



Published in final edited form as:

Psychol Aging. 2021 March ; 36(2): 186–199. doi:10.1037/pag0000595.

Are Age Differences in Recognition-based Retrieval Monitoring an Epiphenomenon of Age Differences in Memory?

Christopher Hertzog,

School of Psychology, Georgia Institute of Technology

Taylor Curley,

School of Psychology, Georgia Institute of Technology

John Dunlosky

Department of Psychology, Kent State University

Abstract

Older adults often demonstrate a monitoring deficit by producing more high-confidence memory errors on recognition memory tests. To eliminate lower memory performance by older adults (OA) as a candidate explanation, we studied how distinctive encoding enhances the retrieval-monitoring accuracy in older adults and younger adults (YA) under different delays (2-day delay for OA, 7-day delay for YA). Individuals viewed items consisting of four randomly selected exemplars (e.g., SALMON, BASS, PERCH, SHARK) from a taxonomic category (e.g., FISH), one being designated as the to-be-remembered target. Participants were randomly assigned to two encoding conditions: Shared (generate a shared feature of all exemplars, e.g., GILLS) or Distinctive (generate a distinctive feature of the designated target). We collected retrospective confidence judgments (RCJs) after a five-alternative forced-choice (5AFC) recognition test, with the lures being either previously presented (old) exemplars or new category exemplars. Recall and recognition memory were better with distinctive encoding, with shared feature generation producing more high-confidence false alarms (HCFA). Distinctive encoding dramatically reduced HCFAs and improved RCJ resolution. Comparison of OA with 2-day delay YA revealed age differences in HCFA consistent with previous studies. As important, age differences in memory for OA and 7-day delay YA were minimized, eliminating age deficits in HCFAs. Matching OAs to a subset of 7-day delay YAs on recognition memory produced additional evidence favoring the null hypothesis of age-equivalence in HCFAs. The results therefore indicated that age differences in recognition-based retrieval monitoring in a forced-choice recognition test are an epiphenomenon of age differences in memory.

Keywords

recognition memory; confidence judgments; high-confidence false alarms; retrieval monitoring

Introduction

Older adults experience age-related decline in episodic memory (Balota, Dolan, & Duchek, 2000; Hertzog & Shing, 2011; Zacks, Hasher, & Li, 2000), even in the absence of memory-related pathologies such as Alzheimer's disease. Older adults have also been reported to have deficits in metacognitive monitoring of retrieval, including a tendency towards high-confidence false alarms (Castel, Middlebrooks, & McGillivray, 2016). In the present investigation, we used sets of categorizable nouns to investigate how distinctive processing at encoding influences age differences in memory errors and metacognitive monitoring illusions generated by studying and remembering confusable same-category exemplars.

Age Differences in Memory Errors

Older adults are more prone to a variety of memory illusions, often leading to elevated false alarms in recognition memory tasks (Devitt & Schacter, 2016; Gallo, 2006). In particular, age differences in recognition memory – when they occur – are usually more attributable to older adults' elevated false alarm rates rather than reduced hit rates (e.g., Koutstaal, 2003; Trelle et al., 2017; see reviews by Devitt & Schacter, 2016, and Light, Prull, LaVoie, & Healy, 2000). This phenomenon is more likely when older adults misattribute the experience of item familiarity (a recognition process that appears to be relatively unaffected by aging, Koen & Yonelinas, 2014) as signifying the test item was previously studied (Jacoby & Rhodes, 2006). Naveh-Benjamin (2000; Olds & Naveh-Benjamin, 2008) identified an age-related associative recognition deficit after people study paired associates (e.g., concrete, normatively unrelated nouns). Older adults show relatively spared old-new recognition memory for each word but perform more poorly when asked to discriminate previously studied (intact) pairs from rearranged pairs (word pairs constructed by taking words from two different paired-associate items).

Aging and Metacognitive Retrieval Monitoring

Evidence is inconsistent as to whether older adults are also impaired in metacognitive monitoring of episodic memory retrieval (Castel et al., 2016; Hertzog, 2016). Monitoring of retrieval during tests of multiple-item lists can be measured with judgments about each item including confidence that a memory decision (e.g., a recognition memory response identifying a candidate target) is correct. Historically, retrospective confidence judgments (RCJs) have most often been used in Signal Detection Theory (SDT) models of recognition memory (e.g., Mickes, Johnson, & Wixted, 2010; Koen & Yonelinas, 2014), but they also figure prominently in metacognitive research on retrieval monitoring (Dunlosky & Metcalfe, 2009). From the latter perspective, RCJs are metacognitive judgments influenced by accessibility of cues generated during the recognition test, including the fluency of retrieval (Hines, Hertzog, & Touron, 2009).

The accuracy of RCJs can be measured by alternative indices of within-person associations of RCJs with recognition memory success, including Goodman-Kruskal gamma correlations (Nelson, 1984) or measures based on Type 2 receiver operating characteristics (Higham & Higham, 2019)¹. Older adults have been reported to have deficits in accuracy of RCJs (e.g., Kelley & Sahakyan, 2003). They are prone to being falsely confident when making

a recognition memory false alarm (e.g., Chua, Schacter, & Sperling, 2009; Dodson, Bawa, & Kruger, 2007; Fandakova et al., 2013, 2018; Shing et al, 2009), especially in tasks designed to produce memory errors (e.g., Tun et al., 1998; but see Panuswan et al., 2020). In source monitoring experiments, they are less likely to reconsider inaccurate source memory attributions by searching for disconfirming evidence (Henkel et al., 1998; Johnson, 2005). They are also prone to familiarity-based memory errors (Jennings & Jacoby, 2003; Jacoby, 1999) and are less likely to use post-retrieval monitoring strategies like recall-to-reject (Light et al, 2000) or the distinctiveness heuristic (Gallo et al, 2006; Schacter, Israel, & Racine, 1999) to avoid such memory errors (Fandakova et al., 2018).

The tendency for older adults to produce high-confidence false alarms (HCFAs) should translate to lower resolution of RCJs with recognition memory, and it does (e.g., Kelley & Sahakyan, 2003). However, a major concern regarding any such effect is whether it merely reflects the consequence of well-known recollection deficits in older adults' recognition experiences (Perfect & Stollery, 1993). Older adults' recognition memory successes are more likely than younger adults' successes to be based on phenomenal experiences of familiarity rather than specific recollection of episodic detail (e.g., Hay & Jacoby, 1999; Perfect & DasGupta, 1997; see Light et al., 2000). Age-deficits in phenomenal recollection during memory performance could reduce the availability of valid cues during retrieval (including noncriterial recollection, Brewer et al., 2010; Hertzog et al., 2014) that discriminate veridical memories from familiarity-induced recognition errors.

Several studies point to the value of equating age groups on memory performance when evaluating RCJ resolution. Hines, Touron, and Hertzog (2009) equated older and younger adults on yes-no associative recognition performance by giving older adults additional study time for each item, yet still detected age differences in RCJ resolution, ruling out the possibility that the monitoring deficit was an artifact of age differences in recollection. Wong, Cramer, and Gallo (2012) found age differences in RCJ resolution for names of common objects. They manipulated resolution in old adults and young adults by contrasting repeated presentation of the object name with a condition that presented pictures of the common objects, which had equivalent effects in both age groups. Furthermore, age differences in RCJ resolution did not disappear when subsets of older and younger adults were matched on memory performance. In contrast, Hertzog, Dunlosky, and Sinclair (2010) equated age groups in associative memory by using a longer retention interval for younger adults and found no age differences in RCJ resolution for a forced-choice recognition memory test.

This investigation brought the issue of age-related retrieval monitoring deficits into clearer focus. We used a category cued-recall task that challenges memory for contextual details by producing semantically based reconstructive memory errors (Brainerd & Reyna, 2015; Healey & Kahana, 2016; Hertzog, Fulton, Mandviwala, & Dunlosky, 2013). It produces high-confidence (recall and recognition) false alarms after individuals engage in relational processing of four exemplars from the same noun category (e.g., animals) presented

¹Higham & Higham (2019) showed that SDT-based measures and gamma correlations are closely related and in some cases asymptotically equivalent measures of resolution.

at study, with only one of them designated as the to-be-remembered target. We used a potent encoding manipulation of either relational (shared feature identification) or distinctiveness encoding; distinctiveness reduces false alarms in this task (Smith & Hunt, 1996). We also used different retention intervals for the two of our age groups that eliminated age differences in memory performance. This approach enabled us to evaluate whether distinctive processing differentially repaired high confidence false alarms and RCJ resolution in older adults. Furthermore, when differences in memory performance were eliminated (by comparing age groups with different retention intervals), we could evaluate whether typical age differences in high-confidence memory errors were eliminated, implying the age deficit in metacognitive monitoring were driven by memory deficits.

Distinctive Processing in the Category Memory Task

Distinctiveness-based encoding is a powerful manipulation known to affect memory (Hunt, 2012). It requires the explicit identification of distinctive, stimulus-specific features of studied items. Older adults are known to benefit from distinctiveness-based encoding (e.g., Carr, Castel, & Knowlton, 2015; Geraci et al., 2009; Luszcz, Roberts, & Mattiske, 1990; see Smith, 2006, for a review), perhaps because it enhances access to specific details that is otherwise deficient following older adults' typical processing at encoding (Aizpurua & Koutstaal, 2010; Greene & Naveh-Benjamin, 2020). Recently, Huff and Aschenbrenner (2018) showed that distinctive encoding reduces age differences in false memories in the Deese-Roediger-McDermott task.

This study evaluated distinctiveness effects on retrieval monitoring by adapting a category memory task (Smith & Hunt, 1996). Distinctiveness can be defined in this context as a process of identifying distinguishing features of specific target items relative to semantically similar distractors (Hunt, 2012). The task is shown in Figure 1. During Session 1, each item involves presentation of four concrete nouns from a single taxonomic category (e.g., birds, fruit, or fish). Four randomly selected nouns were displayed in a vertical grid; one of the items was designated as the to-be-remembered target when receiving a category label to cue recall (i.e., SALMON, in Figure 1). Critically, participants were randomly assigned to a between-subjects condition of orienting task (Shared vs. Distinctive). In the Shared-feature condition, they generated a feature shared by all four exemplars (e.g., "fish have gills"). In the distinctive condition, they generated a feature that differentiated the target from the other three exemplars (e.g., "a salmon has pink flesh"). Smith and Hunt (1996) showed that recall was substantially better following distinctive processing than shared-feature processing. For the purposes of the current investigation, we focus on recognition memory and RCJs in Session 2. A key feature of the task is the use of a five-alternative forced choice (5AFC) recognition memory test (for an example, see the bottom right panel of Figure 1), in which the original target, two exemplars presented as context at study, and two exemplars not previously studied are provided as candidate answers.

Why does this task produce memory errors? We offer the following conceptual model. Under the shared-feature orienting task, encoding the categorizable stimuli creates explicit memory representations of category exemplars that include traces from spreading activation to other category members. Shared processing reinforces category-level information and

further strengthens memory for all displayed exemplars including those not designated as the target. Distinctive processing overlays these graded representations with item-specific features that differentiate the target from its same-category exemplars. Cueing recall by presenting the category name initiates a search process that activates the taxon through two overlapping processes in cascade: (a) reconstruction governed by semantic activation and (b) sampling of candidate answers from the taxonomic category (as in Nelson, McKinney, Gee, & Janczura's [1998] PIER2 model) generates candidates that may be correctly recognized (Hunt, Smith, & Toth, 2016), often on the basis of familiarity experiences that arise both from prior exposure and semantic activation, or alternatively, episodic recollection. If the target is identified via a strong episodic recollection (e.g., Higham & Tam, 2005), then other candidates from the category, even ones generating a familiarity signal, will be discounted or ignored (Brainerd, Reyna, & Howe, 2009). Otherwise, generated candidates will be evaluated on the basis of metacognitively accessed cues such as episodic familiarity.

Likewise, presentation of candidate category exemplars during the recognition test trigger a reconstruction process that will be overridden by a strong recollection of the target. Otherwise, recognition memory errors will occur if either (a) episodic familiarity of co-presented lures is mistakenly attributed to deriving from the target, or (b) semantically-based activation of category exemplars creates an illusion of memory. Semantic reconstruction can create an experience of fluency and familiarity, even for unstudied lures, and that sense of fluency can be mistakenly attributed to being produced by the original study episode. Distinctive processing during study increases the likelihood of episodic recollection of the target, but it also generates traces of item-specific features that can be accessed at test when recollection fails, increasing the likelihood that reconstruction will result in selecting the correct target in a recognition memory test.

Access to memory representations based on an item's original encoding decreases as the retention interval between study and test increases for adults of all ages (Kausler, 1994; MacDonald et al., 2006), increasing the likelihood that candidates will be offered for recall responses based on semantic activation and confusion of semantic familiarity with episodic sources (Healey & Kahana, 2016). Such effects are overlaid on age differences in associative strength generated by encoding processes and strategies (e.g., Dunlosky & Hertzog, 2001; Hertzog et al., 2013; Naveh-Benjamin, Brav, & Levy, 2007). We used different retention intervals (7 days for younger adults, 2 days for older adults) that we hypothesized would generate similar recall and recognition memory performance (see Hertzog, Dunlosky, & Sinclair, 2010). To demonstrate typical age differences in variables like HCFA, we later added a comparison group of young adults tested after a 2-day delay.

Hypotheses

We tested several hypotheses. First, shared-feature encoding will result in substantial rates of recall intrusion errors and HCFAs in the forced-choice recognition test for both 7-day delay young adults 2-day delay older adults. Second, HCFAs will be more likely for old lures than new category lures, establishing familiarity as a basis for the HCFAs. Third, distinctive encoding will reduce these effects, and if it does, then distinctive encoding is also expected to increase RCJ resolution (because HCFAs would reduce the resolution of

RCJs). Fourth, with different retention intervals equating memory performance for 7-day delay young adults with 2-day delay older adults, age differences will be absent in HCFAs and RCJ resolution. Fifth, consistent with prior literature, a comparison of age groups given a common 2-day retention interval will, in contrast, produce age differences in recall, recognition, recall intrusion errors, recognition HCFA, and RCJ resolution. The latter two hypotheses, if confirmed, would argue that age differences in retrieval monitoring are an epiphenomenon based on reduced access to recollection-based cues for older adults when making RCJs.

Method

Participants

This research was conducted under a research protocol approved by the Georgia Institute of Technology's Institutional Review Board.

A total of 176 adults participated in this study. Young adult participants ($N = 120$, $N_{\text{Female}} = 66$) were recruited using the undergraduate research participant pool at the Georgia Institute of Technology and received course credit for their participation. Older adult participants ($N = 56$, $N_{\text{Female}} = 34$) between the ages of 60 and 80 ($\text{Range} = 61\text{--}80$, $M = 70.52$, $SD = 4.59$) were recruited from a volunteer pool of community-dwelling adults living in the Atlanta Metro area and compensated \$50 for participation. Older participants were independent, community-dwelling adults who visited our laboratory on the Georgia Tech campus for two separate test sessions. Thirty-eight older adult participants identified their race as Caucasian while the rest identified as African-American. On average, older participants were well-educated (years of education: $M = 16.13$, $SD = 2.80$), comfortable with using a keyboard and mouse, had good or corrected vision, and did not self-report any disorder that could affect cognitive performance (e.g., Alzheimer's disease, stroke, or mild cognitive impairment). Additionally, they demonstrated normal cognitive abilities, as demonstrated by performance on the 40-question Shipley Vocabulary task (Zachary 1986), $M = 35.69$, $SD = 3.27$.

Materials

Experimental Task.—This investigation was conducted on standard PCs running the Microsoft Windows operating system, with the experimental portions programmed using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). For each participant, the program randomly drew six items from 40 categories using the Van Overschelde et al. (2003) category norms, using categories containing at least seven exemplars. We excluded some categories we considered too abstract or indistinct. We also excluded the highest typicality exemplar for each category (e.g., apple from the category fruit). The six selected items were randomly assigned to be targets, co-presented exemplars, and new lures for the five-alternative forced choice recognition memory test. One exemplar was selected to serve as the target, one exemplar was only be displayed during study, two exemplars were displayed both during study and again at recognition (serving as “old” lures), and two exemplars were used only during recognition (serving as “new” lures).

Additional materials.—Older adult participants completed a short demographic questionnaire in which they reported their age, years of education, and other relevant variables. They also completed a paper-and-pencil version of the recognition vocabulary test from the Shipley-Institute of Living Scale (Zachary, 1986).

Design and Procedure

All participants were randomly assigned to the between-subjects condition of type of orienting task at study (Encoding Condition: Shared vs. Distinctive). They participated in two sessions, separated by a retention interval (delay). Thirty-four young adults and all older adults participated in a 2-day (48-h) delay between Sessions 1 and 2; 86 young adults participated in a 7-day (1-week) delay. The 2-day delay young adult group was added later to enable age comparisons under equivalent delays.

In Session 1 (Figure 1), participants studied 40 items (categories) presented in a random order. For each item, participants first viewed 4 nouns from a taxonomic category. They were oriented to the category by being asked to respond with the word or phrase that best categorized the words. This procedure ensured initial relational processing by all participants. Immediately thereafter, participants were shown the normative category label for the item, as given by the Van Overschelde et al. (2004) norms, displayed above the four randomly selected exemplars. One noun from this set was randomly designated as the target word and highlighted in red. All other stimuli were presented in large black font against a white background (see Figure 1). The locations of the presented exemplars were randomly assigned within the column. Participants were told they would be tested for associative memory using the label we provided.

Next, participants generated a memory aid – a word or phrase to help them remember the target word associated with each category. Given that participant responses were required during the study phase, study of the items was self-paced and terminated with participants' manual entry of the memory aid. The instructions differed depending on the between-subjects condition: In the Shared Condition, participants generated a memory aid capturing a feature shared by all four presented exemplars. In the Distinctive Condition, participants generated a memory aid identifying a feature of the designated target that differentiated it from the other co-presented exemplars. To evaluate whether participants were performing the orienting task correctly, we coded each generated memory aid as describing a shared or distinctive feature of the target word. For example, an individual in the Shared Condition would be 100% compliant if every aid described a feature shared by all four exemplars. Any response that was blank, unintelligible, or highly idiosyncratic to the participant (e.g., "Mom has these") was treated as uncategorizable and ignored. Two coders rated each response with disagreements resolved by consensus or adjudicated by author Curley. Compliance was excellent overall ($M = 93\%$) but individuals in the Distinctive Condition ($M_{\text{Young}} = 96\%$, $M_{\text{Old}} = 93\%$) were slightly more compliant than those in the Shared Condition ($M_{\text{Young}} = 93\%$, $M_{\text{Old}} = 87\%$), and younger adults were more compliant.

After the designated retention interval, participants returned for Session 2. First, participants were given a cued-recall test, with the normative category label used as the cue. Item order was randomized for each participant. Participants entered an exemplar in a text box they

believed to be the target word. We scored recall accuracy by evaluation of the recall response via qualitative coding, counting unambiguous misspellings or apparent typographical errors as correct. Immediately after the recall test for each item participants gave a feeling-of-knowing judgment forecasting future recognition of the target word. We do not report on these judgments in this paper.

Next, participants were immediately given a 5AFC recognition memory task. The 40 items were tested in a new randomized order. The normative category label (e.g., Type of Fish) was presented above 5 candidate exemplars from that category (see Figure 1). One candidate was the designated target. The other four exemplars were recognition lures. Two lures had been previously co-presented with the target during study (“old” lures) and two lures were category exemplars previously unseen by the participant (“new” lures). Alternatives were randomly ordered in a vertical column similar to how items had been originally studied. Participants used the mouse to click on the alternative that they believed was the target. Immediately thereafter they provided a RCJ indicating their level of confidence they had chosen the correct target. RCJs were based on a continuous 0 to 100 scale, where “0” indicated no confidence that the correct target word was chosen during the 5AFC task and “100” indicated absolute confidence that the correct target word had been chosen.

Statistical Approach

We used IBM SPSS Statistics for Windows, version 25.0 (2017) to analyze age and encoding condition differences in aggregated cued recall, recognition, RCJs, and gamma correlations measuring resolution of RCJs in predicting recognition memory. All post hoc pairwise comparisons were computed using a Bonferroni correction on the critical alpha values. Given that we are arguing for nil effects of age on HCFAs and RCJ resolution when controlling on memory performance, we also report Bayes Factor (BF) likelihood ratios computed in JASP (JASP Team, 2018) favoring the null hypothesis that can be used to claim evidence for the equivalence of population means (e.g., Moray & Rouder, 2011). JASP does not report this measure, but it is easily computed as the reciprocal of the BF for the alternative hypothesis that is provided by the program. A benchmark of $BF = 3$ (3:1 odds favoring the null hypothesis) can be taken as moderate evidence for the null hypothesis of age-group equivalence (see Brydges & Bielak, 2020), whereas $1 < BF < 3$ can be taken as ‘anecdotal’ support for the null hypothesis, and $BF = 10$ is regarded as strong evidence for the null hypothesis. It is important to emphasize that these benchmarks are both arbitrary and sample-size dependent. We also report sample effect size statistics, either Cohen’s (1988) $d = (M_1 - M_2) / [MS_{\text{Error}}]^{-0.5}$ or partial η^2 (denoted η^2_p) provided by SPSS GLM. These estimated effect sizes are sample-size independent. Very small effect sizes (e.g., $d < .20$) can be considered evidence that any population mean differences are either completely nil or trivially small.

Results

Given the number of dependent measures analyzed in the results section, we provide a summary overview of the inferential outcomes for the most critical comparisons in Table 1. Measures appear in this table in the order that they are presented below. Effect sizes

for Condition and comparisons of young adults and old adults were scaled in Cohen's d , calculated as. Interaction effect sizes are reported as η^2_p . One clear outcome seen in Table 1 is the robust effect of distinctive vs. shared feature processing (Condition) on all outcome measures, given Cohen's (1988) suggested benchmarks: (0.5 – 0.8 is a moderate effect size, greater than 0.8 is a large effect size). The table includes BFs favoring the null hypothesis for comparisons in which a null effect was predicted, namely those involving a comparison between the 7-day delay YA and 2-day delay OA groups. Full details from all analyses are presented next.

Cued-Recall Performance

Average cued recall for all participants was submitted to a 2 (Encoding Condition: Shared vs. Distinctive) x 3 (Group: Older Adult 2-day vs. Young Adult 2-day vs. Young Adult 7-day) GLM. Figure 2 plots the cell means as a function of these conditions. A significant main effect of Encoding Condition occurred, $F(1,169) = 16.34, p < 0.01, \eta^2_p = 0.09$, indicating that learners in the Distinctive condition correctly recalled more target words, $M = 0.32, SE = 0.02$, than those in the Shared condition, $M = 0.23, SE = 0.02, d = 0.65$. The analysis also yielded a significant main effect of Group, $F(2,169) = 11.37, p < 0.01, \eta^2_p = 0.12$. Recall accuracy was significantly higher for young adults on a 2-day delay ($M = 0.35, SE = 0.02$) than for young adults on a 7-day delay ($M = 0.22, SE = 0.02, p < 0.01, d = 0.95$) and older adults on a 2-day delay ($M = 0.24, SE = 0.02, p < 0.01, d = 0.79$).² 7-day delay young adults did not differ reliably from 2-day delay older adults (see Table 1). The interaction between Condition and Group was not significant, $F < 1$.

Thus, age differences in cued recall observed with identical 2-day delays were eliminated when younger adults were evaluated at a longer 7-day delay, with distinctive encoding resulting in reliably better cued recall for both age groups.

Recall Errors

The category task is prone to recall errors. In standard associative recall, the most common error is an omission error (e.g. Hertzog et al., 2013), defined as not responding with a candidate answer (as well as providing an answer like "Next"). Commission errors were defined as category-consistent exemplars, including either co-presented exemplars or new exemplars from the same category. These recall intrusion errors are less common in standard associative recall tests, but we expected a substantial rate of commission errors given the nature of the task. Extra-category intrusions or idiosyncratic responses were coded as "Other."

A 2 (Encoding Condition: Shared vs. Distinctive) x 3 (Group: Older Adult 2-day vs. Young Adult 2-day vs. Young Adult 7-day) x 3 (Error Type: Omission vs. Commission vs. Other) mixed-effects ANOVA to examine recall error rates (Figure 3). The Greenhouse-Geisser epsilon was greater than 0.9 ($\epsilon = 0.92$), indicating that the degree of violation of the sphericity assumption was inconsequential on inflating the Type I error rate (Hertzog,

²The comparison of the two young groups merely indicates an uninteresting effect of retention interval. Henceforth we focus only on the comparisons of young adult groups to the older adults when evaluating Group, as featured in Table 1.

1994). We therefore used standard mixed-model F-tests and df for effects involving the within-subjects factor of Error Type.

The main effects of Group, $F(2,169) = 9.98, p < 0.01, \eta^2_p = 0.11$, and Encoding Condition, $F(1, 169) = 16.46, p < 0.01, \eta^2_p = 0.09$, were significant. Overall, Shared condition participants had significantly higher recall error rates ($M = 0.26, SE = 0.01$) than those in the Distinctive condition ($M = 0.22, SE = 0.01, d = 0.38$). Post-hoc comparisons revealed that young adults that had a 2-day delay between study and test made significantly fewer recall errors ($M = 0.21, SE = 0.01$) than older adults under a 2-day delay ($M = 0.25, SE = 0.01, p < 0.01, d = 0.38$). Recall error rates were not significantly different between the older adults 2-day and young adult 7-day delay groups ($p = 0.57, d = -0.13$). The associated BF favoring the null hypothesis (see Table 1) indicated moderately strong evidence for the null hypothesis. The interaction between Group and Encoding Condition was also not significant, $F < 1$.

The Error Type main effect was also significant, $F(2,338) = 152.49, p < 0.01, \eta^2_p = 0.47$. Multiple comparisons on the different levels of the factor show that learners reported significantly higher rates of errors of commission ($M = 0.51, SE = 0.02$) than errors of omission ($M = 0.14, SE = 0.02, p < 0.01, d = 1.79$) or other errors ($M = 0.07, SE = 0.02, p < 0.01, d = 2.08$), which were significantly lower than omission error rates ($p = 0.04, d = 0.37$). No higher-order interactions involving Error Type were significant.

The data confirmed the expectation of high levels of intrusion errors with this category cued-recall task, which could impact both recognition memory and RCJ resolution while also indicating that the 7-day delay young and 2-day delay old adult groups were quite similar on free recall errors.

Recognition Performance

Average recognition memory accuracy was also analyzed in a 2 (Encoding Condition: Shared vs. Distinctive) x 3 (Group: Older Adult 2-day vs. Young Adult 2-day vs. Young Adult 7-day) GLM. Figure 4 plots the relevant cell means. The main effect of Condition was significant, $F(1,169) = 45.25, p < 0.01, \eta^2_p = 0.21$, reflecting higher recognition accuracy for individuals in the Distinctive condition, $M = 0.64, SE = 0.02$, compared to the Shared condition, $M = 0.46, SE = 0.02, d = 1.09$. The main effect of Group was also significant, $F(2,169) = 11.90, p < 0.01, \eta^2_p = 0.12$. Recognition accuracy for young adults in a 7-day delay, $M = 0.47, SE = 0.02$ was reliably lower than older adults, $M = 0.56, SE = 0.02, p < 0.01, d = -0.55$. Recognition accuracy levels were not significantly different between older adults and young adults given a 2-day delay, $p = 0.22, d = 0.36$. The interaction between Group and Encoding condition was also not significant, $F(2,174) = 1.58, p = 0.21, \eta^2_p = 0.02$.

RCJs

We analyzed mean RCJs in a 2 (Encoding Condition: Shared vs. Distinctive) x 3 (Group: Older Adult 2-day vs. Young Adult 2-day vs. Young Adult 7-day) GLM. Figure 5 plots the relevant means. The pattern in the RCJs closely aligned with patterns in recognition memory accuracy. The main effect of Encoding Condition was significant, $F(1,169) = 40.29,$

$p < 0.01$, $\eta_p^2 = 0.19$; individuals who were asked to generate Distinctive memory aids provided significantly higher RCJs, $M = 72.28$, $SE = 2.03$, than did participants in the Shared condition, $M = 53.85$, $SE = 2.08$, $d = 1.03$. The main effect of Group was also significant, $F(2,169) = 9.56$, $p < 0.01$, $\eta_p^2 = 0.10$. Mean RCJs for young adult participants in the 7-day delay condition ($M = 55.08$, $SE = 1.93$) were significantly lower than older adults ($M = 65.08$, $SE = 2.41$, $p < 0.01$, $d = -0.56$) in the 2-day delay condition, whereas participants' RCJs from the two 2-day delay conditions were not significantly different from each other ($p = 0.55$, $d = 0.22$). The interaction between Encoding Condition and Group was not significant, $F(2,169) = 0.56$, $p = 0.57$, $\eta_p^2 = 0.01$.

High-Confidence False Alarms

A major motivation of this investigation was to explore age differences in high-confidence false alarms (HCFAs) in the 5AFC recognition test. We defined HCFAs as incorrect recognition trials that were accompanied by RCJs greater than or equal to each participant's own median RCJ. We also distinguished HCFAs in terms of whether the false alarm was to a *new* lure (i.e. a recognition choice for an item that had not been seen before) or to an *old* lure (an item that had been co-presented with the target during study).

HCFAs were analyzed in a 2 (Encoding Condition: Shared vs. Distinctive) \times 3 (Group: Older Adult 2-day vs. Young Adult 2-day vs. Young Adult 7-day) \times 2 (Error Type: New Lure vs. Old Lure) mixed GLM (Figure 6 depicts the cell means and standard errors).

The main effect of Condition was significant, $F(1, 169) = 24.40$, $p < .001$, $\eta_p^2 = 0.13$. HCFAs were on average more common under shared-feature encoding ($M = .11$, $SE = .01$) than under distinctive encoding ($M = .06$, $SE = .01$), $d = 0.56$. This outcome confirmed the hypothesis that distinctive encoding would lower rates of HCFAs. Critically, the Group main effect was also reliable, $F(2, 169) = 31.18$, $p < .001$, $\eta_p^2 = 0.27$. HCFAs were more common for older adults ($M = .11$, $SE = .01$) relative to 2-day delay younger adults' HCFAs ($M = .01$, $SE = .01$), $t(169) = 6.25$, $p < .001$, $d = 0.94$, but did not differ from 7-day delay younger adults ($M = .13$, $SE = .01$), $t(169) = -1.31$, $p = .19$, $d = -0.16$. These effects confirmed the major age-related hypotheses; older adults had more HCFAs than younger adults when assessed with equal retention intervals, but when age groups were more similar in recognition memory performance, older adults had no greater tendency to commit HCFAs.

The within-subject main effect of Error Type was significant, $F(1, 169) = 60.74$, $p < .001$, $\eta_p^2 = 0.26$. HCFAs were more likely to be made for old lures (exemplars previously co-presented with targets at study) ($M = .12$, $SE = .01$) than for new lures ($M = .06$, $SE = .01$), $d = 0.53$. This effect confirmed the hypothesis that the familiarity of old lures was a potent amplifier of confidence in the false alarms. The type of lure opted for in false alarms both qualified and helped to clarify the main effects. The three-way interaction between Error Type, Condition, and Group was significant, $F(2,169) = 3.69$, $p = 0.03$, $\eta_p^2 = 0.04$, as were the two-way interactions between Error Type and Group, $F(2,169) = 6.38$, $p = 0.002$, $\eta_p^2 = 0.07$, Error Type and Condition, $F(1,169) = 11.21$, $p = 0.001$, $\eta_p^2 = 0.06$, and Group and Condition, $F(2,169) = 6.34$, $p < 0.01$, $\eta_p^2 = 0.07$.

As shown in Figure 6, the Error Type X Condition interaction reflected much higher HCFA rates for old lures, relative to new lures, in the shared encoding condition. We focused on decomposing the 3-way interaction. Planned comparisons tested the hypothesis that shared encoding (vs. distinctive) was the primary generator of the differentially higher HCFA effects for 2-day delay older adults, contrasted with 2-day delay younger adults³. The contrast was significant, $t(169) = 1.85$, one-tailed $p = .031$, $\eta_p^2 = 0.02$. The combination of old lures and shared encoding that generated the greatest mean HCFA. The same comparison comparing older adults with 7-day delay younger adults was not reliable, $t < 1$. The largest source of the 3-way interaction involved the same contrast, but comparing the two young adult groups, $t(169) = 2.73$, $p < .001$. As can be seen in Figure 6, the longer retention interval elevated HCFAs for younger adults, especially in combination with shared feature encoding.

We evaluated the null hypothesis stipulating no HCFA differences between 7-day delay younger adults and 2-day delay older adults using Bayesian analysis (see Table 1). The nonsignificant group difference ($F < 1$) was associated with a $BF = 4.35$, which indicated moderate evidence favoring the null hypothesis. When separated by Error Type, the BF for old lures was 2.60, whereas the BF for new lures was 3.89. These statistics were consistent with the argument that HCFAs did not differ when recognition memory performance was similar in the two age groups.

Matched sample analysis.—Our effort to use differential delays equated seven-day delay young adults and 2-day delay older adults on cued recall performance but resulted in reliably better recognition memory performance for the older adult group. To provide better control on level of recognition memory between age groups, we created a subsample of 7-day delay younger adults fully matched on recognition memory performance (see Wong et al., 2012, for a similar approach). We did so by pairing each older adult with one young adult in the 7-day delay group whose mean recognition accuracy was either the same as or closest to the older adults' recognition accuracy. In cases where there were multiple matches for a given older adult, one young adult was randomly chosen from the candidate set. This procedure generated groups of 55 persons with equal 5AFC recognition memory proportion correct ($M = 0.55$, $SD = .21$).

A repeated measures GLM detected robust effects of Error Type and Error Type X Condition interactions, as before. However, there were no effect of Group on HCFAs ($F < 1$), no effect of Group X Condition ($F < 1$), and no 3-way Group X Condition by Error Type interaction, $F(1, 106) = 1.56$, $p = .21$, $\eta_p^2 = 0.01$. The BF s for Group, Group X Condition, and Group X Condition X Error Type were 5.75, 4.72, and 6.37, respectively, constituting moderately strong evidence for the null hypothesis of age-group equivalence.

³These partial interaction contrasts were generated by analyzing difference scores (Old Lure – New Lure) to capture the within-subjects component of the comparisons. The difference scores were analyzed in an X Condition GLM, with the following partial interaction contrasts: The contrasts on the (OA, 7-day YA, 2-day YA) X Condition (Shared, Distinctive) subspace of the parameter vector **B** assigned L-matrix weights [-1 1 0, 1 -1, 0] to contrasting OA w 7-day YA, and weights [-1 0 1, 1, 0, -1] contrasting OA w 2-day YA.

Summary.—The analyses of HCFAs revealed three major findings. First, as hypothesized, age differences in HCFA occurred when both age groups were evaluated after a 2-day delay. However, this difference was eliminated when different delays more nearly equated the two age groups on memory performance, consistent with the hypothesis that the problem is one of memory, not faulty metamemory. The matched-sample analysis provided strong evidence for no age differences in HCFAs controlling on recognition memory accuracy. Second, for both age groups distinctive encoding resulted in much lower HCFAs, primarily as a function of reducing false alarms to previously co-presented exemplars. Third, the elevated rates of HCFAs under shared encoding for older adults were triggered by old lures, indicating misleading episodic familiarity deriving from original study of co-presented exemplars as a major source of the effects.

RCJ Resolution

To assess relationships between item-level RCJs and recognition test accuracy, we calculated Goodman-Kruskal gamma correlations computed for each participant (see Nelson 1984). Gamma correlation values can range anywhere between -1.0 (perfect negative correlation) and 1.0 (perfect positive correlation). We provide results from alternative indices of metacognitive resolution accuracy in Supplementary Materials, Table S1 (see Benjamin & Diaz, 2008, and Higham & Higham, 2019, for discussions regarding the limitations of gamma). Given the consistency in outcomes across the measures, we report only detailed inferential statistics on gamma correlations in the main paper.

A 2 (Encoding Condition: Shared vs. Distinctive) \times 3 (Group: Older Adult 2-day vs. Young Adult 2-day vs. Young Adult 7-day) GLM revealed a reliable main effect of Condition, $F(1,166) = 34.37, p < 0.01, \eta^2_p = 0.17$, indicating that individuals in the Distinctive condition had higher resolution, $M = 0.62, SE = 0.04$, than did those in the Shared condition, $M = 0.29, SE = 0.04, d = 0.96$ (see Figure 7). A significant main effect of Group, $F(2,166) = 3.58, p = 0.03, \eta^2_p = 0.04$, reflected only the effect of retention interval between the 2-day ($M = 0.54, SE = 0.06$) and 7-day ($M = 0.37, SE = 0.04, p = 0.03, d = 0.51$) young adult groups. Gamma correlations for older adults with a 2-day delay ($M = 0.46, SE = 0.06$) were not significantly different than the two young adult groups ($ps > .25$; see Table 1). The Condition \times Group interaction was not significant, $F(2,166) = 0.15, p = 0.86, \eta^2_p = 0.002$.

We also analyzed data from the matched samples to evaluate age differences in resolution fully controlling on age differences in recognition memory. There was no reliable Group difference in resolution, $F < 1, M_{Younger} = .43, SE = .05, M_{Old} = .46, SE = .05$, with a small sample mean difference favoring older adults. The BF favoring the null hypothesis was 5.43. There was also no Group \times Condition interaction, $F < 1$, with a BF favoring the null hypothesis of 5.03. There was still a robust effect of Condition, $M_{Shared} = .29, SE = .05, M_{Distinctive} = .61, SE = .05, F(1, 104) = 23.27, p < .001, d = 1.32$, with distinctive encoding generating much higher gamma correlations.

These outcomes confirmed the hypothesis that distinctive encoding enhances RCJ resolution. However, contrary to our hypothesis, we found no evidence of age differences in RCJ resolution, even when comparing older and younger adults with an equivalent 2-day delay. Given the effects observed on HCFAs, this outcome was unexpected. Furthermore,

the data for the matched samples favored the null hypothesis of no age differences in RCJ resolution when controlling on recognition memory accuracy.

Correlations of HCFAs and Gammas

To further assess the different outcomes, we tested our idea that HCFAs and gammas should be substantially (and negatively) correlated across individuals (after all, HCFAs are a major source of discordances that should generate lower gamma correlations). We aggregated the data across the three groups, and computed HCFA-Gamma correlations separately for the Shared and Distinctive encoding conditions. The Pearson r 's were large and reliably different from zero: -0.63 , 95% CI: $[-0.74 -0.49]$, for the Shared condition and -0.77 , 95% CI: $[-0.84 -0.66]$, for the Distinctive condition. A one-tailed r -to- z test of the condition difference was significant, $z = 1.74$, $p = .042$, indicating a higher negative correlation when distinctive encoding reduced the likelihood of HCFAs. Thus, the apparent discrepancy between the two outcome measures occurred despite the fact that these variables were highly intercorrelated.

Discussion

This investigation demonstrates that older and younger adults benefit from distinctive encoding on memory performance and retrieval monitoring in a complex category exemplar task prone to memory illusions. Individuals had difficulty in delayed recall and recognition of the designated target; they were highly likely to make category-consistent intrusion errors in cued-recall; and were prone to high-confidence recognition memory errors, often selecting co-presented (old) lures as correct in forced-choice recognition. Distinctive encoding dramatically repaired these memory illusions and greatly improved RCJ resolution.

These results cast a somewhat different light on the phenomenon of age-related elevations in HCFAs reported in prior studies (e.g., Dodson et al., 2007; Fandakova et al., 2013; Shing et al., 2009). As noted by Perfect and Stollery (1993), an open question is whether these errors merely reflect an indirect consequence of age-related memory impairments or an additional age-related deficit in the integrity of retrieval monitoring. By that account, reduced accessibility to diagnostic cues, such as recollection about aspects of the original encoding, could generate higher HCFAs for older adults. The present results are consistent with the memory-deficit account of HCFAs, at least in this task. Older adults showed equivalent benefits to younger adults of distinctive encoding for RCJ resolution. Furthermore, when age differences in memory performance was rendered similar by using different retention intervals (2 days for older adults, 7 days for younger adults), young adults and older adults showed similar levels of HCFAs and similar benefits of distinctive encoding for reducing HCFAs. A matched-sample analysis equating the age groups on recognition memory generated even stronger evidence in favor of age-equivalence in HCFAs. By contrast, when, as is typically done, retention intervals were identical for both age groups we observed age differences favoring young adults in HCFAs, consistent with previous studies.

We also found that lures that were originally presented with the target during encoding were a primary source of HCFAs, indicating that familiarity-based recognition of previously seen

exemplars in the Shared condition was a common route for generating elevated confidence when false alarming to those stimuli. Again, that tendency was repaired by distinctive encoding.

Age differences in HCFAs are often interpreted as a reflection of structural deficits in the aging brain (Devitt & Schacter, 2016), including especially the hippocampal formation (Fandakova et al., 2018; Shing et al., 2009), consistent with arguments about pattern-separation mechanisms that require hippocampal integrity (Lael & Yassa, 2018; Stark, Yassa, & Stark, 2010). The large effects associated with distinctive encoding for older adults indicate such effects are malleable depending on how the information was originally encoded. As argued by Hunt (2012), distinctive encoding has dramatic effects on reducing false memories because it renders differentiating details processed at study accessible during the subsequent memory test. Distinctive encoding, like other orienting tasks such as self-referencing (e.g., Hamami et al., 2011; Leshikar, Dulas, & Duarte, 2015) or intentional mediational strategies like interactive imagery (Hertzog, Sinclair, & Dunlosky, 2010; Kausler, 1994; Paivio, 2007; Richardson, 1998), benefit older adults' formation and retention of new memories. In our category task environment, older adults were able to capitalize on distinctive processing to avoid memory errors, just like younger adults.

Our results question whether the HCFA metric is the best way to capture judgment-based illusions in memory task environments. Although retention interval influenced HCFAs in younger adults, it had little impact on resolution as measured by Goodman-Kruskal gamma correlations. Nelson (1984; Gonzalez & Nelson, 1996) argued that ordinal within-person correlations generate indices of resolution that are independent of individual differences in level of memory performance, which seems to be supported by age-invariance in gammas, as well as the age-invariance in the strong effect of distinctive encoding on gammas. HCFAs, in contrast, appear to be sensitive to level of memory performance. These differences occurred despite the fact that HCFAs and gamma correlations were highly correlated, more so under distinctive encoding.

As noted in the Introduction, at least two processing pathways may produce memory errors in this task. Our results confirm that familiarity-based illusions based on prior exposure to recognition memory lures is an important route to memory errors. These results are fully consistent with demonstrations that misattributed familiarity is a source of memory illusions (e.g., Jacoby, 1999; Jacoby & Rhodes, 2006). Another pathway is semantically-based reconstruction (e.g., Brainerd & Reyna, 2015), a process reflected in false alarms to new category exemplars. In this study, false alarms to new lures were less common than false alarms to old lures. Nevertheless, rates of selecting category-consistent exemplars were elevated above the baseline under shared-feature encoding seen in 2-day delay younger adults for both older adults and 7-day delay young adults. The most likely explanation for this pattern is a generation of candidate (category-consistent) answers, a process apparently opposed by creation of target-specific memory traces under distinctive encoding. Furthermore, semantic reconstruction errors can be substantial in the absence of old lures that trigger familiarity-based memory errors. We previously presented results with a four-alternative forced-choice recognition test (in young adults only) in which all three lures were previously unseen exemplars, and false-alarm rates to these lures was substantial, despite the

fact that episodic familiarity would lead to a correct target identification (Hertzog, Curley, & Dunlosky, 2017).

Limitations and Future Directions

We acknowledge that these results may not generalize to other instances of age-related increases in HCFAs. The complex nature of the multiple-exemplar items differs from more traditional recognition memory tasks in other studies. Furthermore, age differences in recognition false alarms are more likely with yes-no recognition tests than forced-choice recognition memory tasks (e.g., Guerin, Robbins, Gilmore, & Schacter, 2012; Trelle et al., 2017), perhaps owing to greater susceptibility to erroneous generate-recognize heuristics (Hunt, Smith, & Toth, 2016). It is an open question whether we would have seen age differences in HCFAs when equating for overall memory performance in this task had we employed a yes-no recognition test. Age differences in RCJ resolution seen in other studies (e.g., Hines et al., 2009; Wong et al., 2012) were found with yes-no recognition tests, but not with the forced-choice tests in this study and Hertzog et al. (2010). Thus, the age-invariance in RCJ resolution found in this study may be due in part to presentation of the correct target among alternative lures in the 5AFC task that reduces older adults' susceptibility to memory illusions (Guerin et al., 2012).

The effect of retention intervals on HCFAs argues for an investigation of a broader range of retention intervals in studying memory illusions in general, and HCFAs and RCJ resolution in particular. Our investigation only examined two retention intervals that were expected (based on prior data) to produce similar memory performance in both age groups, and an age-equivalent 2-day retention interval expected to produce age differences in memory and metamemory. It is possible that age differences in retrieval monitoring will vary non-monotonically as a function of retention interval. We conceptualize this issue as akin to the metaphor of a Goldilocks Zone in astronomers' search for habitable planets in other solar systems. Relative to the energy released by its star, a habitable planet's orbit must be within a zone that produces temperatures conducive to life, such as liquid water. By analogy, retention intervals must be long enough to avoid ceiling effects (especially in the Distinctive Condition), short enough to avoid floor effects, but more critically, also of a duration that still provides accessibility to valid (diagnostic) metacognitive cues (like noncritical recollection of aspects of original encoding; Brewer et al., 2010; Hertzog, Fulton, Sinclair, & Dunlosky, 2014) enabling discrimination of incorrect candidates from the actual target. There may be age differences in the so-called 'sweet spot' for retention intervals in which cued-recall and recognition test performance is sufficiently error-prone yet access to diagnostic cues can avoid FAs and maximize RCJ resolution. Only a parametric study randomly varying retention intervals for persons of different ages could reveal the nature of these potentially non-monotonic tradeoffs.

We also acknowledge that our method included a prior cued-recall test with substantial intrusion errors that could have impacted 5AFC performance (Pansuwan et al., 2020). It is possible that the cued-recall test altered underlying memory representations in a way that affected recognition performance, such as by reinforcing recall intrusions errors.

Conclusion

Age differences in HCFAs in this category memory task depend on level of memory, and hence, accessibility to memory-based cues at the time of the recognition test. Thus, age differences in HCFAs observed in the literature do not necessarily implicate a metacognitive deficit in retrieval monitoring during the process of selecting a recognition test response option. Instead they may reflect consequences of age changes in the quality of memory representations that affect the cues available for making RCJs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was funded in part by a Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Research Training Grant (T32) from the National Institutes of Health (National Institute on Aging) Grant 5T32AG000175. The data reported here were featured in a 2018 paper at the Annual Convention of the Gerontological Society of America. The data associated with this paper have been archived in the Open Science Framework repository: <https://osf.io/87n4r/>. We thank our undergraduate research assistants for their help in data collection and analysis, especially Jayna Glover, Omer Oncul, Hannah Shotwell, Kirsten Reynolds, Joshua Parades, Kenley Tyler, Alysha Naran, Skyler Sigua, Aliyah Steele, Aiman Waris, Mackenzie Roy, Faizah Asif, Yusra Asif, Ana Supariwala, and Caroline Dalluge. For more information on our overall research program, please visit <http://www.hertzoglab.psychology.gatech.edu/>.

References

- Aizpurua A, & Koutstaal W. (2010). Aging and flexible remembering: Contributions of conceptual span, fluid intelligence, and frontal functioning. *Psychology and Aging, 25*, 193–207. doi: 10.1037/a0018198 [PubMed: 20230139]
- Balota DA, Dolan PO, & Duchek JM (2000). Memory changes in healthy older adults. In Tulving E & Craik FIM (Eds.), *The Oxford Handbook of Memory* (pp. 395–409). New York, NY US: Oxford University Press.
- Benjamin AS, & Diaz M. (2008). Measurement of relative metamnemonic accuracy. In Dunlosky J & Bjork RA (Eds.), *Handbook of memory and metamemory* (pp. 73–94). New York: Psychology Press.
- Brainerd CJ, & Reyna VF (2015). Fuzzy-trace theory and lifespan cognitive development. *Developmental Review, 38*, 89–121. doi: 10.1016/j.dr.2015.07.006 [PubMed: 26644632]
- Brainerd CJ, Reyna VF, & Howe ML (2009). Trichotomous processes in early development, aging, and neurocognitive impairment: An integrative theory. *Psychological Review, 116*, 783–832. doi: 10.1037/a0016963 [PubMed: 19839684]
- Brewer GA, Marsh RL, Clark-Foos A, & Meeks JT (2010). Noncriterial recollection influences metacognitive monitoring and control processes. *Quarterly Journal of Experimental Psychology, 63*, 1936–1942. doi:10.1080/17470210903551638
- Brydges CR, & Bielak AM (2020). A Bayesian analysis of evidence in favor of the null hypothesis in gerontological psychology (or lack thereof). *Journal of Gerontology B: Psychological Sciences, 75*, 58–66. doi:10.1093/geronb/gbz033
- Carr VA, Castel AD, & Knowlton BJ (2015). Age-related differences in memory after attending to distinctiveness or similarity during learning. *Aging, Neuropsychology, and Cognition, 22*, 155–169. doi: 10.1080/13825585.2014.898735
- Castel AD, Middlebrooks, & McGillivray S. (2016). Memory monitoring in old age: Impaired, spared, and aware. In Dunlosky J & Tauber SK (Eds), *Oxford Handbook of Metamemory* (pp. 519–535). Oxford, England: Oxford University Press.

- Chua EF, Schacter DL, & Sperling RA (2009). Neural basis for recognition confidence in younger and older adults. *Psychology and Aging*, 24, 139–153. [PubMed: 19290745]
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (rev. ed.). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.
- Craik FIM, & Jennings JM (1992). Human memory. In Craik FIM & Salthouse TA (Eds.), *The Handbook of Aging and Cognition* (pp. 51–110). Hillsdale, NJ: Erlbaum.
- Devitt AL, & Schacter DL (2016). False memories with age: Neural and cognitive underpinnings. *Neuropsychologia*, 346–359. doi: 10.1016/j.neuropsychologia.2016.08.030 [PubMed: 27592332]
- Dodson CS, Bawa S, & Krueger LE (2007). Aging, metamemory and high confidence errors: A misrecollection account. *Psychology and Aging*, 22, 122–133. doi: 10.1037/0882-7974.22.1.122 [PubMed: 17385989]
- Dunlosky J, & Hertzog C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition*, 29, 247–253. [PubMed: 11352207]
- Dunlosky J, & Metcalfe J. (2009). *Metacognition*. Thousand Oaks, CA: Sage Publications, Inc.
- Fandakova Y, Shing YL, & Lindenberger U. (2013). High confidence memory errors in old age: The roles of monitoring and binding processes. *Memory*, 21, 732–750. doi: 10.1080/09658211.2012.756038 [PubMed: 23305088]
- Fandakova Y, Sander MC, Grandy TH, Cabeza R, Werkle-Bergner M, & Shing YL (2018). Age differences in false memory: The importance of retrieval monitoring processes and their modulation by memory quality. *Psychology and Aging*, 33, 119–133. doi: 10.1037/pag0000212 [PubMed: 29494183]
- Gallo DA (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. New York: Psychology Press.
- Gallo DA, Bell DM, Beier JS, & Schacter DL (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory*, 14, 730–741. doi: 10.1080/09658210600648506 [PubMed: 16829489]
- Geraci L, McDaniel MA, Manzano I, & Roediger HL (2009). The influence of age on memory for distinctive events. *Memory & Cognition*, 37, 175–180. doi: 10.3758/MC.37.2.175 [PubMed: 19223567]
- Gonzalez R, & Nelson TO (1996). Measuring ordinal association in situations that contain tied scores. *Psychological Bulletin*, 119, 159–165. doi: 10.1037/0033-2909.119.1.159 [PubMed: 8559859]
- Greene NR, & Naveh-Benjamin M. (2020). A specificity principle of memory: Evidence from aging and associative memory. *Psychological Science*, 31, 316–331. doi: 10.1177/095679762090176 [PubMed: 32074021]
- Guerin SA, Robbins CA, Gilmore AW, & Schacter DL (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, 66, 68–78. doi: 10.1016/j.jml.2011.07.002 [PubMed: 22125357]
- Hamami A, Serbun SJ, & Gutchess AH (2011). Self-referencing enhances memory specificity with age. *Psychology and Aging*, 26, 636–646. doi: 10.1037/a0022626 [PubMed: 21480719]
- Hay JF, & Jacoby LL (1999). Separating habit and recollection in young and older adults: Effects of elaborative processing and distinctiveness. *Psychology and Aging*, 14, 122–134. doi: [PubMed: 10224637]
- Healey MK, & Kahana MJ (2016). A four-component model of age-related memory change. *Psychological Review*, 123, 23–69. doi: 10.1037/rev0000015 [PubMed: 26501233]
- Henkel LA, Johnson MK, & DeLeonardis DM (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General*, 127, 251–268. [PubMed: 9742716]
- Hertzog C. (1994). Repeated measures analysis in developmental research: What our ANOVA text didn't tell us. In Cohen SH & Reese HW (Eds.), *Life-span developmental psychology: Methodological contributions* (pp. 187–222). New York: Erlbaum.
- Hertzog C. (2016). Development of control processes in adulthood. In Dunlosky J & Tauber SK (Eds), *Oxford Handbook of Metamemory* (pp. 537–558). Oxford, England: Oxford University Press.

- Hertzog C, Curley T, Dunlosky J. (2017). Effects of a distinctiveness manipulation on the accuracy of retrieval monitoring. Paper presented at the 58th annual meeting of the Psychonomic Society, Vancouver, British Columbia, CA.
- Hertzog C, Dunlosky J, & Sinclair SM (2010). Episodic feeling-of-knowing resolution derives from the quality of original encoding. *Memory & Cognition*, 38, 771–784. [PubMed: 20852240]
- Hertzog C, Fulton EJ, Mandviwala L, & Dunlosky J. (2013). Older adults show deficits in retrieving and decoding associative mediators generated at study. *Developmental Psychology*, 49, 1127–1131. doi:10.1037/a0029414 [PubMed: 22799582]
- Hertzog C, Fulton EK, Sinclair SM, & Dunlosky J. (2014). Recalled aspects of original encoding strategies influence episodic feeling of knowing. *Memory & Cognition*, 42, 126–140. doi: 10.3758/s13421-013-0348-z [PubMed: 23835601]
- Hertzog C, & Shing YL (2011). Memory development across the lifespan. In Fingerman K, Berg CA, Antonucci T, & Smith J (Eds), *Handbook of Lifespan Developmental Psychology* (pp. 299–330). NY, NY: Springer.
- Higham PA, & Higham DP (2019). New improved gamma: Enhancing the accuracy of Goodman-Kruskal's gamma using ROC curves. *Behavior Research Methods*, 51, 108–125. doi: 10.3758/s13428-018-1125-5 [PubMed: 30264365]
- Higham PA, & Tam H. (2005). Generation failure: Estimating metacognition in cued recall. *Journal of Memory and Language*, 52, 595–617. doi: 10.1016/j.jml.2005.01.015
- Hines JC, Touron DR, & Hertzog C. (2009). Metacognitive influences on study time allocation in an associative recognition task: An analysis of adult age differences. *Psychology and Aging*, 24, 462–475. doi:10.1037/a0014417 [PubMed: 19485662]
- Huff MJ, & Aschenbrenner AJ (2018). Item-specific processing reduces false recognition in older and younger adults: Separating encoding and retrieval using signal detection and the diffusion model. *Memory and Cognition*, 46, 1287–1301. doi:10.3758/s13421-018-0837-1 [PubMed: 29959616]
- Hunt RR (2012). Distinctive processing: The co-action of similarity and difference in memory. In Ross BH (Ed.), *The psychology of learning of motivation* (Vol. 56, pp. 1–46). Oxford, United Kingdom: Elsevier. 10.1016/B978-0-12-394393-4.00001-7
- Hunt RR, Smith RE, & Toth J. (2016). Category cued recall evokes a generate-recognize retrieval process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 339–350. doi: 10.1037/xlm0000136
- IBM SPSS Statistics for Windows, version 25.0 (2017).
- Jacoby LL (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25, 3–22.
- Jacoby LL, & Rhodes MG (2006). False remembering in the aged. *Current Directions in Psychological Science*, 15, 49–53.
- JASP Team (2018). JASP. [Computer software]. Retrieved from <https://jasp-stats.org>.
- Jennings JM & Jacoby LL (2003). Improving memory in older adults: training recollection. *Neuropsychological Rehabilitation*, 13, 417–440. doi:10.1080/09602010244000390
- Johnson MK (2006). Memory and reality. *American Psychologist*, 61, 760–771. doi: 10.1037/0003-066x.61.8.760
- Kausler DH (1994). *Learning and memory in normal aging*. San Diego, CA: Academic Press.
- Kelley CM, & Sahakyan L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory & Language*, 48, 704–721.
- Koen JD, & Yonelinas AP (2014). The effects of aging, amnesic mild cognitive impairment, and Alzheimer's disease on recollection and familiarity: A meta-analytic review. *Neuropsychology Reviews*, 24, 332–354. doi: 10.1007/s11065-014-9266-5
- Koutstaal W. (2003). Older adults encode - but do not always use - perceptual details: Intentional versus unintentional effects of detail on memory judgments. *Psychological Science*, 14, 189–193. doi: 10.1111/1467-9280.01441 [PubMed: 12661684]
- Leal S, & Yassa MA (2018). Integrating new findings and examining clinical applications of pattern separation. *Nature & Neuroscience*, 21, 163–173. doi: [PubMed: 29371654]

- Leshikar ED, Dulas MR, & Duarte A. (2015). Self-referencing enhances recollection in both young and older adults. *Aging, Neuropsychology, and Cognition*, 22, 388–412. doi: 10.180/13725585.2014.957150
- Light LL, Prull MW, La Voie DJ, & Healy MR (2000). Dual-process theories of memory in old age. In Perfect TJ & Maylor EA (Eds.), *Models of cognitive aging* (238–300). London: Oxford University Press.
- Luszcz MA, Roberts TH, & Mattiske J. (1990). Use of relational and item-specific information in remembering by older and younger adults. *Psychology and Aging*, 5, 242–249. [PubMed: 2378689]
- MacDonald SWS, Stigsdottir-Neely A, Derwinger A, & Bäckman L. (2016). Rate of acquisition, adult age, and basic cognitive abilities predict forgetting: New views on a classic problem. *Journal of Experimental Psychology: General*, 135, 368–390. doi: 10.1037/0096-3445.135.3.368
- Masson MEJ, & Rotello CM (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 509–527. doi: 10.1037/a0014876
- Mickes L, Johnson EM, & Wixted JT (2010). Continuous recollection versus unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 843–863. doi: 10.1037/a0019755
- Morey RD, & Rouder JN (2011). Bayes Factor approaches for testing interval null hypotheses. *Psychological Methods*, 16, 406–419. doi: 10.1037/a0024377 [PubMed: 21787084]
- Naveh-Benjamin M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1170–1187. Doi: 10.1037/0278-7393.26.5.1170
- Naveh-Benjamin M, Brav T, & Levy O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging*, 22, 202–208. doi:10.1037/0882-7974.22.1.202 [PubMed: 17385995]
- Nelson DL, McKinney VM, Gee NR, & Janczura GA (1998). Interpreting the influence of implicitly activated memories and recognition. *Psychological Review*, 105, 299–324. [PubMed: 9577240]
- Nelson TO (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109. [PubMed: 6544431]
- Old SR, & Naveh-Benjamin M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, 23(1), 104–118. doi:10.1037/0882-7974.23.1.104 [PubMed: 18361660]
- Panuswan T, Breuer F, Gazder T, Lau Z, Cueva S, Swanson L, Taylor M, Wilson M, & Morcom AM (2020). Evidence for age-invariance in false recognition. *Memory*, 28, 172–186. doi: 10.1080/09658211.2019.1705351 [PubMed: 31868124]
- Perfect TJ, & Dasgupta ZZR (1997). What underlies the deficit in reported recollective experience in old age? *Memory & Cognition*, 25, 849–858. [PubMed: 9421571]
- Perfect TJ, & Stollery B. (1993). Memory and metamemory performance in older adults: One deficit or two? *The Quarterly Journal of Experimental Psychology Section A*, 46, 119–135. doi: 10.1080/14640749308401069
- Pollack I, & Hsieh R. (1969). Sampling variability of the area under the ROC-curve and of d' . *Psychological Bulletin*, 71, 161–173. doi: 10.1037/h0026862
- Psychology Software Tools, Inc. [E-Prime 2.0]. (2016). Retrieved from <https://www.pstnet.com>.
- Richardson JTE (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review*, 5, 597–614.
- Schacter DL, Israel L, & Racine C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, 40, 1–24. doi: 10.1006/jmla.1998.2611
- Shing YL, Werkle-Bergner M, Li S-C, & Lindenberger U. (2009). Committing memory errors with high confidence: Older adults do but children don't. *Memory*, 17, 169–179. doi: 10.1080/09658210802190596 [PubMed: 18608975]

- Smith RE (2006). Adult age differences in episodic memory: Item specific, relational, and distinctive processing. In Hunt RR & Worthen J (Eds.), *Distinctiveness and memory* (pp. 259–287). Oxford, U.K.: Oxford University Press.
- Smith RE, & Hunt RR (1996). Accessing the particular from the general: The power of distinctiveness in the context of organization. *Memory & Cognition*, 24, 217–225. [PubMed: 8881324]
- Stanislaw H, & Todorov N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31, 137–149. doi: 10.3758/BF03207704
- Stark SM, Yassa MA, & Stark CEL (2010). Individual differences in pattern separation performance associated with healthy aging in humans. *Learning and Memory*, 17, 284–288. doi: 10.1101/lm.1768110 [PubMed: 20495062]
- Trelle A, N., Henson RN, Green DAE, & Simons JS (2017). Declines in representational quality and strategic retrieval processes contribute to age-related increases in false recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 43, 1883–1897. doi: 10.1037/xlm0000412
- Tun PA, Wingfield A, Rosen MJ, & Blanchard L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging*, 13, 230–241. doi:10.1037/0882-7974.13.2.230 [PubMed: 9640584]
- Van Overschelde J, Rawson K, & Dunlosky J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language*, 50(3), 289–335.
- Wong JT, Cramer SJ, & Gallo DJ (2012). Age-related reduction in the confidence-accuracy relationship in episodic memory: Effects of recollection quality and retrieval monitoring. *Psychology and Aging*, 27, 1053–1065. doi: 10.1037/a0027686 [PubMed: 22449027]
- Zachary R. (1986). *Shipley Institute of Living Scale Revised manual*. Los Angeles, CA: Western Psychological Services.
- Zacks RT, Hasher L, & Li KZH (2000). Human memory. In Craik FIM & Salthouse TA (Eds.), *The handbook of aging and cognition* (pp. 293–357), Mahwah, NJ: Erlbaum.

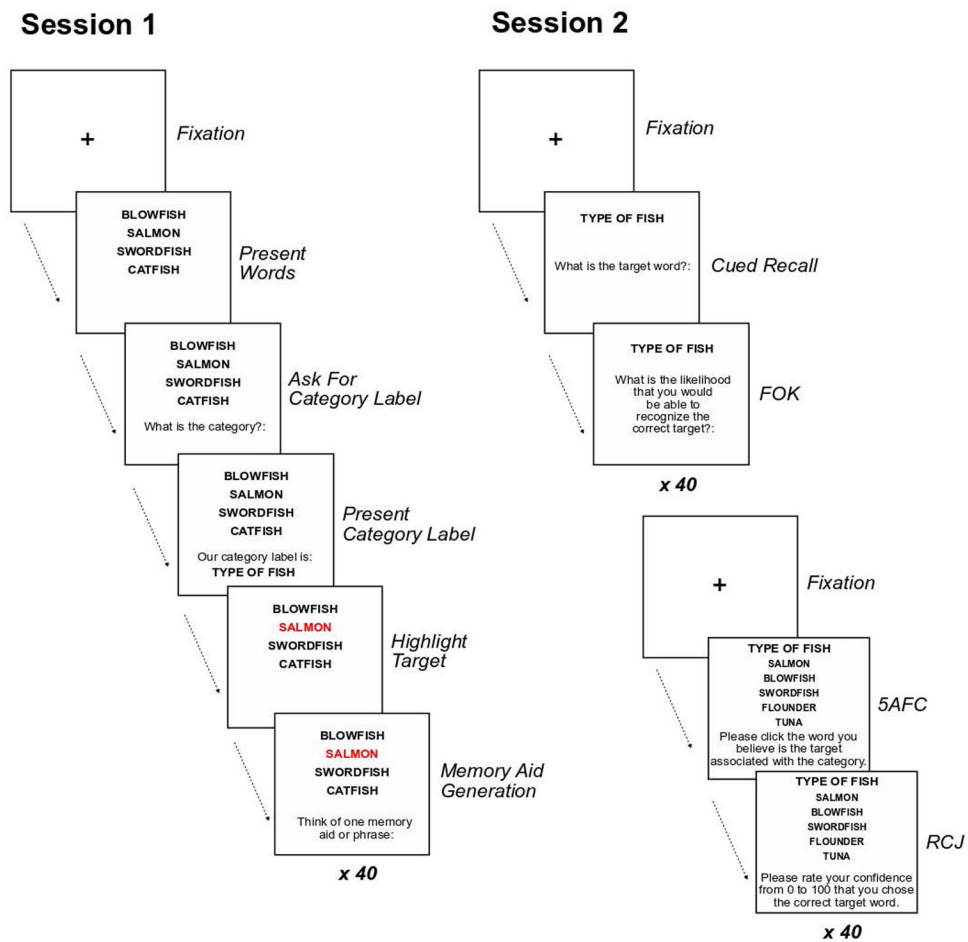


Figure 1. A graphical representation of Sessions 1 and 2 of the category memory task. For this example, SALMON is the target word highlighted in red.

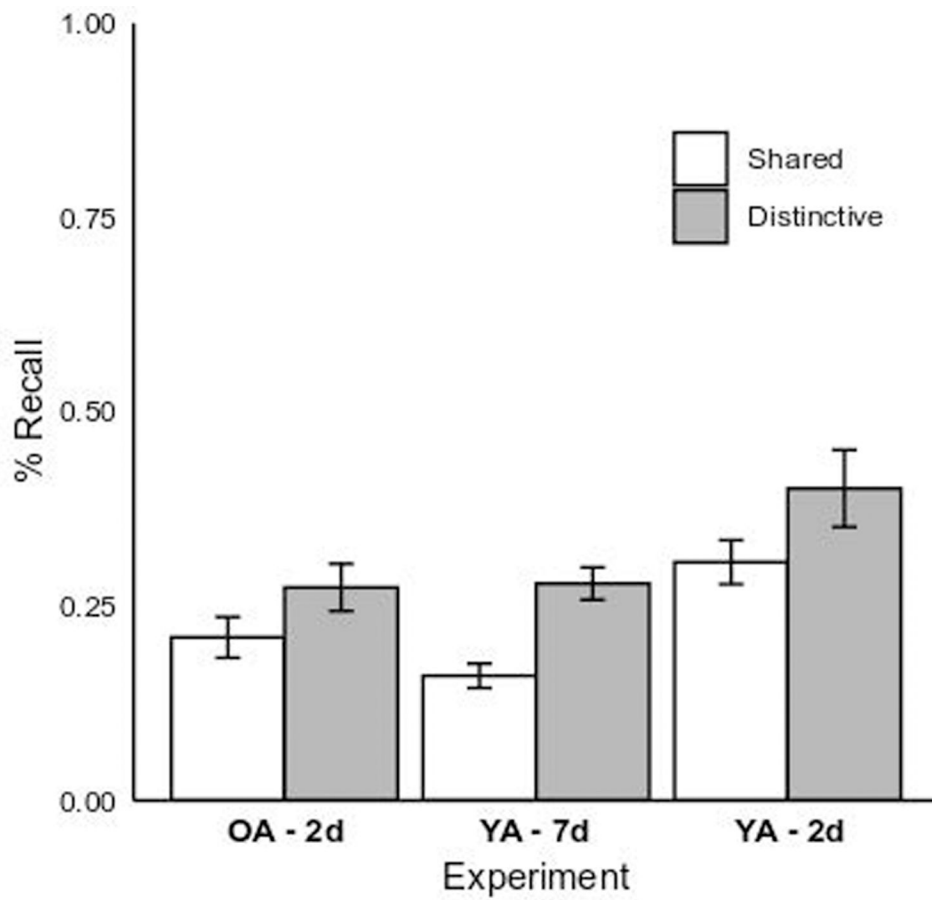


Figure 2. Mean cued-recall memory performance for the young adult 7-day (“YA – 7d”) and older and young adult 2-day (“OA – 2d” and “YA – 2d”) experimental groups by memory aid condition (Shared vs. Distinctive). Error bars represent 1 SE of the fitted least-squares means.

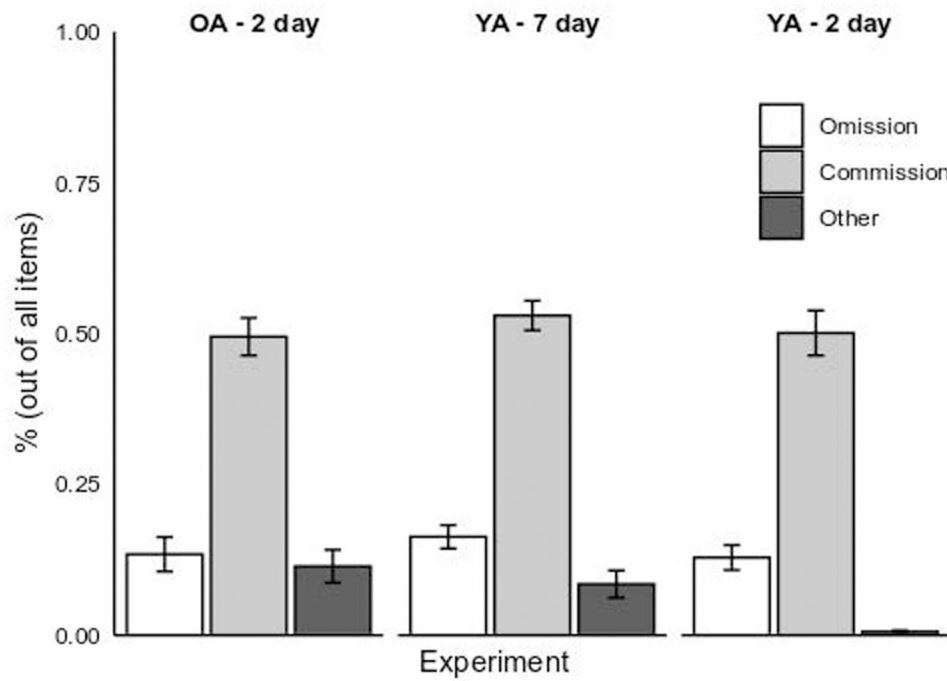


Figure 3. Recall error rates for the young adult 7-day (“YA – 7 Day”) and older and young adult 2-day (“OA – 2 day” and “YA – 2 day”) experimental groups. The error rates are broken down by memory aid condition (Shared vs. Distinctive) and error type (Omission vs. Commission, vs. Other). Error bars represent 1 SE of the fitted least-squares means.

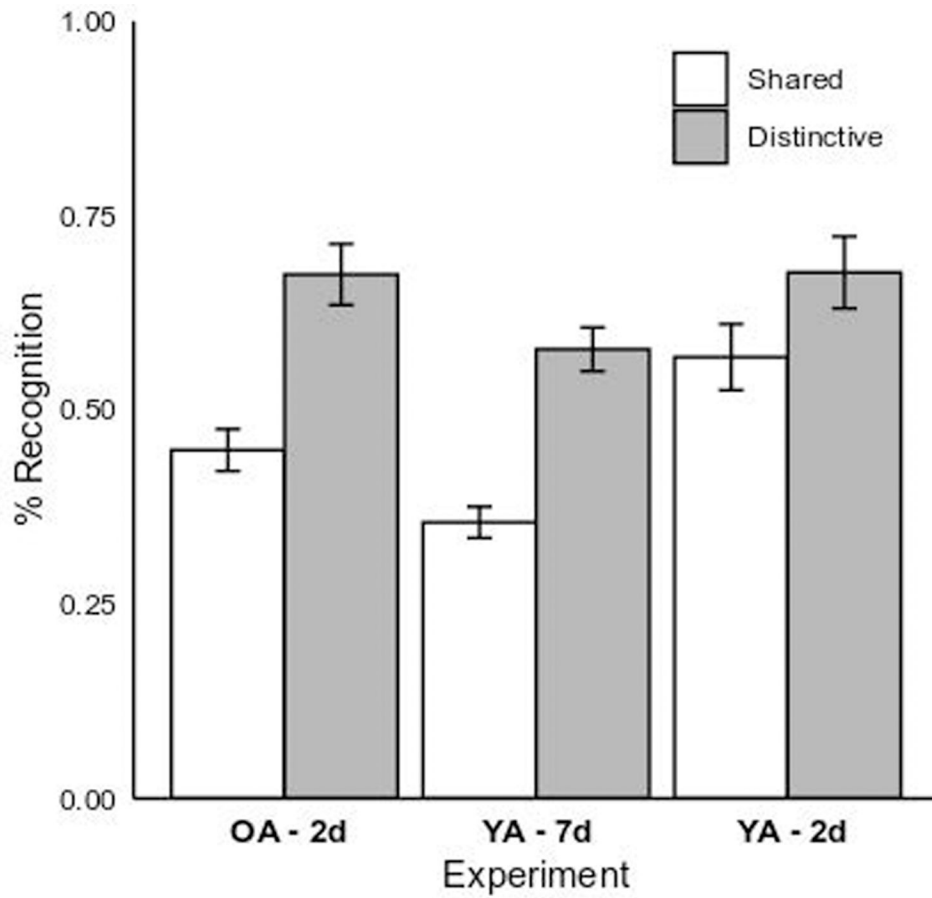


Figure 4. Mean recognition memory performance for the young adult 7-day (“YA – 7d”) and older and young adult 2-day (“OA – 2d” and “YA – 2d”) experimental groups by memory aid condition (Shared vs. Distinctive). Error bars represent 1 SE of the fitted least-squares means.

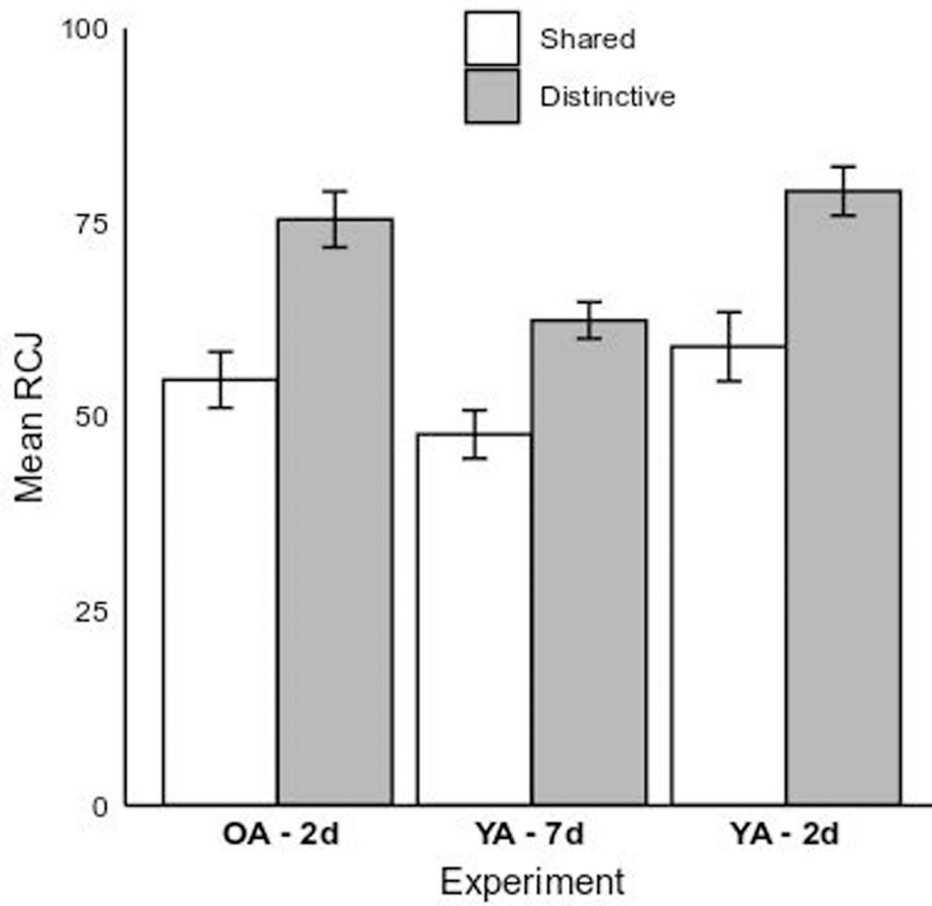


Figure 5. Mean RCJs for the young adult 7-day (“YA – 7d”) and older and young adult 2-day (“OA – 2d” and “YA – 2d”) experimental groups by memory aid condition (Shared vs. Distinctive). Error bars represent 1 SE of the fitted least-squares means.

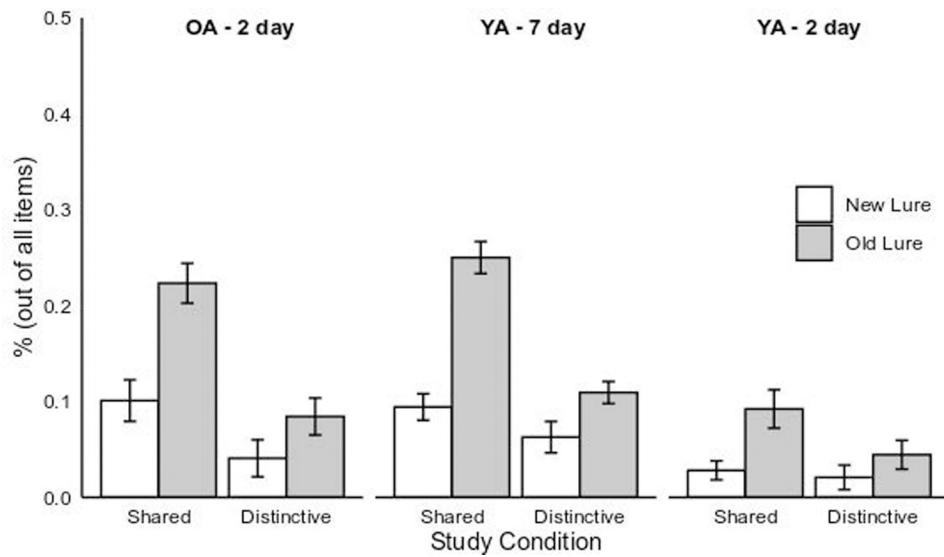


Figure 6. High-confidence false alarms rates for the young adult 7-day (“YA – 7 Day”) and older and young adult 2-day (“OA – 2 day” and “YA – 2 day”) experimental groups. The HCFA rates are broken down by memory aid condition (Shared vs. Distinctive) and error type (New vs. Old Lures). Error bars represent 1 SE of the fitted least-squares means.

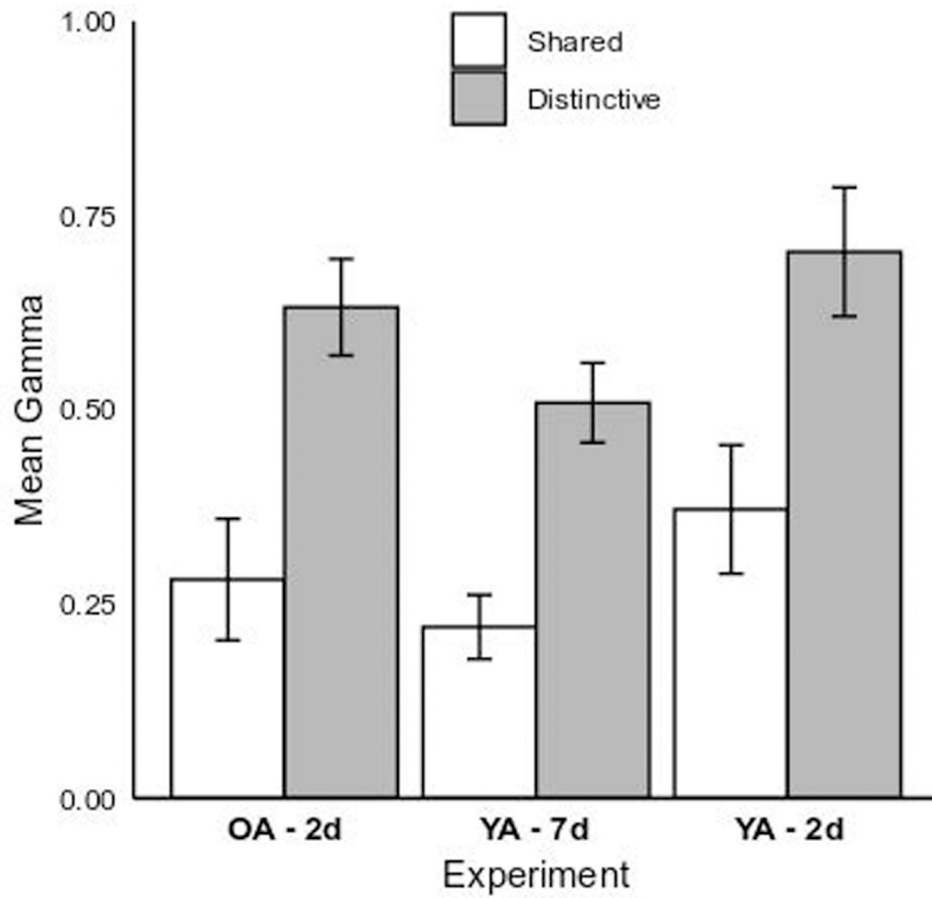


Figure 7. Mean gamma correlations of RCJs with recognition memory performance for the young adult 7-day (“YA – 7d”) and older and young adult 2-day (“OA – 2d” and “YA – 2d”) experimental groups by memory aid condition (Shared vs. Distinctive). Error bars represent 1 SE of the fitted least-squares means.

Effect sizes, significance, and Bayes Factors for Condition, Group contrasts (young adult vs. old adult), and Group contrasts X Condition interaction for all dependent measures.

Table 1.

Measure	d^a	d^b	d^b	$BF(H_0)$	η^2	$BF(H_0)$	Condition x 7-d YA vs. 2-d OA	Condition x 2-d YA vs. 2-d OA	η^2	$BF(H_1)$	$BF(H_1)$
Cued Recall	0.65*	0.79*	-0.16	3.77	0.01	2.56	0.39	<0.01	0.79	0.04	0.04
Recall Errors	0.38*	0.38*	-0.13	8.33	0.01	100	0.01	<0.01	0.04	1.2	1.2
Recognition	1.09*	0.36*	-0.55*	0.22	<0.01	1.05	0.95	<0.01	0.32	48.4	48.4
RCJs	1.03*	0.22	-0.56*	0.16	<0.01	0.83	1.21	<0.01	0.32	0.32	0.32
HCFAs	0.56*	0.94*	-0.16	4.35	<0.01	5.56	0.18	0.10*	48.4	0.32	0.32
RCJ											
Resolution	0.96*	0.25	-0.27	2.45	<0.01	2.44	0.41	<0.01	0.32	0.32	0.32

* $p < 0.01$

Condition refers to Distinctive vs. Shared encoding; YA = young adults; OA = older adults; 2-d = 2 day retention interval; 7-d = 7 day retention interval; d = Cohen's d ; $BF(H_0)$ = Bayes Factor in favor of the null hypothesis; $BF(H_1)$ = Bayes Factor in favor of the alternative hypothesis; RCJ = retrospective confidence judgments; HCFAs = high-confidence false alarms. See text for further details and results.

^a - Positive effect size indicates better performance in the Distinctive Condition

^b - Positive effect size indicates better performance by young adults