-confidence correct responses than younger adults, comparing either the full attention groups (.17 vs. .42,  $p < .001$ ) or the matched groups (.19 vs. .32,  $p < .05$ ), but there was no evidence that older adults made more high-confidence errors than younger adults on the font test, and in fact, older adults made fewer errors than the younger adults (.10 vs. .18 in the full attention groups,  $p < .01$ , both means = .11 in the matched groups). On the picture test, older adults again made fewer high-confidence correct responses than younger adults in full attention (.52 and .76,  $p < .001$ ), but this difference was eliminated in the matched groups (.54 and .55,  $p = .84$ ). Neither of the

group differences in high-confidence incorrect responses on the picture test was significant (both means = .26 in the full attention groups, .27 vs. .18 in the matched groups,  $p = .13$ ). Thus, as with the preceding analysis, these analyses provided little evidence for an age-related increase in high-confidence errors on the picture test, and no evidence for such an increase on the font test.<sup>5</sup>

### Recollection-Metamemory Correlations

To further explore the relationship between recollection and metamemory accuracy, we correlated 2AFC recollection accuracy with each of the metamemory measures across participants. These correlations were calculated separately for the picture test and font test, and also for each of the three experimental groups, resulting in 18 possible correlations. As can be seen from Table 3, each of these correlations was significant at  $p < .05$  (uncorrected), and almost all of them (15 of 18) remained significant after a Bonferroni correction. These correlations demonstrate that recollection accuracy was positively related to metamemory accuracy at the individual level, potentially because recollection quality affected metamemory accuracy. By this account, participants with greater recollection were more likely to make accurate 2AFC responses, and they also were more likely to recollect detailed features for studied items that helped them differentiate correct from incorrect responses with confidence judgments.

We also correlated 2AFC recollection accuracy with the average confidence assigned to correct and incorrect responses, across participants. If participants with greater recollection were more likely to recollect detailed features for studied items, and if these features influenced confidence judgments, then we would expect positive correlations between recollection accuracy and the average confidence judgment assigned to correct responses. As can be seen in Table 4, this relationship was observed for all of the relevant comparisons for correct responses, supporting the idea that recollection accuracy affected confidence judgments. In contrast, the relationship between recollection accuracy and the average level of confidence assigned to incorrect responses was smaller or non-existent, most likely because these responses were made in the absence of accurate recollection (i.e., guesses based on noncritical information, such as familiarity or false recollection).<sup>6</sup> To the degree that incorrect responses were based on a mixture of these different kinds of information (e.g., familiarity or false recollection), one would not necessarily expect a strong correlation between recollection accuracy and confidence for incorrect responses. Whereas familiarity-based guesses should elicit low-confidence errors, false recollection should elicit high-confidence errors, and both processes may have increased with decreases in successful recollection.

### Discussion

We found that the age-related impairment in episodic memory confidence judgment accuracy was strongly associated with declines in recollection. Older adults had reduced recollection and metamemory accuracy relative to younger adults in the full attention conditions (cf. Kelley & Sahakyan, 2003), and they also demonstrated reduced confidence

<sup>5</sup>A reviewer wondered whether our use of shorter study/test blocks in older adults limited our ability to find age-related differences in high-confidence errors, so that making the task harder in older adults would have increased high-confidence errors. We do not believe this is the case for two reasons. First, although we used shorter study/test blocks in older adults, they still found the task more difficult than younger adults. Second, our other manipulations suggested that making the task harder actually decreased high-confidence errors (i.e., the font test relative to the picture test, or the younger divided relative to the full attention conditions, see Table 2).

<sup>6</sup>In most of these conditions the individual variability in incorrect responses was similar to that for correct responses (see Figure 4), so that differences in variability are unlikely to explain the different patterns of correlations between correct and incorrect responses within each group. In contrast, because there were group differences in overall accuracy and confidence (especially for the font test), we avoid interpretation of the size of these correlations across the groups.

for correct trials. Further evidence for the association between recollection and metamemory accuracy was that metamemory accuracy was greater for pictures than font colors in all groups, and within each group individual differences in recollection accuracy correlated with metamemory accuracy and also with the average level of confidence assigned to correct responses. As discussed in the Introduction, age-related reductions in metamemory accuracy may be due to impaired recollection and/or monitoring processes, either of which could affect the ability to retrieve and use detailed features that can accurately differentiate correct from incorrect responses when making confidence judgments.

We also found age-related impairments in metamemory accuracy even when the groups were matched on 2AFC recollection accuracy, conceptually replicating Dodson and Krueger (2006) and Dodson et al. (2007a, 2007b) and extending this pattern to a forced-choice recollection test. To illustrate the potential importance of matching recollection accuracy across age groups, consider a situation where aging impairs recollection accuracy and thereby increases the number of test responses that are based on guessing. Because correct responses on a recollection test are based on some combination of accurate recollection and guessing, this situation could result in a relatively greater proportion of correct responses that were based on guessing as opposed to accurate recollection in older adults compared to younger adults. Moreover, if metamemory were intact in older adults, then these guesses would elicit relatively lower confidence than responses based on accurate recollection, potentially lowering the average confidence made to correct trials relative to incorrect trials. As a result, even if metamemory ability were intact in older adults, increased guessing could reduce estimates of metamemory accuracy, because metamemory measures depend on the confidence difference between correct and incorrect trials. By matching the age groups on 2AFC accuracy, and hence guessing rates, our study avoided these interpretative issues.

Our matched-accuracy findings are theoretically important because they indicate that the age-related reduction in metamemory accuracy cannot be entirely due to differences in recollection quantity or guessing rates across the age groups. Instead, these matched-accuracy results indicate that the age-related reduction in metamemory accuracy was due to some other mechanism. As discussed in the Introduction, there are at least two other factors that might contribute to the age-related reduction in metamemory accuracy. The first factor is recollection quality (i.e., the number of unique features associated with retrieved items), which we differentiate from recollection quantity (i.e., the frequency of successfully retrieving items, cf. Scimeca et al., 2011). Even if recollection quantity were matched across the age groups, yielding similar recollection test accuracy, age-related reductions in the quality of successfully recollected information could restrict the range of retrieved features that could help to differentiate correct from incorrect responses when making confidence judgments. In a sense, this recollection-based explanation simply assumes that confidence judgments are a more sensitive measure of fine-grained differences in recollected detail compared to the 2AFC judgment. The second factor that could contribute to the age-related reduction in metamemory accuracy is impairment in a metacognitive monitoring process. It may be that a metacognitive monitoring process is independent from the recollection processes, in the sense that it is supported by a different network of brain regions (e.g., prefrontal cortex more than hippocampus, cf. Schacter, Norman, & Koutstaal, 1998), and/or could be affected by processes other than recollection (such as working memory or anxiety, as discussed in the Introduction). In these cases, a metacognitive impairment could reduce the accuracy of confidence judgments even if accurate features were otherwise available for retrieval.

Although the age-related metamemory impairment that we observed could have been due to differences in recollection quality or metacognitive monitoring (or both), other findings could be construed as favoring a recollection quality interpretation. The finding that older

adults are relatively good at metamemory judgments for non-recollected information (e.g., general knowledge, Dahl, Allwood, & Hagberg, 2009; Lachman, Lachman, & Thronesbery, 1979; Pliske & Mutter, 1996) suggests that they do not have a generalized metacognitive deficit. By extension, these findings suggest that the age-related impairment in metamemory that we observed was due to differences in recollection quality. Our finding that metamemory accuracy was greater for picture recollections than for font color recollections in both age groups further implicates the importance of recollection quality in metamemory accuracy. In fact, the differences in metamemory accuracy that we observed with our manipulation of recollection quality cannot be attributed to individual differences in a metacognitive monitoring process, because our manipulation of recollection quality was conducted within-subjects.

Our finding that metamemory accuracy was reduced for font recollections compared to picture recollections is analogous to distinctiveness effects that have been observed in the false recognition literature (e.g. Dodson & Schacter, 2001; Gallo et al., 2007; Schacter et al., 1999; Schacter & Wiseman, 2006). These findings are consistent with the idea that the same processes that affect accuracy in false recognition tasks also can enhance the accuracy of confidence judgments in a 2AFC recollection task. Because distinctiveness effects in false recognition have been attributed to recollection quality as opposed to recollection quantity (see Scimeca et al., 2011), they also bolster the conclusion that the stimulus effects observed in the present study were due to differences in recollection quality. More generally, these findings illustrate the close relationship between recollection quality and retrieval monitoring processes across a variety of testing situations, as predicted by the source monitoring framework (Johnson et al., 1993).

Although some aspects of our results were consistent with previous false recognition effects, our results did not support the prediction that age-related metamemory impairments would be greater for font color relative to pictures. This prediction was based on the idea that less distinctive stimuli are more prone to the effects of familiarity and false recollection, especially in older adults (e.g., Gallo et al., 2007; Schacter et al., 1999). In fact, there was no evidence that incorrect responses were made with greater confidence for font color than pictures in either age group, and older adults did not make more high-confidence errors for font color than younger adults. We instead found that the differences between stimuli and age groups were primarily localized in confidence judgments for correct responses, as were the correlations between recollection accuracy and confidence within each condition. These patterns suggest that differences in recollection quality primarily influenced confidence judgments by affecting the likelihood that detailed features from target items would be retrieved when making a correct response, as opposed to false recollections associated with incorrect responses. Older adults may have been equally impaired on accuracy for the font test and the picture test because both tests required the recollection of specific details associated with studied items (i.e., remembering the color of prior presentation), and this recollection in turn depended on a general binding process that was impaired with age (e.g., Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000).

It has been shown that older adults are more prone to false recollection in other testing contexts, such as more typical source memory tests (e.g., Dodson et al., 2007a). As discussed above, we did not find consistent evidence for this effect on the font test. One possible explanation for this discrepancy is that the 2AFC test affected our ability to observe high-confidence errors. Because two items were presented on each test trial, an otherwise compelling false recollection for the lure may have been offset by some degree of accurate recollection for the target (i.e., a relative comparison may have lowered confidence). However, it is important to note that we found some evidence for age-related increases in confidence to incorrect responses on the picture test, and age effects on calibration error

scores were greatest on the picture test. This finding suggests that age-related increases in false recollection might depend on the kind of to-be-recalled information (e.g., Gopie, Craik, & Hasher, 2010; May et al., 2005), with more false recollection for more distinctive stimuli (pictures instead of font). Although this effect was not predicted by the false recognition literature as a whole, a similar pattern in confidence judgments has been reported in at least one prior false recognition task (Gallo et al., 2009), suggesting that it is not unique to the 2AFC task. Taken together, these findings suggest that errors are overall less likely for pictures than font color, as documented in the false recognition literature, but when errors do occur, older adults are more likely than younger adults to make high-confidence errors for pictures relative to font color, potentially because pictures have more features to erroneously recombine into a false recollection. Of course, this account is only speculative, and additional research will be necessary to determine the extent that different materials and tasks might differentially affect false recollection in younger and older adults.

In sum, the results of the current study suggest that aging reduces the confidence-accuracy relationship by impairing recollection quality, above and beyond the potential influence of recollection quantity. This recollection quality impairment restricts the range of the to-be-recalled details upon which monitoring processes can differentiate correct from incorrect responses. These results are consistent with a recent overview of the metacognitive literature by Hertzog and Dunlosky (2011), where it was concluded that aging generally spares monitoring processes in other contexts, but that aging impairs some metamemory judgments due to differences in recollection quality. Adding to this conclusion, our results suggest that these age-related differences are not entirely due to false recollection effects, but instead can be largely due to differences in the recollection and/or monitoring of specific details for studied items.

## Acknowledgments

The authors are grateful to Laura Ransin for assisting with data collection, Sasha Cervantes for compiling the stimuli, and Chad Dodson for the formulas to calculate calibration error scores. This work was supported by the National Institute on Aging Grant AG030345.

## References

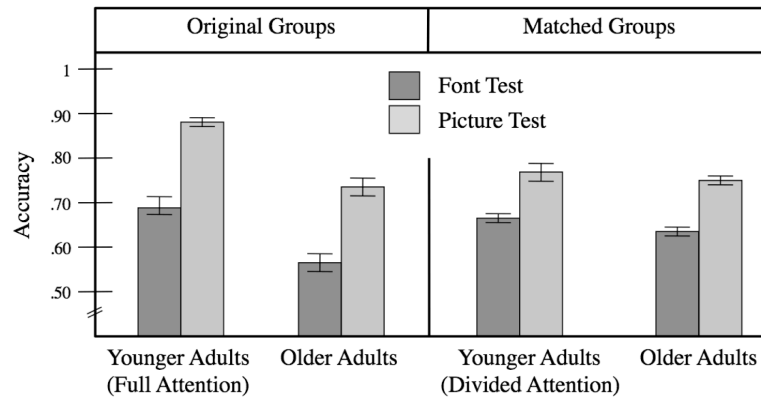
- Balota, DA.; Dolan, PO.; Duchek, JM. Memory changes in healthy older adults. In: Tulving, E.; Craik, FIM., editors. *The Oxford handbook of memory*. Oxford University Press; New York: 2000. p. 395-409.
- Brewer GA, Marsh RL, Clark-Foos A, Meeks JT. Noncriterial recollection influences metacognitive monitoring and control processes. *Quarterly Journal of Experimental Psychology*. 2010; 63:1936–1942.
- Brink TL, Yesavage JA, Lum O, Heersema P, Adey MB, Rose TL. Screening tests for geriatric depression. *Clinical Gerontologist*. 1982; 1:37–44.
- Chalfonte BL, Johnson MK. Feature memory and binding in young and older adults. *Memory & Cognition*. 1996; 24:403–416.
- Chua EF, Schacter DL, Sperling RA. Neural basis for recognition confidence in younger and older adults. *Psychology & Aging*. 2009; 24:139–153. [PubMed: 19290745]
- Connor LT, Dunlosky J, Hertzog C. Age-related differences in absolute but not relative metamemory accuracy. *Psychology & Aging*. 1997; 12:50–71. [PubMed: 9100268]
- Dahl M, Allwood CM, Hagberg B. The realism in older people's confidence judgments of answers to general knowledge questions. *Psychology & Aging*. 2009; 24:234–238. [PubMed: 19290758]
- Daniels, K.; Toth, J.; Jacoby, L. The aging of executive functions. In: Bialystok, E.; Craik, FIM., editors. *Lifespan cognition: Mechanisms of change*. Oxford University Press; New York, NY: 2006. p. 96-111.

- Dodson CS, Bawa S, Krueger LE. Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology & Aging*. 2007a; 22:122–133. [PubMed: 17385989]
- Dodson CS, Bawa S, Slotnick SD. Aging, source memory, and misrecollections. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2007b; 33:169–181.
- Dodson CS, Krueger LE. I misremember it well: Why older adults are unreliable witnesses. *Psychonomic Bulletin & Review*. 2006; 13:770–775. [PubMed: 17328371]
- Dodson CS, Schacter DL. “If I had said it I would have remembered it”: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*. 2001; 8:155–161. [PubMed: 11340861]
- Dunlosky J, Baker JMC, Rawson KA, Hertzog C. Does aging influence people’s metacomprehension? Effects of processing ease on judgments of text learning. *Psychology & Aging*. 2006; 21:390–400. [PubMed: 16768583]
- Dunlosky J, Connor LT. Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition*. 1997; 25:691–700.
- Flavell JH. First discussant’s comments: What is memory development the development of? *Human Development*. 1971; 14:272–278.
- Folstein MF, Folstein SE, McHugh PR. “Mini-Mental State”: A practical method for grading the mental state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198. [PubMed: 1202204]
- Gallo DA, Cotel SC, Moore CD, Schacter DL. Aging can spare recollection-based retrieval monitoring: The importance of event distinctiveness. *Psychology & Aging*. 2007; 22:209–213. [PubMed: 17385996]
- Gallo DA, Foster KT, Johnson EJ. Elevated false recollection of emotional pictures in young and older adults. *Psychology & Aging*. 2009; 24:981–988. [PubMed: 20025411]
- Gallo DA, Roediger HL III. The effects of associations and aging on illusory recollection. *Memory & Cognition*. 2003; 31:1036–1044.
- Gonzalez R, Nelson TO. Measuring ordinal association in situations that contain tied scores. *Psychological Bulletin*. 1996; 119:159–165. [PubMed: 8559859]
- Goodman LA, Kruskal WH. Measures of association for cross classifications. *Journal of the American Statistical Association*. 1954; 49:732–764.
- Gopie N, Craik FIM, Hasher L. Destination memory impairment in older people. *Psychology & Aging*. 2010; 25:922–928. [PubMed: 20718537]
- Hertzog, C. Metacognition in older adults: Implications for application. In: Perfect, TJ.; Schwartz, BL., editors. *Applied metacognition*. Cambridge University Press; London: 2002. p. 169-196.
- Hertzog C, Dunlosky J. Metacognition in later adulthood: Spared monitoring can benefit older adults’ self regulation. *Current Directions in Psychological Science*. 2011; 20:167–173.
- Hertzog C, Kidder D, Powell-Moman A, Dunlosky J. Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology & Aging*. 2002; 17:209–225. [PubMed: 12061407]
- Hultsch DF, Hertzog C, Dixon RA. Age differences in metamemory: Resolving the inconsistencies. *Canadian Journal of Psychology*. 1987; 41:193–208. [PubMed: 3502896]
- Jacoby LL, Bishara AJ, Hessels S, Toth J. Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General*. 2005; 134:131–148. [PubMed: 15869342]
- Jacoby LL, Rhodes MG. False remembering in the aged. *Current Directions in Psychological Science*. 2006; 15:49–53.
- Jacoby LJ, Wahlheim CN, Rhodes MG, Daniels KA, Rogers CS. Learning to diminish the effects of proactive interference: Reducing false memory for young and older adults. *Memory & Cognition*. 2010; 38:820–829.
- Jennings JM, Jacoby LL. Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology & Aging*. 1993; 8:283–293. [PubMed: 8323731]
- Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. *Psychological Bulletin*. 1993; 114:3–28. [PubMed: 8346328]

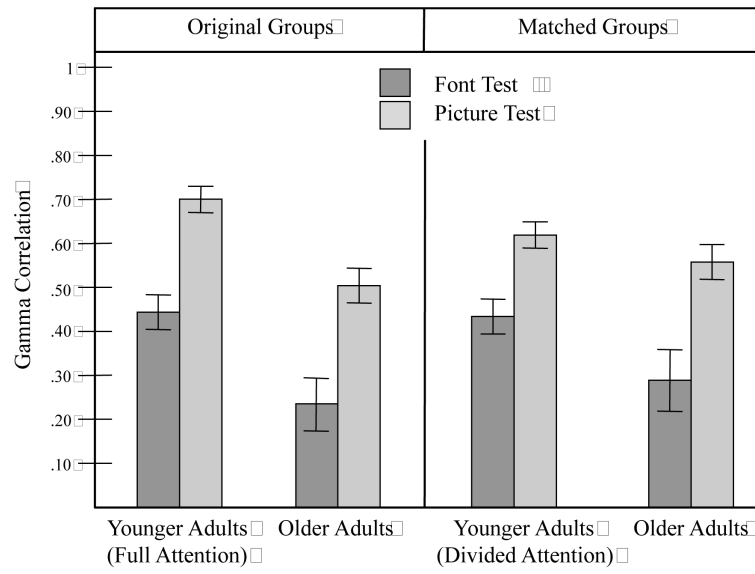
- Jonker C, Geerlings MI, Schmand B. Are memory complaints predictive for dementia? A review of clinical and population-based studies. *International Journal of Geriatric Psychiatry*. 2000; 15:983–991. [PubMed: 11113976]
- Karpel ME, Hoyer WJ, Togli MP. Accuracy and qualities of real and suggested memories: Nonspecific age differences. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. 2001; 56:P103–110.
- Kelley CM, Sahakyan L. Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language*. 2003; 48:704–721.
- Koriat A, Goldsmith M. Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*. 1996; 103:490–517. [PubMed: 8759045]
- Kuhlman BG, Touron DR. Older adults' use of metacognitive knowledge in source monitoring: Spared monitoring but impaired control. *Psychology & Aging*. 2011; 26:143–149. [PubMed: 21058865]
- Lachman JL, Lachman R, Thronesbery C. Metamemory throughout the adult lifespan. *Developmental Psychology*. 1979; 15:543–551.
- Lampinen JM, Meier CR, Arnal JD, Leding JK. Compelling untruths: content borrowing and vivid false memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2005; 31:954–963.
- Lyle KB, Johnson MK. Importing perceived features into false memories. *Memory*. 2006; 14:197–213. [PubMed: 16484110]
- May CP, Rahhal T, Berry EM, Leighton EA. Aging, source memory, and emotion. *Psychology & Aging*. 2005; 20:571–578. [PubMed: 16420132]
- Naveh-Benjamin M. Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 2000; 26:1170–1187.
- Nelson TO. A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*. 1984; 95:109–133. [PubMed: 6544431]
- Nelson TO. Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item: Comments on Shraw (1995). *Applied Cognitive Psychology*. 1996; 10:257–260.
- Nelson, TO.; Narens, L. Metamemory: A theoretical framework and new findings. In: Bower, GH., editor. *The psychology of learning and motivation*. Academic Press; New York: 1990. p. 125-173.
- Nilsson LG. Memory function in normal aging. *Acta Neurologica Scandinavica*. 2003; 107:7–13. [PubMed: 12542507]
- Norman KA, Schacter DL. False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*. 1997; 25:838–848.
- Pansky A, Goldsmith M, Koriat A, Pearlman-Avni S. Memory accuracy in old age: Cognitive, metacognitive, and neurocognitive determinants. *European Journal of Cognitive Psychology*. 2009; 21:303–329.
- Parkin AJ, Walter BM. Recollective experience, normal aging, and frontal dysfunction. *Psychology & Aging*. 1992; 7:290–298. [PubMed: 1610518]
- Pliske RM, Mutter SA. Age differences in the accuracy of confidence judgments. *Experimental Aging Research*. 1996; 22:199–216. [PubMed: 8735153]
- Pull MW, Dawes LLC, Martin AM III, Rosenberg HF, Light LL. Recollection and familiarity in recognition memory: Adult age differences and neuropsychological test correlates. *Psychology & Aging*. 2006; 21:107–118. [PubMed: 16594796]
- Rotello, CM.; Macmillan, NA. Response bias in recognition memory: Skill and strategy in memory use. In: Benjamin, AS.; Ross, BH., editors. *The psychology of learning and motivation: Advances in research and theory*. Elsevier Academic Press; San Diego: 2008. p. 61-94.
- Ryan EB. Beliefs about memory changes across the life span. *Journal of Gerontology*. 1992; 47:41–46.
- Schacter DL, Israel L, Racine C. Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*. 1999; 40:1–24.
- Schacter DL, Norman KA, Koutstaal W. The cognitive neuroscience of constructive memory. *Annual Review of Psychology*. 1998; 49:289–318.



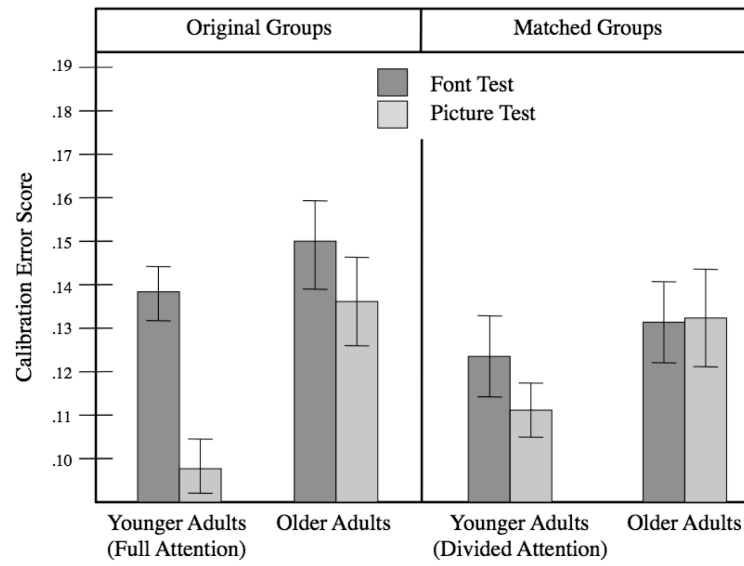
- Schacter, DL.; Wiseman, AL. Reducing memory errors: The distinctiveness heuristic. In: Hunt, RR.; Worthen, JB., editors. *Distinctiveness and memory*. Oxford University Press; New York: 2006. p. 89-107.
- Scimeca JM, McDonough IM, Gallo DA. Quality trumps quantity at reducing memory errors: Implications for retrieval monitoring and mirror effects. *Journal of Memory and Language*. 2011; 65:363–377.
- Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning*. 1980; 6:174–215.
- Souchay C, Isingrini M. Age related differences in metacognitive control: Role of executive functioning. *Brain and Cognition*. 2004; 56:89–99. [PubMed: 15380879]
- Souchay C, Isingrini M, Espagnet L. Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology*. 2000; 14:299–309. [PubMed: 10791869]
- Szekely A, D'Amico S, Devescovi A, Federmeier K, Herron D, Iver G, Jacobsen T, Arevalo AL, Vargha A, Bates E. Timed action and object naming. *Cortex*. 2005; 41:7–25. [PubMed: 15633703]
- Thomas AK, Bulevich JB, Dubois SJ. Context affects feeling-of-knowing accuracy in younger and older adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2011; 37:96–108.



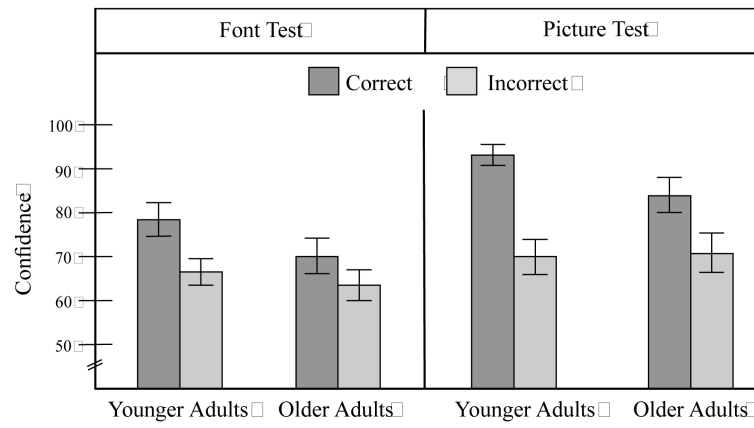
**Figure 1.** Proportion correct (and standard errors) on the 2AFC recollection tests for the two full attention groups (left panel) and two groups matched on accuracy (right panel).



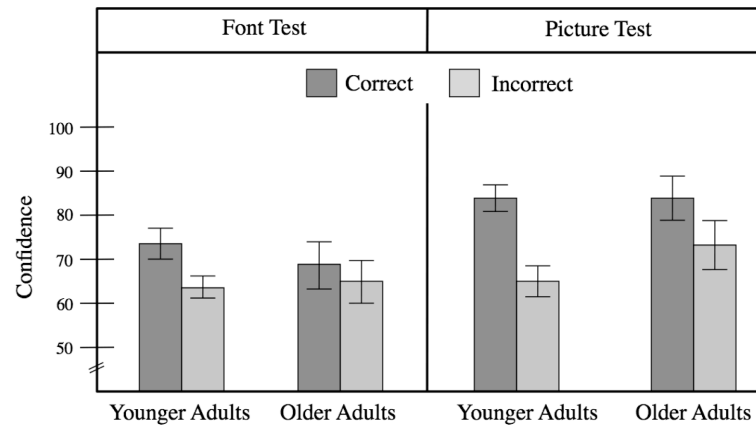
**Figure 2.** Mean gamma scores (and standard errors) for the full attention groups (left panel) and matched groups (right panel) on both test types.



**Figure 3.** Mean calibration error scores (and standard errors) for the full attention groups (left panel) and matched groups (right panel) on both test types.



**Figure 4.** Mean confidence ratings (and standard errors) for the full attention groups on the font test (left panel) and picture test (right panel).



**Figure 5.** Mean confidence ratings (and standard errors) for the matched groups on the font test (left panel) and picture test (right panel).

**Table 1**

Mean proportion of total test trials in each of three different confidence bins for each participant group.

| Group           | Confidence |           |           |
|-----------------|------------|-----------|-----------|
|                 | 50/60      | 70/80     | 90/100    |
| Font Test       |            |           |           |
| Younger Full    | .38 (.03)  | .26 (.02) | .36 (.03) |
| Younger Divided | .45 (.03)  | .23 (.02) | .32 (.03) |
| Older Full      | .54 (.04)  | .31 (.04) | .15 (.03) |
| Younger Matched | .49 (.03)  | .26 (.02) | .25 (.03) |
| Older Matched   | .51 (.04)  | .32 (.04) | .17 (.03) |
| Picture Test    |            |           |           |
| Younger Full    | .12 (.01)  | .16 (.02) | .72 (.03) |
| Younger Divided | .28 (.03)  | .22 (.02) | .50 (.03) |
| Older Full      | .27 (.04)  | .27 (.03) | .46 (.04) |
| Younger Matched | .30 (.02)  | .23 (.02) | .48 (.03) |
| Older Matched   | .27 (.04)  | .25 (.02) | .48 (.04) |

Notes. Standard error of each mean in parenthesis.

**Table 2**

Mean proportion of test trials in each of three different confidence bins for each participant group, calculated separately for correct and incorrect trials.

| Font Test    |           | Confidence      |           |           |           |
|--------------|-----------|-----------------|-----------|-----------|-----------|
|              |           | 50/60           | 70/80     | 90/100    |           |
| Font Test    | Correct   | Younger Full    | .32 (.03) | .26 (.02) | .42 (.04) |
|              |           | Younger Divided | .38 (.03) | .24 (.02) | .38 (.04) |
|              |           | Older Full      | .52 (.05) | .31 (.04) | .17 (.03) |
|              |           | Younger Matched | .41 (.04) | .27 (.03) | .32 (.03) |
|              |           | Older Matched   | .48 (.05) | .32 (.04) | .19 (.04) |
|              | Incorrect | Younger Full    | .55 (.03) | .27 (.03) | .18 (.03) |
|              |           | Younger Divided | .64 (.03) | .24 (.03) | .13 (.02) |
|              |           | Older Full      | .59 (.05) | .31 (.04) | .10 (.03) |
|              |           | Younger Matched | .65 (.04) | .24 (.03) | .11 (.02) |
|              |           | Older Matched   | .57 (.05) | .32 (.04) | .11 (.03) |
| Picture Test | Correct   | Younger Full    | .08 (.01) | .15 (.02) | .76 (.03) |
|              |           | Younger Divided | .21 (.02) | .21 (.02) | .57 (.03) |
|              |           | Older Full      | .23 (.04) | .26 (.02) | .52 (.04) |
|              |           | Younger Matched | .22 (.02) | .23 (.03) | .55 (.04) |
|              |           | Older Matched   | .22 (.03) | .24 (.02) | .54 (.04) |
|              | Incorrect | Younger Full    | .46 (.04) | .28 (.03) | .26 (.04) |
|              |           | Younger Divided | .55 (.03) | .26 (.03) | .19 (.02) |
|              |           | Older Full      | .41 (.05) | .34 (.03) | .26 (.04) |
|              |           | Younger Matched | .59 (.04) | .23 (.03) | .18 (.03) |
|              |           | Older Matched   | .43 (.05) | .30 (.03) | .27 (.04) |

Notes. Standard error of each mean in parenthesis.



**Table 3**

Bivariate correlations between 2AFC recollection accuracy and each metamemory accuracy measure, across the individuals in each group.

|                   | Younger Full | Younger Divided | Older Full |
|-------------------|--------------|-----------------|------------|
| Font Test         |              |                 |            |
| Resolution        | + .41 **     | + .56 **        | + .43 **   |
| Discrimination    | + .51 **     | + .65 **        | + .62 **   |
| Calibration Error | - .46 **     | - .54 **        | - .38 *    |
| Picture Test      |              |                 |            |
| Resolution        | + .37 *      | + .55 **        | .54 **     |
| Discrimination    | + .35 *      | + .62 **        | + .60 **   |
| Calibration Error | - .70 **     | - .54 **        | - .43 **   |

Notes. Resolution and discrimination were positively coded (higher scores indicated greater metamemory accuracy), whereas calibration was negatively coded (higher scores indicate greater metamemory error).

\*  $p < .05$  (uncorrected),

\*\*  $p < .05$  (Bonferroni corrected, or uncorrected  $p < .0027$ ).

**Table 4**

Bivariate correlations between 2AFC recollection accuracy and the average confidence to correct and incorrect responses, across the individuals in each group.

|              | Younger Full        | Younger Divided     | Older Full          |
|--------------|---------------------|---------------------|---------------------|
| Font Test    |                     |                     |                     |
| Correct      | + .63 <sup>**</sup> | + .70 <sup>**</sup> | + .45 <sup>**</sup> |
| Incorrect    | + .32 <sup>*</sup>  | + .31 <sup>*</sup>  | + .17               |
| Picture Test |                     |                     |                     |
| Correct      | + .39 <sup>*</sup>  | + .64 <sup>**</sup> | + .41 <sup>**</sup> |
| Incorrect    | -.15                | + .09               | + .06               |

\* Notes.  $p < .05$  (uncorrected),

\*\*  $p < .05$  (Bonferroni corrected, or uncorrected  $p < .0042$ ).