

The Effect of Age on Recollection is not Moderated by Differential Estimation Methods

Saad A. Alghamdi^{a,b,c*} Michael D. Rugg^{a,b,d,e}

^aCenter for Vital Longevity, University of Texas at Dallas, Dallas, TX

^bSchool of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, TX

^cDepartment of Psychology, King Saud University, Riyadh, Kingdom of Saudi Arabia

^dDepartment of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX

^eSchool of Psychology, University of East Anglia, Norwich, UK

***Correspondence:** Saad A. Alghamdi, University of Texas at Dallas, Center for Vital Longevity, 1600 Viceroy Drive, Suite 800, Dallas, TX 75235. saad.alghamdi@utdallas.edu

The Effect of Age on Recollection is not Moderated by Differential Estimation Methods

Episodic memory performance declines with increasing age. It has sometimes been reported that this decline is more marked when episodic recollection is estimated by 'objective' measures such as source memory performance than when it is estimated by 'subjective' measures such as the 'Remember/Know' procedure. Here, our main goal was to directly contrast recollection estimates derived from these procedures in the same samples of young and older participants (24 adults per age group, within-subjects manipulation of test procedure). Following identical study phases in which words were paired with either faces or scenes, participants' memories were assessed in separate test blocks using either Remember/Know or source memory procedures. Contrary to several prior reports, the deleterious effects of age on recollection estimates did not differ according to test type. Thus, we found no evidence that age differentially impacts subjective and objective recollection estimates. Additionally, and consistent with prior findings, effects of age on estimates of familiarity-driven recognition were small and non-significant.

Keywords: cognitive ageing, familiarity, Remember/Know, source memory, subjective memory

Introduction

It is well established that long-term memory declines with advancing age (e.g. Craik & McDowd, 1987; for review see Drag & Bieliauskas, 2009; Fraundorf, Hourihan, Peters, & Benjamin, 2019). Episodic memory - memory for personally experienced unique events - is especially vulnerable to the effects of age (e.g. Nilsson, 2003; Old & Naveh-Benjamin, 2008; Park et al., 1996; Spencer & Raz, 1995). However, there is evidence suggesting that the size of this age effect differs according to how episodic memory is tested. For example, in one study (Mark and Rugg, 1998), episodic memory was examined with both 'objective' (source memory) and 'subjective' (Remember/Know, or R/K) procedures for assessing recollection performance. Whereas the accuracy of source memory judgments was lower in older than in young adults (consistent with numerous other findings; see Cansino, 2009, for review), recollection performance as estimated from Remember judgments did not differ between the two age groups. Of importance, though, Mark and Rugg (1998) did not report false alarm rates for their R/K test, leaving open the possibility that recollection estimates were overestimated in their older participants. Somewhat analogous findings have been reported subsequently however (Duarte, Henson, & Graham, 2006; Duarte, Ranganath, Trujillo, & Knight, 2008). In both of these studies Duarte and colleagues divided their older age groups into two sub-groups based on item recognition performance, forming high- versus low-performing groups. The high-performing groups had equivalent recognition performance to that of young adults. The test phase was a two-step task in which subjects first performed an R/K judgment and then a source memory judgment on each trial. The findings from both studies indicated that while low-performing older

adults demonstrated impaired recollection for both R/K and source memory judgments, high-performing older adults demonstrated a significant reduction in recollection only when recollection was operationalized in terms of source accuracy.

In another study, an analogous dissociation between ‘objective’ and ‘subjective’ memory judgments was reported (Gallo, Korthauer, McDonough, Teshale, & Johnson, 2011). Here, the dissociation was between the reported vividness of a retrieved memory and source memory accuracy. Whereas older participants demonstrated lower source memory performance than young individuals, vividness and confidence ratings did not differ according to age group. A more recent study (Folville, D’Argembeau, & Bastin, 2020) reported a similar dissociation between subjective and objective measures of memory. In this study, the authors obtained both subjective (vividness ratings) and objective memory measures (free recall and a temporal order judgment) on each test trial. Whereas older adults performed less well than the young adults on the source and free recall tasks, memory vividness ratings did not differ between the two age groups.

The above-reviewed findings are consistent with the proposal that the effects of age on episodic memory differ according to whether recollection is operationalised through objective or subjective memory tests. However, other findings are inconsistent with this proposal. For instance, Prull, Dawes, Martin, Rosenberg, & Light (2006) contrasted recollection estimates derived from R/K and process dissociation (essentially, source memory) procedures, and reported equivalent age effects on the two classes of recollection estimate. In another study, Koen and Yonelinas (2016) also compared recollection estimates derived from R/K and process dissociation procedures, albeit with different study conditions for each procedure, and with age as a continuous variable (ranging between 40 and 81 yrs). The authors reported that the strength of the relationship between age and recollection

performance did not differ according to the estimation method. The findings from these two studies are consistent with those of a meta-analysis (Koen and Yonelinas, 2014) that yielded little or no evidence for weaker age effects on subjective than objective recollection estimates.

In contrast to measures of recollection, familiarity-based memory judgments (i.e. memory judgments made in the absence of qualitative information about the prior episode) have frequently (e.g. Jennings & Jacoby 1993; Jacoby, 1999; Titov and Knight 1997; Caldwell & Masson 2001; Howard, Bessette-Symons, Zhang, & Hoyer, Mancuso, Kliot, Arnold, & Dickerson, 2006; Wolk et al., 2013; Koen and Yonelinas, 2016), but not always (e.g. Prull et al., 2006; Toth and Parks 2006; Duarte, Graham, & Henson, 2010; Duzel, Yonelinas, & Heinze, 2010; Wang, De Chastelaine, Minton, & Rugg, 2012), been reported to be preserved with increasing age. In their meta-analytic review, Koen and Yonelinas (2014) reported that familiarity estimates tended to decline with increasing age in studies employing the R/K procedure, but not in studies using the Process Dissociation (PD) and Receiver-Operating Characteristics (ROC) procedures for separately estimating recollection and familiarity. Thus, the effect of age on familiarity also appears to be moderated by the estimation method. Although the main focus of the present study was to investigate whether age has a differential impact on different estimates of recollection, we took advantage of our employment of the R/K procedure to address the question of whether age impacts familiarity estimates derived from this procedure.

In summary, studies examining whether the effects of age on recollection performance differ according to whether recollection is estimated with objective or subjective measures have yielded inconsistent findings. Whereas five studies that employed a within-subject design to address this question reported larger age effects for objective

measures (Mark & Rugg, 1998; Duarte et al., 2006; Duarte et al., 2008; Gallo et al., 2011; Folville et al., 2020), other studies did not report this result (e.g. Prull et al., 2006; Koen & Yonelinas, 2016). The findings from these latter studies are consistent with the outcome of the meta-analysis conducted by Koen and Yonelinas (2014), in which across-study estimates of effect sizes were estimated. In light of these inconsistent findings, the goal of the present study was to further examine whether the effects of age on recollection performance differ according to whether recollection is estimated with subjective or objective methods. The study employed a within-subject design in which the sole manipulation (other than age group) was the nature of the memory test (R/K vs. source memory). Thus, we held constant both the nature of the experimental materials and the study task, allowing a direct contrast between the effects of age on the two classes of recollection estimate (unlike, for example, in the case of Koen and Yonelinas, 2016, where different study conditions were employed with each type of memory test). At issue is whether, relative to their younger counterparts, older individuals are less disadvantaged when recollection is estimated with an R/K than with a source memory procedure (Mark & Rugg, 1998; Duarte et al., 2006; Duarte et al., 2008; Gallo et al., 2011; Folville et al., 2020), or whether instead the two classes of recollection estimate yield equivalent age-related impairments (Prull et al., 2006; Koen and Yonelinas, 2014; Koen & Yonelinas, 2016).

Materials and Methods

Participants

Participants' characteristics and neuropsychological test scores are detailed in Table 1.

Participants were recruited from the University of Texas at Dallas and surrounding communities

and compensated \$15 per hour for their participation. They gave informed consent to their participation in accordance with University of Texas at Dallas Institutional Review Board procedures. All participants learned English during childhood and had normal or corrected to normal vision. Exclusion criteria included reported cardiovascular disease other than treated hypertension, a history of psychiatric or neurological disorder, recently used psychotropic medication, history of substance abuse, or inadequate neuropsychological test performance (see below). The neuropsychological test battery comprised the following tests: The Mini-Mental State Examination (MMSE), the California Verbal Learning Test-II (CVLT) (Delis, Kramer, Kaplan, & Ober, 2000), the Symbol Digit Modality Test (SDMT) (Smith, 1982), Trail Making Tests A and B (Reitan & Wolfson, 1985), the Digit Span Forward and Backward test of the Wechsler Adult Intelligence Scale- Revised (Wechsler, 1981), the Wechsler Test of Adult Reading (WTAR; Wechsler, 2001), the UDS Category Fluency test (Benton, 1968), and List 1 of the Raven's Progressive Matrices (J. Raven, Raven, & Courth, 2000). Potential participants were excluded if they scored < 26 on the MMSE, or > 1.5 standard deviations (SDs) below age-appropriate norms on any CVLT sub-test or on any two of non-memory tests. The rationale for these criteria was to exclude participants who might be in the early stages of mild cognitive impairment and hence at risk of pathological cognitive decline.

A total of 30 older and 27 young participants were initially recruited to the study. Nine participants were excluded for the following reasons: 2 young and 2 older participants were excluded because of below-criterion neuropsychological test performance; 2 older adults were excluded due to their participation in a prior similar study; one young participant was excluded due to experimenter error that compromised data collection, and 2 older participants were excluded due to their item memory performance falling outside cut-offs determined on the basis

of pilot data ($Pr < .2$ in one case, and a false alarm rate $> .4$ in the other. Inclusion of these participants did not, however, alter the findings reported below in any substantive way). We are unable to report the neuropsychological test results of one young subject whose scores were missing. Also, we were unable to administer the neuropsychological test battery to two older subjects who did not return for the test session. The final analyses included twenty-four participants in each age group.

The critical contrast in the present study is that between the z-scores of the older adults according to the nature of the retrieval test (Figure 3). This contrast directly tests the hypothesis that the effects of age on subjective estimates of recollection are weaker than age effects on objectively derived estimates. Whereas prior studies have reported such a dissociation (Duarte et al., 2006, 2009; Rugg and Mark, 1998) none employed a direct contrast analogous to that employed in the present study, disallowing the estimation of a predicted effect size. However, in each study the reported effect sizes (Cohen's d) exceeded 1.2 for age effects on source memory performance and were coupled with null effects for subjective (Remember) estimates. A sensitivity analysis indicated that the N of 24 for our older sample affords 80% power to detect an effect size of 0.52 (one-tailed repeated measures t -test, $p < .05$).

Materials

The experimental materials comprised 480 experimental items and 58 practice items. The experimental items were divided between two study-test cycles. Each study phase included 96 object names paired with a picture of either a face or a scene. An equal number of older and young female and male faces of mixed ethnicity, and equal numbers of urban and rural scenes were included in each study block. In the test phase, we used a 2:1 ratio of old to new words such

that all words from the previously shown study list were used (96 old words per block) along with an additional 48 new words that served as lures. Each study and test block contained a 30 second rest interval at its halfway point. The blocks always began with 2 filler items, and a further 2 fillers immediately followed each of the mid-way rests.

Procedure

Overview

Each participant undertook two study/test cycles, each of which comprised a single study and test block. In the study phase of each cycle (Figure 1A) participants were presented with word-picture pairs, one pair at a time, and were required to form a mental image associating the object denoted by the word with the depicted picture. Specifically, if the picture was a face, instructions were to form a mental image of the depicted person interacting with the denoted object. If the picture was a scene, the requirement was to imagine the denoted object moving around within the scene. Participants rated the vividness of the images they formed using the scale described below. Immediately after a study phase, participants undertook a two-stage memory test that involved an initial test of item memory followed, for positive judgments, by either an R/K or a source memory judgment.

Table 1

	Young	Old	p-value
N	24*	24**	-
Age	22.7 (3.3)	70.9 (3.81)	-
Sex	8M; 16F	12M; 12F	.24***
Yrs of Full-Time Education	15.74 (2.47)	16.91 (2.56)	.13

MMSE	29.22 (1.20)	29.14 (1.04)	.81
CVLT Short Delay – Free	13.22 (2.41)	10.32 (2.75)	<.001
CVLT Short Delay – Cued	13.65 (1.9)	11.82 (2.17)	.005
CVLT Long Delay – Free	14 (1.91)	11.68 (2.78)	.003
CVLT Long Delay – Cued	14.04 (1.94)	12.14 (2.44)	.006
CVLT Recognition – Hits	15.65 (.78)	14.73 (1.49)	.01
CVLT Recognition – FA	.35 (.65)	2.59 (2.34)	<.001
SDMT	62.96 (6.46)	47.32 (9.84)	<.001
Digit Span Total	19.26 (5.61)	16.73 (3.81)	.1
Trail Making A (secs)	18.56 (4.95)	32.67 (11.72)	<.001
Trail Making B (secs)	42.95 (15.89)	76.64 (32.9)	<.001
Category Fluency	23.78 (5.89)	20.41 (4.01)	.02
WTAR	40.96 (5.41)	42.64 (6.03)	.33
Ravens List 1	12.34 (23.3)	23.23 (78.89)	<.001

* Data from one subject were missing for the neuropsychological analysis.

** Two participants did not undertake the neuropsychological tests.

*** Assessed with Chi-square test

Note. Standard deviations are given in parentheses. The p-values were computed using Welch t-tests. Digit span total equals the sum of forward and backward span. MMSE = Mini-mental State Exam; CVLT = California Verbal Learning Test II; FA: False Alarm; SDMT = Symbol-Digit Modalities Test; WTAR = Wechsler Test of Adult Reading.

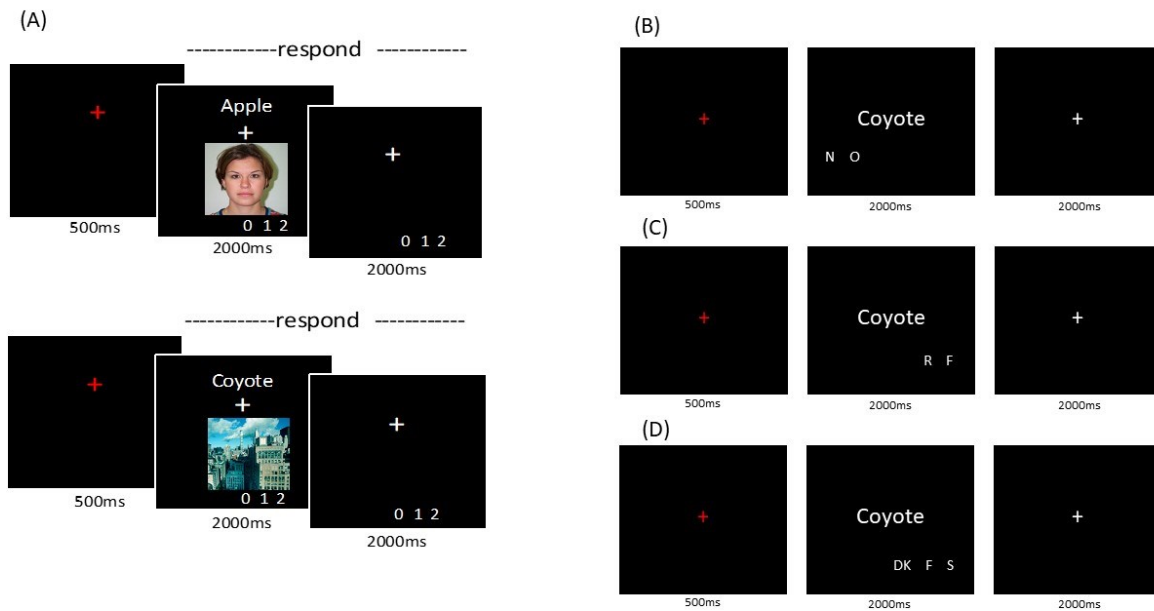


Figure 1. (a) Schematic of the study task with an example of a word-face pair (upper) and a word-scene pair (lower). (b) Schematic of the item memory test. (c) Schematic of the R/K procedure. (d) Schematic of the source task.

Instructions and practice

Before each study-test cycle, participants were given a script describing the study and test procedures. After they indicated that they understood the instructions, they were asked to explain the tasks thoroughly in their own words and were corrected if necessary. Participants were not informed of the nature of the study or test procedures to be undertaken in the second study-test cycle until after they had completed the first cycle.

In an initial self-paced practice study session, participants were presented with 12 word-picture pairs and were required to verbalize the mental image they formed during each trial and the rationale behind their vividness judgment. This session was followed by a timed session

using an additional 12 word-picture pairs. Subsequently, participants were instructed, and given practice, on the test to be employed in that study-test cycle. As for the study practice sessions, participants were first presented with a self-paced test session using 12 old and 5 new words, followed by a timed practice phase using a different set of 12 old and 5 new words.

Study phase

Each study pair was preceded by a red fixation cross for a duration of 500ms. The pairs were presented for a duration of 2s and followed by a white fixation cross for a further 2s., providing a total of 4000 ms in which to make a response. As noted above, study instructions for word-face pairs were to imagine the depicted person interacting with the object denoted by the word. For word-scene pairs, instructions were to imagine the denoted object moving around within the scene. Vividness ratings were made via a laptop keyboard and comprised the options of “not vivid” (scored as 0), “somewhat vivid” (score of 1), or “very vivid” (score of 2).

Test phases

As schematized in Figure 1B, each test trial began with a red fixation cross that was displayed for 500ms. The cross was replaced by the test word, which was presented for 2s and replaced in turn by a 2s duration white fixation cross. Instructions were to first make an item memory (old/new) judgment on the test word (Figure 1B). If the word was endorsed ‘New’, the requirement was simply to wait for the onset of the next trial. If the judgment was ‘Old’ (i.e. studied), a second judgment was undertaken, which in different study-test cycles was either the source memory or the R/K test (Figures 1C & 1D). The duration of each test trial was held constant and participants were instructed to make both responses prior to the change of color of the fixation cross, which occurred 4 s after the onset of the test item. Responses were, however,

logged until 4.5 s following the onset of the test item. Belying the possibility that older participants were under greater time-pressure on the two memory tests than the young adults, the number of missed responses in the older age group was very small (a mean of approximately 1 trial for the R/K task and 2 trials for the source task), and did not differ significantly from the number of omissions made by the young subjects.

As just noted, for trials on which an old judgment was given, the subsequent judgment differed between the two study-test cycles. In one cycle, participants were required to make an R/K judgment (Figure 1C). They were instructed to respond ‘Remember’ (R) when recognition of the test word was accompanied by memory for at least one qualitative, reportable detail from the test item’s study episode. If recognition was not accompanied by recollection of any such detail, a ‘Familiar only’ (K) judgment was required. During the initial practice phase participants were requested to verbally describe the basis of the Remember judgments they made.

In the other test cycle, participants undertook a source memory task (Figure 1D). The requirement was to judge whether the test word had been associated at study with a face or a scene; a ‘don’t know’ option was also available.

The ordering of the two test procedures was counterbalanced across participants. Thus, half of the older and young subjects were assigned to perform the source task in the first study-test cycle, and the Remember/Know task in the second cycle. The remaining half of each group was assigned to perform the R/K task in the first cycle, and the source task in the second cycle¹.

Data Analysis

Item recognition

Item recognition accuracy (Pr) was measured as $p_{Hit} - p_{False\ Alarm}$, where p_{Hit} is the probability of a Hit (i.e. studied words correctly judged 'old') and $p_{False\ Alarm}$ is the probability of a false alarm (unstudied words incorrectly judged 'old'). Pr estimates were subjected to a three-way mixed ANOVA with the factors of age group, study material (face vs. scene), and memory task (R/K vs. Source).

Remember/Know

Memory performance on the R/K test was estimated according to the independent R/K procedure. (Yonelinas & Jacoby, 1995). Thus, recollection was operationalized as: $p_{RHit} - p_{RFalse\ Alarm}$, where p_{RHit} is the probability of an R judgment to an old word, and $p_{RFalse\ Alarm}$ is the probability of an R judgment to a new word. Familiarity was estimated as: $(p_{KHit}/1 - p_{RHit}) - (p_{KFalse\ Alarm}/1 - p_{RFalse\ Alarm})$, where p_{KHit} is the probability of a 'K' judgment to an old word, and $p_{KFalse\ Alarm}$ is the probability of a K judgment to a new word. Each of these estimates was subjected to a two-way mixed ANOVA with the across-participants factor of age group) and the within-participants factor of study material (face vs. scene).

Source Memory

Source recollection (p_{SR}) was estimated with a measure (Park, Uncapher, & Rugg, 2008) derived from single high threshold theory (Snodgrass and Corwin, 1988), using the equation: $p_{SR} = (p_{Sr} - .5 * (1 - p_{DK})) / (1 - (.5 * (1 - p_{DK})))$. This equation is a modification of the standard single high threshold model, in which the probability of a correct binary response is assumed to reflect the sum of the probability that a test probe will return an above-threshold signal diagnostic of its source, and the probability that a below threshold signal will be associated with

a ‘lucky guess’ (i.e. $p_{\text{Correct}} = p_{\text{Recollect}} + (1 - p_{\text{Recollect}}) * .5$), to accommodate a ‘don’t know’ response option. Group differences in pSR were examined with a Welch t-test.

Relative effect sizes for subjective and objective recollection estimates

To contrast the older group’s recollection estimates according to the test procedure, estimates of source accuracy (pSR) and recollection (pR) for each older participant were z-scored relative to the means and standard deviations of the corresponding estimates of the young group. A paired t-test was then used to contrast the older participants’ z-scores according to the test procedure.

In the case of theoretically important null findings, evidence supporting the null hypothesis was estimated by computing a Bayes factor for the relevant contrast. We employed the BF_{10} metric, according to which a factor < 1 indicates evidence favoring the null, while a factor > 1 is indicative of evidence favoring the alternate hypothesis (Wagenmakers et al., 2018).

Results

Study phase

Vividness Ratings

Means (and SDs) for the vividness ratings from the study phase are reported in Table 2, segregated by memory task, subsequent memory judgment, and age group. The data were subjected to a three-way mixed effects ANOVA with the factors of age group, subsequent memory judgment and memory task. The ANOVA revealed a main effect of subsequent memory, $F(2, 92) = 22.22$, $MSe = .04$, $p < .001$, $\text{partial-}\eta^2 = .33$, with no other significant effects (the main effect of age group approached significance, $p = .066$, while all other effects were far from significant, $\text{min } p = .37$). Post-hoc tests revealed that mean ratings were significantly higher for

study items that went on to be recollected (i.e. later recognized items accorded Remember or accurate source judgments) than they were either for items attracting accurate item memory judgments only ($t(47) = 6.06, p < .001, d = .88$) or items that went on to be missed ($t(47) = 5.43, p < .001, d = .78$). The ratings for these latter two classes of item did not significantly differ ($p = .53$).

Table 2. Means (and SDs) vividness ratings at study segregated by memory task, subsequent memory judgment, and age group.

Task	Source Task			Remember/Know Task		
	S Hit	S incorr	Miss	R Hit	K Hit	Miss
Young	1.32 (.31)	1.15 (.34)	1.20 (.4)	1.31(.30)	1.14(.39)	1.14(.41)
Older	1.17 (.34)	0.98 (.38)	1.01 (.36)	1.14(.44)	0.97(.44)	0.95(.35)

RTs

The means of median RTs associated with the study judgments are given in Table 3. These data were submitted to a three-way ANOVA with the factors of age group, memory task, and vividness rating (not, somewhat, and very vivid). The ANOVA revealed a main effect of rating, $F(2, 92) = 6.40, MSe = 66512.64, p = .002$, partial- $\eta^2 = .12$, with no other effects approaching significance. Post-hoc comparisons showed that RTs averaged across tasks and age groups were significantly lower for ‘very vivid’ than for ‘not vivid’ ($t(47) = 3.05, p = .007, d = .44$) or ‘somewhat vivid’ ($t(47) = 3.14, p = .007, d = .45$). RTs to these latter two classes of judgments did not significantly differ ($p = .93$).

Table 3. Means (and SDs) of median RTs at study segregated by memory task, vividness ratings, and age group.

	Source Task			Remember/Know Task		
	0	1	2	0	1	2
Young	2386(643)	2475(561)	2373(572)	2460(663)	2435(736)	2337(684)
Older	2587(393)	2526(433)	2413(428)	2556(493)	2566(488)	2414(487)

Test Phase

RTs

The means of median RTs for judgments made during the test phase are reported in Table 4. RTs are shown for studied items accorded correct item memory judgments and represent the latencies from the onset of the test item to the second judgment (i.e. either the R/K or the source memory judgment, depending on the retrieval test). The data from each test task were submitted to two-way ANOVAs with the factors of Age (Older vs. Young) and judgment (either R hit vs K hit, or Source correct vs source incorrect/don't know). In the case of the R/K data, the ANOVA revealed main effects of age, $F(1, 46) = 15.95$, $MSe = 384776.06$, $p < .001$, $\text{partial-}\eta^2 = .26$, and judgment, $F(1, 46) = 67.92$, $MSe = 89641.94$, $p < .001$, $\text{partial-}\eta^2 = .60$, and a significant interaction between these two factors, $F(1, 46) = 8.43$, $MSe = 89641.94$, $p = .006$, $\text{partial-}\eta^2 = .16$. Post-hoc contrasts revealed that older adults' R and K judgments were significantly slower than those of young adults ($t(47) = 4.86$, $p < .001$ and $t(47) = 2.34$, $p = .045$, respectively). Within group contrasts revealed that both young and older age participants demonstrated slower responses for K judgments ($t(23) = -7.88$, $p < .001$ and $t(23) = -3.78$, $p = .001$, respectively). As is evident from

Table 4, the age group x judgment interaction reflected the fact that the age differences were greater for R than for K judgments.

The ANOVA for the RTs in the source task revealed main effects of age, $F(1, 43)=26.22$, $MSe=451698.18$, $p<.001$, $\text{partial-}\eta^2=.38$, and judgment, $F(1.68, .72.03)=9.72$, $MSe=278972.16$, $p<.001$, $\text{partial-}\eta^2=.18$, and a non-significant interaction between the two factors, $F(1.68, 72.03)=2.8$, $MSe=278972.16$, $p=.08$, $\text{partial-}\eta^2=.06$. These findings reflect faster responses on the part of the young adults and, in both age groups, for correct source judgments.

Table 4. Means (and SDs) of RTs at test segregated by memory task, judgment, and age group.

	Source Task			Remember/Know Task	
	S Correct	S incorr	DK	R Hit	K Hit
Young	1895(657)	2250(490)	2493(538)	1643(342)	2324(510)
Older	2758(466)	2906(493)	3021(523)	2326(557)	2659(511)

Item recognition

Means (and SDs) for item discriminability (Pr) are given in Table 5, segregated by age group, material and memory task. A three-way mixed effects ANOVA with the factors of age group (young vs. older), study material (face vs. scene), and memory task (RK vs. source) revealed main effects of age, $F(1,46) = 34.04$, $MSe= .07$, $p< .001$, $\text{partial-}\eta^2= .43$, and Study Material, $F=(1, 46) = 5.78$, $MSe= .01$, $p<.02$, $\text{partial-}\eta^2= .11$, but no effect of Study Task, $F(1, 46) = .69$, $MSe= .01$, $p= .41$, $\text{partial-}\eta^2= .02$, and no interaction effects. The results reflect lower performance for older adults compared to young adults, and for test words paired with scenes compared to those paired with faces.

Table 5. Means (and SDs) for item memory according to study material, memory task, and age groups.

Task	Remember/Know		Source Task	
Item type	Face	Scene	Face	Scene
Young	.71 (.11)	.70 (.14)	.72 (.15)	.69 (.17)
Older	.52 (.17)	.48 (.17)	.50 (.18)	.45 (.13)

Remember/Know task

Means (and SDs) for Recollection and Familiarity estimates are reported in Table 6 for the different age groups and study materials (note that one older participant failed to make any K judgments. Thus, familiarity estimates are based on the remaining 23 older participants). A two-way mixed ANOVA on the recollection estimates revealed main effects of age, $F(1, 46) = 28.19$, $MSe = .06$, $p < .001$, $\text{partial-}\eta^2 = .38$, and study material, $F(1, 46) = 11.21$, $MSe = .01$, $p = .002$, $\text{partial-}\eta^2 = .20$, but no interaction between these factors. Recollection estimates were lower for older compared to young adults and were also lower for words paired with scenes than with faces. In light of the near-significant age differences in mean vividness ratings accorded at study, we repeated this analysis as an ANCOVA employing the ratings for items that went on to receive R judgments as a covariate of no interest. The effect of age remained significant ($F(1, 45) = 23.84$, $MSe = .027$, $p < .001$, $\text{partial-}\eta^2 = .346$).

ANOVA of the familiarity estimates revealed no significant effects. Notably, the effect of age was far from significant, $F(1, 44) = .17$, $MSe = .04$, $p = .68$, $\text{partial-}\eta^2 = .004$, and this negative finding was supported by a Bayes Factor that anecdotally favored the null hypothesis ($BF_{10} = .35$).

Table 6. Means (and SDs) for Recollection and Familiarity estimates by study material and age group.

Item type	Recollection			Familiarity		
	Face	Scene	All	Face	Scene	All
Young	63. (.14)	.57 (.17)	.60 (.14)	.30 (.16)	.37 (.16)	.34(.14)
Old	.38 (.22)	.30 (.18)	.34(.19)	.29 (.18)	.30 (.19)	.29(.17)

Source task

Mean probability of source correct (SC), source incorrect (SI), and source DK judgments, segregated according to study material and age group, can be found in Table 7. We used these data to derive source accuracy estimates (pSR), which were contrasted across age groups using an unpaired t-test. The outcome of the test was significant, $t(43) = 5.49$, $SE_{Diff} = 0.06$, $p < .001$, Cohen's $d = 1.58$, with older adults demonstrating lower source accuracy ($M = .34$, $SD = .18$) than young adults ($M = .67$, $SD = .23$). As with the pR metric (see above), this age effect remained significant when the mean vividness ratings accorded study items that went on to attract source correct judgments were employed as a covariate ($F(1, 45) = 27.11$, $MSe = .042$, $p < .001$, $\text{partial-}\eta^2 = .376$).

We also examined whether participants in either group were biased in favor of one of the two sources in the absence of source information. To investigate this, we conducted an ANOVA on the source judgments associated with item false alarms, segregated by the factors of age group and study material. No significant effects were revealed.

Table 7. Means (and SDs) for the probability of source correct, source incorrect, Don't Know segregated by age group, study material.

Study material	Young		Older	
	Face	Scene	Face	Scene
SC	.80 (.15)	.78 (.22)	.64 (.20)	.55 (.19)
SI	.05(.06)	.06 (.05)	.19 (.17)	.24 (.12)
DK	.15 (.14)	.17 (.20)	.17 (.17)	.21 (.18)
pSR	.67 (.23)		.34 (.18)	

Note. SC: studied items correctly judged as old and attributed to their source. SI: studied items correctly judged as old but were attributed to the wrong source. DK: studied items correctly judged as old and received DK judgment on the source task. pSR: source accuracy estimated from the source task.

Correlation of recollection estimates derived from the source and the R/K procedures

Correlations between the recollection estimates for the two age groups derived from each procedure were computed to examine whether the strength of the associations between the two estimates differed by age. As depicted in Figure 2, there were significant correlations between the estimates in both groups (young: $r(23) = .71, p < .001$; older: $r(23) = .61, p = .001$). Fisher's procedure revealed that the two correlations did not significantly differ ($z = .55, p = .58$).

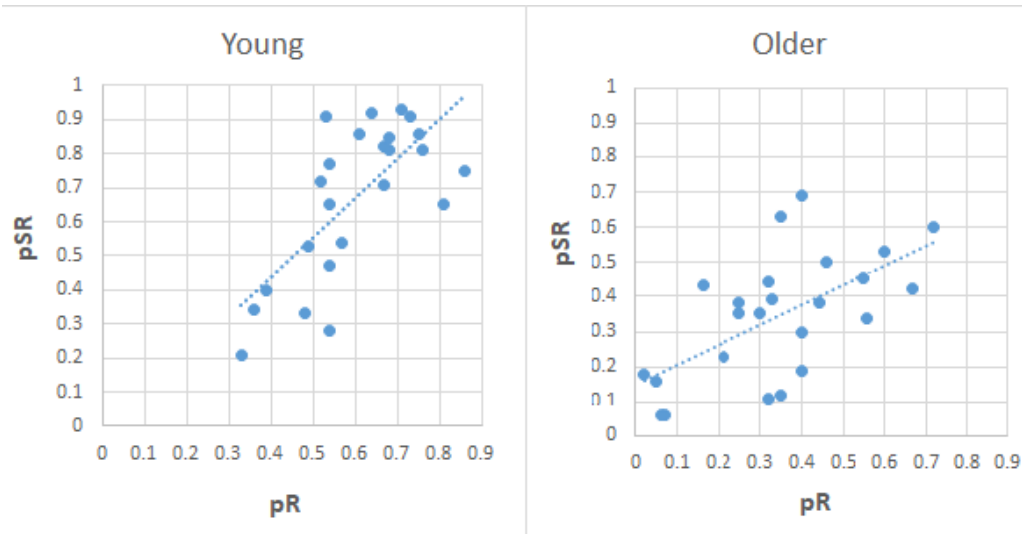


Figure 2. Scatter plots depicting recollection estimates derived from the source task (y-axis) and the R/K task (x-axis) for young and older participants.

Relative effect sizes for age differences in subjective and objective recollection

To allow a direct comparison between the effects of age on the two recollection estimates, older participant's pR and pSR scores were z-transformed with respect to the corresponding scores of the young group (see Figure 3). As is evident from the figure, the z-scores for pR were numerically *lower* than those for pSR ($M = -1.86$, $SD = 1.4$ and $M = -1.43$, $SD = .79$, respectively). The scores did not however differ significantly, $t(23) = -1.901$, $SE_{Diff} = .23$, $p = .07$, Cohen's $d = -.39$. The associated Bayes factor strongly favored the null hypothesis against the pre-experimental, one-tailed alternative hypothesis of lower scores on the source task, $BF_{10} = .08$. Inspection of Figure 3 suggests that older participants' z scores were more variable for the R/K than for the source task. This impression was confirmed with a Pitman's test (Pitman, 1939), $t(22) = 3.52$, $p = .002$. As is evident from comparing the data in Tables 6 and 7, this difference is a consequence of the young participants demonstrating greater variability in their pSR than in their

pR scores (SDs =.23 and .14 respectively, $t(22) = 3.43$, $p < .002$) rather than a disparity in the variances of these metrics in the older adults (SDs =.18 and .19 respectively).

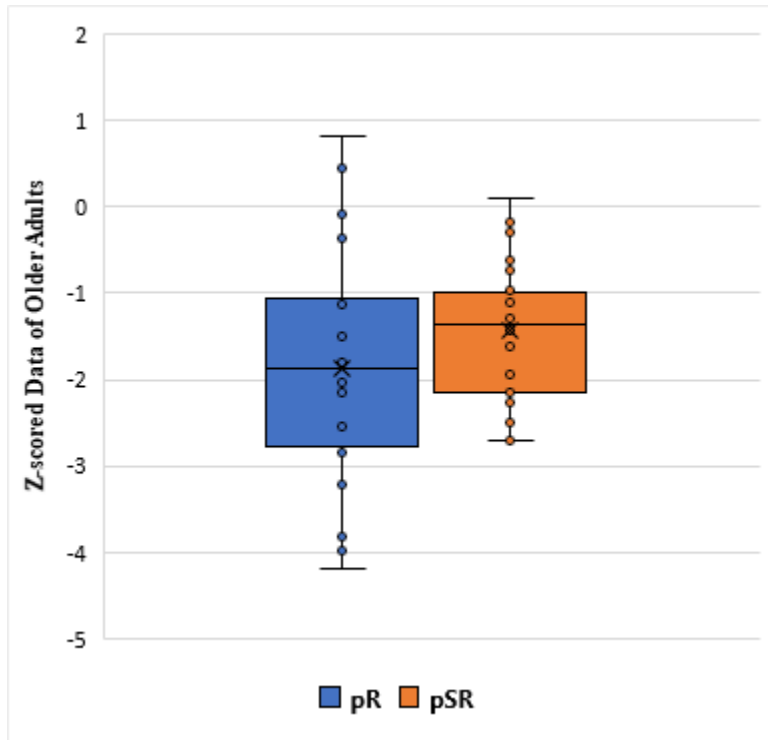


Figure 3. Box and whisker plots and single older subjects' z scores for recollection estimates derived from the R/K (left) and source (right) retrieval tests.

Comparison of recollection estimates between high- and low-performing older adults

In two prior studies (Duarte et al., 2006; Duarte et al., 2008), differential effects of age on subjective and objective recollection estimates were confined to older participants with relatively high item memory performance. To determine whether this was the case here, we divided our older sample on the basis of a median split of their item memory scores, and conducted a two-

way ANOVA on their z-scores with the factors of sub-group (high-performing vs. low-performing) and memory task (R/K vs. source). The results revealed a main effect of group, $F(1, 22) = 17.192$, $MSe = 1.146$, $p < .001$, $\text{partial-}\eta^2 = .44$, reflecting the poorer performance of the low-performing sub-group, and a marginal effect of task, $F(1, 22) = 3.81$, $MSe = .58$, $p = .06$, $\text{partial-}\eta^2 = .15$, but no interaction between the two factors, $F(1, 22) = 2.41$, $MSe = 5.6$, $p = .13$, $\text{partial-}\eta^2 = .1$. Thus, we found no evidence that item memory performance moderated the effects of age on estimates of subjective vs. objective recollection.

Effects of task order

In a further subsidiary analysis, we examined whether the factor of test order impacted the foregoing findings. For example, participants who performed the source task prior to the R/K task might have been biased to base their Remember judgments on successful recall of source information. We examined this possibility with two ANOVAs, one for each task, and in each case employing the factors of age group and task order. For the R/K procedure, the results revealed a main effect of age group, $F(1, 44) = 27.48$, $MSe = .03$, $p < .001$, $\text{partial-}\eta^2 = .38$, but no evidence of an order effect, $F(1, 44) = .56$, $MSe = .03$, $p = .46$, $\text{partial-}\eta^2 = .01$, or an interaction between the order and age group $F(1, 44) = 3.52$, $MSe = .03$, $p = .07$, $\text{partial-}\eta^2 = .07$. An analogous ANOVA of source accuracy estimates revealed an identical pattern of findings (age: $F(1, 44) = 28.70$, $MSe = .04$, $p < .001$, $\text{partial-}\eta^2 = .4$; order: $F(1, 44) = .44$, $MSe = .04$, $p = .51$, $\text{partial-}\eta^2 = .01$; order by age: $F(1, 44) = .93$, $MSe = .04$, $p = .34$, $\text{partial-}\eta^2 = .02$).

Discussion

The main goal of the present study was to examine whether age has a differential impact on recollection estimates derived from an objective versus a subjective test of memory. In contrast

with the findings of some prior studies (Mark & Rugg, 1998; Duarte et al., 2006; Duarte et al., 2008) we found that the impact of age on recollection was not moderated by the estimation method. Instead, and consistent with other studies (Prull et al. 2006; Koen & Yonelinas, 2016; for review see Koen & Yonelinas, 2014), we identified a robust and equivalent effect of age on recollection estimates derived from each method. In further support of the proposal that the two classes of recollection estimate are supported by a common set of processes in the two age groups, the across-participant correlations between the estimates were highly robust and age invariant (cf. Koen & Yonelinas, 2016).

Relative effect sizes for age differences in subjective and objective recollection

Why did our results differ from those of Mark and Rugg (1998) and Duarte et al. (2006; 2008)? As noted previously, the findings of Mark and Rugg (1998) are ambiguous, since the authors failed either to report false alarm rates or to correct R hit rates for false alarms. Thus, it is possible that appropriately corrected estimates would have revealed an effect of age on subjective recollection estimates. As was also noted previously, Duarte and colleagues segregated their older participants into high- versus low-performing sub-groups based on item memory performance. While they identified an age effect on source accuracy in both sub-groups, an age-related decline in recollection estimates derived from the R/K procedure was evident only in the low-performing group. Thus, the authors suggested that variability among older adults in overall memory capacity might moderate the impact of age on subjective recollection estimates. They proposed two (mutually non-exclusive) explanations for why ‘high performing’ older adults demonstrated a dissociation between subjective and objective estimates. First, these adults might be especially prone to ‘non-criterial’ recollection, that is, to retrieving qualitative information from a study episode that is non-diagnostic for the required source

judgment. Second, the relatively low source accuracy of high performing individuals might reflect an age-related decline not in recollection *per se*, but in the executive/control processes required to use the information to guide behaviour. In the present study, however, we found no evidence for such a dissociation in our older group between the two classes of recollection estimate, even when we adopted Duarte et al.'s (2006, 2008) approach of dividing the participants according to their item memory performance. We note that a critical difference between the experimental procedures of Duarte and colleagues and those adopted here is the manner in which subjective and objective recollection estimates were acquired. In the prior studies, both estimates were obtained on the same test trials: participants were required first to make an R/K judgment, and then a source judgment, on each test trial. By contrast, in the present study the two memory tests were administered in separate study-test cycles. It is possible that – as was entertained but rejected by Duarte et al. (2006) – the employment of successive judgments encouraged participants to treat the R/K judgment as more of a confidence rating than a report of mnemonic content. By this argument, the failure in those studies to find an age effect on subjective recollection estimates is a reflection of the high levels of confidence that the high performing older adults had in their item memories.

A reviewer drew our attention to another characteristic of our experimental procedure that might account for the discrepant findings between the present and the aforementioned prior studies. In the present study, we employed elaborative study tasks that likely facilitated the formation of item-item associations between study words and their background 'contexts', which comprised trial-unique scenes and faces. Thus, while the source memory task required only a categorical judgment (face vs. scene), there was the potential in both the source and Remember/Know tasks for the recollection of specific item-item associations. This situation contrasts with study tasks

where the source information comprises a simple physical feature of the study event (for example, presentation of the study item above vs. below fixation, as in Duarte et al. 2008, or in a male vs. a female voice, as in Mark and Rugg, 1998). The present finding of equivalent age differences in subjective and objective recollection estimates might therefore be a consequence of the availability of highly differentiated ‘item-context’ associations that, presumably, were more accessible to our young participants in both the source and R/K tasks.

While a full resolution of this issue will require further research, we think it unlikely that the present findings are merely a reflection of the idiosyncrasies of the study task. First, the findings are consistent with those reported in prior studies that compared the effects of age on objective and subjective measures of recollection while manipulating source information through a physical feature of the study task (temporal order in Prull et al., 2006; sensory modality in Koen and Yonelinas, 2016; see also Koen and Yonelinas; 2014). Notably, the effect sizes (Cohen’s d) for the effects of age on the objective and subjective measures reported in Prull et al. (2006) were 1.43 and 1.33 respectively, comparable to those obtained in the present study (1.58 and 1.56, respectively). Second, whereas Duarte et al. (2008) employed study location as a source feature, in Duarte et al. (2006) the source manipulation varied which of two study tasks (manipulability vs. animacy judgment) the test items were subjected to, affording the opportunity to encode richer and more idiosyncratic item-context associations. Nonetheless, the effect sizes for the age differences in recollection estimates in the two studies were broadly comparable: the effects were negligible in both cases for the subjective estimates, but sizeable for the estimates derived from the source judgments (1.62 and 1.23 in the later and earlier study respectively). Together, these different observations suggest that there is scant support for the possibility that findings of

differential effects of age on subjective and objective recollection estimates depend on the nature of the study task.

The foregoing discussion offers a possible resolution of the disparate findings obtained in the present and prior studies contrasting the effects of age on recollection estimates from R/K and source memory procedures. It does not, however, address the question of why older individuals are as likely as young participants to rate retrieved memories as highly vivid, while at the same time demonstrating lower performance on objective measures of memory performance (Gallo et al., 2011; Folville et al., 2020). The present data offer little or no insight into the basis of these findings beyond the strong implication that the processes supporting such meta-memorial judgments are distinct from those underlying subjective reports of recollection elicited by the R/K procedure. We conjecture that the apparent divergence between these two forms of subjective report (we say apparent, since to our knowledge no study has compared vividness and Remember judgments in the same participants while holding study conditions and experimental materials constant) might reflect the differing constraints of the two procedures: whereas vividness instructions allow participants considerable flexibility in how they interpret the task instructions and set criteria for the different ratings, R/K instructions typically (as in the present study) emphasize the need to restrict Remember judgments to trials where a *reportable* contextual detail about the study episode was recollected.

Two other aspects of our findings concerning subjective and objective recollection estimates are worthy of brief mention. First, consistent with prior reports (e.g. Jackson & Schacter, 2004) vividness ratings at study predicted memory performance, such that those study items that went on to be recollected attracted higher ratings than items that were either recognized but not recollected or for which recognition failed. Of importance, this relationship

did not differ according to age group or retrieval test, and, in addition, age differences in recollection estimates remained robust when vividness ratings from study were included as covariates. Thus, we found no evidence that the present findings for the effects of age on subjective and objective recollection estimates were confounded by age differences in the vividness ratings given at study.

Second, unlike in the older age group, the variance in the source recollection estimates (pSR) of the young adults was robustly greater than it was for their subjective recollection estimates (pR). We have no explanation for this disparity in the variances of the two estimates, or why it was evident for the young age group only. We note however that, should they prove replicable, these findings hint at the possibility that young adults are more sensitive to the differential demands of the two retrieval tests than their older counterparts.

Effect of age on familiarity

Familiarity estimates derived from the R/K task did not significantly differ according to age group. This null finding is consistent with numerous prior reports (e.g. Jennings and Jacoby 1993; Jacoby, 1999; Titov and Knight 1997; Caldwell and Masson 2001; Howard et al. 2006; Wolk et al., 2013; Koen and Yonelinas, 2016; but see Prull et al., 2006; Toth and Parks 2006; Duarte et al., 2010; Duzel et al. 2011; Wang et al., 2012 for conflicting findings). In a meta-analysis, Koen and Yonelinas (2014) reported that whereas familiarity estimates derived from process dissociation and ROC procedures were age-invariant, estimates derived from the R/K procedure were not. In an ensuing empirical study, however (Koen & Yonelinas, 2016), the same authors reported age invariance for familiarity estimates derived from all three procedures, a finding consistent with that reported here.

Before concluding, we note some limitations of the present study. First, it employed a cross-sectional design, and it remains to be seen whether similar findings will emerge in future studies based on longitudinal designs. Second, the study employed only a single example of a subjective and an objective memory test and a single set of experimental materials. Thus, it remains to be demonstrated that the present findings generalize to a wider range of tests and materials. In particular, we note that it is possible that there are variants of the R/K procedure (for example, those that do not impose a response deadline, unlike in the present case, or those that employ a more liberal definition of subjective recollection) that are less sensitive to the effects of age than the procedure adopted in the present study.

In conclusion, the main goal of the present study was to address the question of whether age differentially impacts objective and subjective recollection estimates, as operationalized by source memory and R/K procedures respectively. Contrary to some prior reports, we found no evidence of a disproportionate impact. Thus, the present findings converge with those of several prior studies to suggest that subjective and objective recollection estimates are similarly sensitive to the effects of age.

Acknowledgements:

This research was supported by the National Science Foundation [grant number 1633873].

1. Due to experimenter error, there was a one yoked pair of young and older subjects who performed the source rather than the R/K task first, so 13 yoked pairs of young and older subjects performed the source task first, and 11 yoked pairs first performed the R/K task

2. Although these analyses are for the study task, we included the factor of memory task to examine whether ratings differed according to task.
3. For the sake of clarity, we refer to study materials as face (i.e. words paired with face pictures during the study task) or scene (words paired with scene pictures during the study task).

References:

- Benton, A. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia*, 6(1), 53-60.
- Caldwell, J., & Masson, M. (2001). Conscious and unconscious influences of memory for object location. *Memory & Cognition*, 29(2), 285-295.
- Cansino, S. (2009). Episodic memory decay along the adult lifespan: A review of behavioral and neurophysiological evidence. *International Journal of Psychophysiology*, 71(1), 64-69.
- Craik, F., McDowd, J., & Craik, I. (1987). Age Differences in Recall and Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(3), 474-479.
- Delis DC, Kramer JH, Kaplan E, Ober BA. (2000). *California Verbal Learning Test, 2nd ed.* San Antonio, TX: The Psychological Corporation.
- Drag, L., & Bieliauskas, L. (2010). Contemporary review 2009: Cognitive aging. *Journal of Geriatric Psychiatry and Neurology*, 23(2), 75-93.

- Duarte, A., Henson, R., & Graham, K. (2008). The effects of aging on the neural correlates of subjective and objective recollection. *Cerebral Cortex*, 18(9), 2169-2180.
- Duarte, A., Ranganath, C., Trujillo, C., & Knight, R. (2006). Intact recollection memory in high-performing older adults: ERP and behavioral evidence. *Journal of Cognitive Neuroscience*, 18(1), 33-47.
- Folville, A., D'Argembeau, A., & Bastin, C. (2020). Deciphering the relationship between objective and subjective aspects of recollection in healthy aging. *Memory*, 28(3), 362–373.
- Fraundorf, S., Hourihan, K., Peters, R., & Benjamin, A. (2019). Aging and recognition memory: A meta-analysis. *Psychological Bulletin*, 145(4), 339-371.
- Gallo, D., Korthauer, L., McDonough, I., Teshale, S., & Johnson, E. (2011). Age-related positivity effects and autobiographical memory detail: Evidence from a past/future source memory task. *Memory*, 19(6), 641-652.
- Howard, M., Bessette-Symons, B., Zhang, Y., & Hoyer, W. (2006). Aging selectively impairs recollection in recognition memory for pictures: Evidence from modeling and receiver operating characteristic curves. *Psychology and Aging*, 21(1), 96-106.
- Jacoby, L. (1999). Ironic Effects of Repetition: Measuring Age-Related Differences in Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 3-22.
- Jackson, O., & Schacter, D. L. (2004). Encoding activity in anterior medial temporal lobe supports subsequent associative recognition. *NeuroImage*, 21(1), 456–462.
- Jennings, J., & Jacoby, L. (1993). Automatic Versus Intentional Uses of Memory: Aging, Attention, and Control. *Psychology and Aging*, 8(2), 283-293.
- Koen, J., & Yonelinas, A. (2014). The Effects of Healthy Aging, Amnesic Mild Cognitive Impairment, and Alzheimer's Disease on Recollection and Familiarity: A Meta-Analytic Review. *Neuropsychology Review*, 24(3), 332-354.
- Koen, J., & Yonelinas, A. (2016). Recollection, not familiarity, decreases in healthy ageing: Converging evidence from four estimation methods. *Memory*, 24(1), 75-88.
- Mark, R., Rugg, M., & Rugg, . (1998). Age effects on brain activity associated with episodic memory retrieval An electrophysiological study. *Brain*, 121, 861-873.
- Nilsson, L. (2003). Memory function in normal aging. *Acta Neurol Scand Suppl*, 179, 7-13.
- Old, S., & 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.
- Park, D., Smith, A., Lautenschlager, G., Earles, J., Frieske, D., Zwahr, M., & Gaines, C. (1996). Mediators of Long-Term Memory Performance Across the Life Span. *Psychology and Aging*, 11(4), 621-637.

- Park, H., Uncapher, M., & Rugg, M. (2008). Effects of study task on the neural correlates of source encoding. *Learning and Memory*, 15(6), 417-425.
- Pitman, E. (1939). A Note on Normal Correlation. *Biometrika*, 31(1/2), 9-12.
- Prull, M., Dawes, L., Mcleish, A., Iii, M., Rosenberg, H., & Light, L. (2006). Recollection and Familiarity in Recognition Memory: Adult Age Differences and Neuropsychological Test Correlates. *Psychology and aging*, 21(1), 107-118.
- Raven J, Raven JC, Courth JH. (2000). *Manual for Raven's progressive matrices and vocabulary Scales. Section 4: The Advanced Progressive Matrices*. San Antonio, TX: Harcourt Assessment.
- Reitan, R., & Wolfson, D. (1985). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (Vol. 4). Tucson, AZ: Neuropsychological Press.
- Smith A. (1982). *Symbol digit modalities test (SDMT) manual*. Los Angeles, CA: Western Psychological Services.
- Snodgrass, J., & Corwin, J. (1988). Pragmatics of Measuring Recognition Memory: Applications to Dementia and Amnesia. *Journal of Experimental Psychology: General*, 117(1), 34-50.
- Spencer, W., & Raz, N. (1995). Differential Effects of Aging on Memory for Content and Context: A Meta-Analysis. *Psychology and Aging*, 10(4), 527-539.
- Titov, N., & Knight, R. (1997). Adult Age Differences in Controlled and Automatic Memory Processing. *Psychology and Aging*, 12(4), 565-573.
- Toth, J., & Parks, C. (2006). Effects of age on estimated familiarity in the process dissociation procedure. *Memory & Cognition*, 34(3), 527-537.
- Wagenmakers, E., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., . . . Morey, R. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin and Review*, 25(1), 35-57.
- Wang, T. H., de Chastelaine, M., Minton, B., & Rugg, M. D. (2012). Effects of age on the neural correlates of familiarity as indexed by ERPs. *Journal of cognitive neuroscience*, 24(5), 1055-68.
- Wechsler D. (1981). *WAIS-R : Wechsler adult intelligence scale--revised*. New York, NY: Psychological Corp
- Wechsler D. (2001). *Wechsler test of adult reading*. San Antonio, TX: The Psychological Corporation.
- Wolk, D., Mancuso, L., Kliot, D., Arnold, S., & Dickerson, B. (2013). Familiarity-based memory as an early cognitive marker of preclinical and prodromal AD. *Neuropsychologia*, 51(6), 1094-1102.
- Yonelinas, A., & Jacoby, L. (1995). The Relation between Remembering and Knowing as Bases for Recognition: Effects of Size Congruency. *Journal of Memory and Language*, 34(5), 622-643.

