

# Age differences in the precision of memory at short and long delays

Stephen Rhodes<sup>1</sup>, Emily E. Abbene<sup>2</sup>, Ashley M. Meierhofer<sup>2</sup>, & Moshe Naveh-Benjamin<sup>2</sup>

<sup>1</sup> Rotman Research Institute, Toronto, Canada

<sup>2</sup> Department of Psychological Sciences, University of Missouri, USA

Accepted for publication at *Psychology & Aging*, July 2020.

## Abstract

Age differences are well established for many memory tasks assessing both short-term and long-term memory. However, how age differences in performance vary with increasing delay between study and test is less clear. Here we report two experiments in which participants studied a continuous sequence of object-location pairings. Test events were intermixed such that participants were asked to recall the precise location of an object following a variable delay. Older adults exhibit a greater degree of error (distance between studied and recalled locations) relative to younger adults at short (0-2 intervening events) and longer delays (10-25 intervening events). Mixture modeling of the distribution of recall error suggests that older adults do not fail to recall information at a significantly higher rate than younger adults. Instead, what they do recall appears to be less precise. Follow up analyses demonstrate that this age difference emerges following only one or two intervening events between study and test. These findings are consistent with the suggestion that aging does not greatly impair recall from the focus of attention but age differences emerge once information is displaced from this highly accessible state. Further, we suggest that age differences in the precision of memory, but not the probability of successful recall, may be due to the use of more gist-like representations in this task. (216 words)

*Keywords:* Short-Term Memory; Long-Term Memory; Precision; Aging; Mixture Model

Word count: 9500 (approx.)

Age differences in performance are seen across measures of short-term/working memory (for meta-analyses, see Bopp & Verhaeghen, 2005, 2020; Jaroslawska & Rhodes, 2019) and long-term/episodic memory (for meta-analyses, see Old & Naveh-Benjamin, 2008; Rhodes, Greene, & Naveh-Benjamin, 2019; Spencer & Raz, 1995; Verhaeghen, Marcoen, & Goossens,

1993). However, largely distinct literatures have developed around short- and long-term memory, each with their own procedures and tasks. This makes it difficult to assess to what extent age-related decline in performance is common to different delay periods or whether different underlying processes are implicated at different delays. Further, many of the commonly used tasks permit only a coarse characterization of responses as either correct or incorrect, which provides limited insight into the source of age differences in memory (i.e., are older adults more likely to fail to recall *any* information or are their memories just less precise?). In the present work we report two experiments addressing two main objectives: (1) To assess the role of study-test delay in modulating the magnitude of age differences in the recall of object location and (2) to use a fine-grained measure of recall performance that allows the separation of putatively distinct sources of recall error, namely imprecise recall of the target location versus recall failure. As we describe below, addressing these issues is important for establishing to what extent age differences in short-term and long-term memory reflect common or distinct underlying processes.

### Age differences at different study-test delays

Our main motivating question concerns the role that the delay between study and test plays in modulating age differences in memory performance. While there is a long-standing, and ongoing, debate regarding the basis of short- and long-term memory storage, there is general agreement that different processes are engaged at different delays (Atkinson & Shiffrin, 1968; Cowan, 2019; Norris, 2017; Ruchkin, Grafman, Cameron, & Berndt, 2003). For instance, recently studied information is assumed to be held in a highly accessible state, either via a focus of attention (e.g., Cowan, 1988; Oberauer, 2002) or via specialized stores that temporarily buffer information (e.g., Baddeley & Hitch, 1974; Logie, 2011). When tasked with recall, information that is not in this highly accessible state is typically assumed to engage a search process that may or may not successfully reconstruct the relevant information (Raaijmakers & Shiffrin, 2002; Unsworth & Engle, 2007). As such, it may be that this non-privileged information is more susceptible to interference from other encoded memories that may overlap to some degree with the target information (Cowan, Johnson, & Saults, 2005; Oberauer, Awh, & Sutterer, 2017).

Age differences in performance over the short- and long-term may be attributable to a common limitation in the encoding of information. For instance, it has been suggested that an age-related reduction in the capacity of working memory may limit older adults' ability to encode information and, therefore, also limit subsequent long-term memory performance (Craik & Byrd, 1982; Hara & Naveh-Benjamin, 2015). Alternatively, age differences in

---

Emily E. Abbene and Ashley M. Meierhofer contributed equally. Much of this work was completed while the first author was at the University of Missouri supported by Economic and Social Research Council grant ES/N010728/1 (to R.H Logie et al). Part of this work was presented at the 2018 Cognitive Aging Conference in Atlanta and at the 2018 meeting of the Psychonomic Society in New Orleans. Materials, data, and analysis code available at: <https://github.com/stephenrho/delay-precision>.

Correspondence concerning this article should be addressed to Stephen Rhodes, Rotman Research Institute, Baycrest, 3560 Bathurst Street, Toronto, ON M6A 2E1. E-mail: [srhodes@research.baycrest.org](mailto:srhodes@research.baycrest.org)

performance at different delays may reflect the slower encoding of information (Bartsch, Loaiza, & Oberauer, 2019), poor binding of component information (Naveh-Benjamin, 2000), or the formation of less distinctive representations owing to less efficient neural gain (S.-C. Li, Lindenberger, & Frensch, 2000; S.-C. Li, Naveh-Benjamin, & Lindenberger, 2005). This common limiting factor, namely the amount or quality of information encoded, would result in an age difference that does not greatly depend on the delay between study and test.

However, there are reasons to suspect that the difference between younger and older adults may increase with delay. Specifically, older adults have been found to exhibit specific difficulty with retrieval of information from long-term memory (Danckert & Craik, 2013; Healey & Kahana, 2016; Rhodes et al., 2019; Wahlheim & Huff, 2015). In addition, increasing the delay between study and test increases the opportunity for interference from intervening events. Older adults may be particularly susceptible to this interference (Hasher & Zacks, 1988; Hedden & Park, 2003), which would result in a larger age difference with increasing study-test delay. Thus, depending on what aspects of memory we theorize to be most impacted by age, there are different expectations as to how age should interact with study-test delay.

In one paradigm that can be used to probe memory following different study-test delays, participants are presented with a continuous stream of stimuli in which study and test events are intermixed (often referred to as “continuous recall” or “continuous recognition”). For example, in one of the first reports using this task with groups of different ages, Wickelgren (1975) presented a stream of high frequency words and participants had to respond “old” to repeated words (and “new” otherwise). The amount of time between repeated words was varied between 2 minutes and 2 hours and Wickelgren found that age differences in recognition sensitivity did not increase over time (see also, Chen & Naveh-Benjamin, 2012; Ferris, Crook, Clark, McCarthy, & Rae, 1980). Poon and Fozard (1980) used a similar task, but overall shorter study-test intervals, including tests of short-term memory. They found that group differences in recognition performance increased in the first 12 seconds following study and stayed relatively constant up to a delay of around 3 minutes (see also, Le Breck & Baron, 1987).

The majority of studies we could find in the aging literature have used continuous recognition, but Maddox, Balota, Coane, and Duchek (2011) report relevant data from a cued recall task (see also Balota, Duchek, & Paullin, 1989). In their experiments they were primarily interested in the role of spacing between tests of memory for paired associates. Thus, in their task, participants studied a pair of words and were asked to immediately recall one of the words in response to the other as a cue. Following this immediate recall attempt, there could be a delay or additional retrieval attempts. For older adults, performance declined considerably if there was a gap of two or more events between the initial recall attempt and subsequent attempts. For younger adults, a similar drop in performance was not seen until there were five intervening events, suggesting that older adults lose information more quickly over the short-term.

Thus, the work reviewed here suggests that age differences in memory performance are present at fairly short study-test delays (on the order of seconds) and then remain relatively constant at longer delays (on the order of minutes, hours). However, this should be taken

very tentatively as there are studies that find little change in age differences over short delays (Chen & Naveh-Benjamin, 2012) and increasing age differences at longer delays (Le Breck & Baron, 1987; Peterson, Schmidt, & Naveh-Benjamin, 2017). Further, as the majority of studies have looked at recognition, the literature may be missing specific age-related limitations in recall from long-term memory. Finally, these previous reports have mostly used binary measures of performance, where a response is either correct or incorrect. This provides fairly limited insight into the source of age differences in performance; specifically, do older adults make more errors due to a greater incidence of complete memory failure or are there age differences in the degree to which memory is corrupted by noise, leading to less precise representations? Here we adopt a task that better allows us to identify the source of age differences in performance.

### **Distinguishing the success and (im)precision of recall**

In the task used in the present experiments, participants are asked to recall the location of a previously studied object following a variable study-test delay. Previous work has clearly established an age-related deficit in object-location memory in both recognition (Chalfonte & Johnson, 1996; Mitchell, Johnson, Raye, Mather, & D’Esposito, 2000; Naveh-Benjamin, 1988; Noack, Lövdén, & Lindenberger, 2012) and recall (Hampstead, Towler, Stringer, & Sathian, 2018; Korkki, Richter, Jeyarathnarajah, & Simons, 2020; Naveh-Benjamin, 1987; Nilakantan, Bridge, VanHaerents, & Voss, 2018; Sharps & Gollin, 1988). Given this established age difference in object-location memory we wanted to examine whether study-test delay plays any role in modulating this difference. In focusing on object-location memory there is a possibility that the findings will not generalize to other stimuli. However, older adults’ difficulty in retaining associations in memory has been demonstrated across a range of materials (e.g., Bastin & Van Der Linden, 2005; Chen & Naveh-Benjamin, 2012; Naveh-Benjamin, 2000) and the meta-analysis of Old and Naveh-Benjamin (2008) found that age-related differences in object-location were similar in magnitude to other kinds of associations.

More importantly this paradigm allowed us to acknowledge the fact that memory retrieval is not an all-or-none process (Harlow & Donaldson, 2013; Harlow & Yonelinas, 2016; Mickes, Johnson, & Wixted, 2010). When provided with a particular cue it is possible that the strength of the association between the cue and response is not sufficient to result in recall (an assumption shared by many models of memory, see Raaijmakers & Shiffrin, 2002). Indeed, conscious recollection of past events has long been hypothesized to be a threshold process that can fail completely (e.g., Yonelinas, 2002). However, when cue elicited retrieval succeeds, what is retrieved is unlikely to be a perfect read-out of the initial event. The initial encoding and storage of information may lack some detail, associations can be distorted by other study events, and/or the reconstruction of the target at cued recall may be imperfect. Thus, the precision with which a particular location is recalled in response to an object cue will vary.

Tasks like the one used in the present work can allow us to distinguish failure of retrieval from imprecise retrieval of the target location. In particular we can use a *mixture modeling* approach to do so (see, e.g., Biderman, Luria, Teodorescu, Hajaj, & Goshen-Gottstein, 2019;

Brady, Konkle, Gill, Oliva, & Alvarez, 2013; Richter, Cooper, Bays, & Simons, 2016; W. Zhang & Luck, 2008). The model we use here makes two simple assumptions: (1) When a cue object is presented it either elicits successful retrieval of a remembered location or it does not.<sup>1</sup> (2) If retrieval is successful, the remembered location is an imprecise estimate of the presented location and, over a number of trials, these recalled locations would form a (bivariate) distribution centered on their target locations. Thus, the distribution of recalled locations is assumed to reflect a *mixture* of successful recall from memory, with some imprecision, and failures of recall, resulting in guesses. Thus, in addition to looking at recall error, we can use a mixture model to estimate the probability of recall from memory (which we refer to as  $m$ ) and the precision of recall around the target location ( $\sigma$ ) as a function of the delay between study and test.

Much of the previous application of continuous recall tasks and mixture modeling to groups of younger and older adults has been done in the context of short-term memory. In these studies older adults' recall is less precise than that of younger adults (Souza, 2016) and older adults also show a tendency to make "mis-binding" errors (Peich, Husain, & Bays, 2013; Pertzov, Heider, Liang, & Husain, 2015). For example, Pertzov et al. (2015) presented difficult-to-name fractals in different locations and found that older adults were more likely to erroneously locate a cued fractal near the location of another studied fractal.

Korkki et al. (2020, Experiment 3) present relevant data from tasks involving different study-test delays. They had younger and older participants recall the color of previously studied objects in separate short- and long-term memory tasks. In addition, their participants completed a perceptual task where they matched the color of an object currently visible on screen, providing a measure of response imprecision arising from perceptual or motor limitations, rather than memory. In each task color was reported on a circular scale allowing Korkki et al. (2020) to apply the circular mixture models developed to analyze error data in working memory tasks (e.g., Bays, Catalao, & Husain, 2009; W. Zhang & Luck, 2008).

For both groups responses in the perceptual task were very accurate and, interestingly, in both the short- and long-term tasks there were clear age differences in the estimated precision of recall but no clear differences in the probability of successful recall from memory (see also, Nilakantan et al., 2018). More importantly for our purposes, Korkki et al. (2020) examined to what extent age differences in long-term memory performance could be explained by differences in performance on the perceptual and short-term tasks. They found that, for older adults, performance in the short-term task was a predictor of performance over the long-term (accounting for around 35% of variance), perhaps suggesting that limitations in short-term memory precision may, to some extent, determine the fidelity of recall at later delays. However, age differences in long-term memory still remained when including short-term memory performance as a covariate. Performance in the perceptual task did not predict long-term performance for either age group.

These findings may be taken to suggest that there are age differences in processes that operate over the long-term in addition to those limiting performance over the short-term. Nevertheless, there are still important differences between the short- and long-term tasks

<sup>1</sup>Note that the model cannot distinguish failure to recognize the object cue itself from failure to recall the location associated with a particular recognizable object. Experiment 2 aims to address this.

used by Korkki et al. (2020) that limit direct comparison. Specifically, in the short-term task the 3-object arrays were presented for 3 seconds and followed by a pattern mask (to eliminate lingering sensory information), whereas arrays were presented for 3 times this duration in the long-term task and were not masked. Further, 15 colors were recalled during each test of long-term memory, resulting in a much greater chance of output interference (see Biderman et al., 2019) relative to the single test following each short-term trial.

Additionally, as these tasks were blocked, it is possible that participants approach the task differently when given different amounts of time at study or when they know that a particular array must be memorized for a longer or shorter period of time. Relevant data on this comes from Hinrichs and Grunke (1975), who used a continuous cued recall task and provided participants with information on the retention interval for particular pairs at study. Providing information on the length of the retention interval benefited performance over short intervals (2-5 intervening events, around 4-16 s) and had little effect on performance at longer delays. This suggests that participants can employ different encoding or maintenance strategies when the retention interval is known. While age differences in strategic approach across different memory tasks is an interesting question in itself (see, e.g., Bailey, Dunlosky, & Hertzog, 2009; Naveh-Benjamin, Brav, & Levy, 2007), we wanted to rule out differences due to strategic approach when assessing the role of study-test delay.

## Experiment 1

This experiment addressed two main goals. The first was to assess how age differences in memory for object-location vary between study-test delays spanning a few intervening events (a matter of seconds) to many intervening events (a couple of minutes). The delays were chosen to cover the range of study-test intervals often used in studies of short-term/working memory (e.g., Peich et al., 2013; Pertzov et al., 2015) and long-term/episodic memory (e.g., Brady et al., 2013; Korkki et al., 2020). Further, previous work, reviewed above, suggests that this range of delays is most likely to reveal increasing age differences (Le Breck & Baron, 1987; Maddox et al., 2011; Poon & Fozard, 1980). The second goal was to further explore the sources of age differences in object location recall through mixture modeling. Specifically, we wanted to distinguish responses that reflect imprecise memory for object location from complete failures of recall location in response to the object cue.

To this end we used a continuous cued recall task (adapted from Lew, Pashler, & Vul, 2016) in which participants studied objects sequentially in different locations and test events, where a particular object had to be relocated, were intermixed following a variable delay. To allow for a sufficient number of observations at different delays between study and test, we follow the general procedure of Chen and Naveh-Benjamin (2012) and group lags into “*short*”, “*medium*”, and “*long*” delays. To our knowledge, this is the first study to address age differences in the precision of recall following different study-test delays using a paradigm where participants are unaware of how long it will be until a particular item is probed. This is important as it ensures that differences between tests of memory at different delays cannot reflect differences at encoding (e.g., due to different strategic approach, see above).

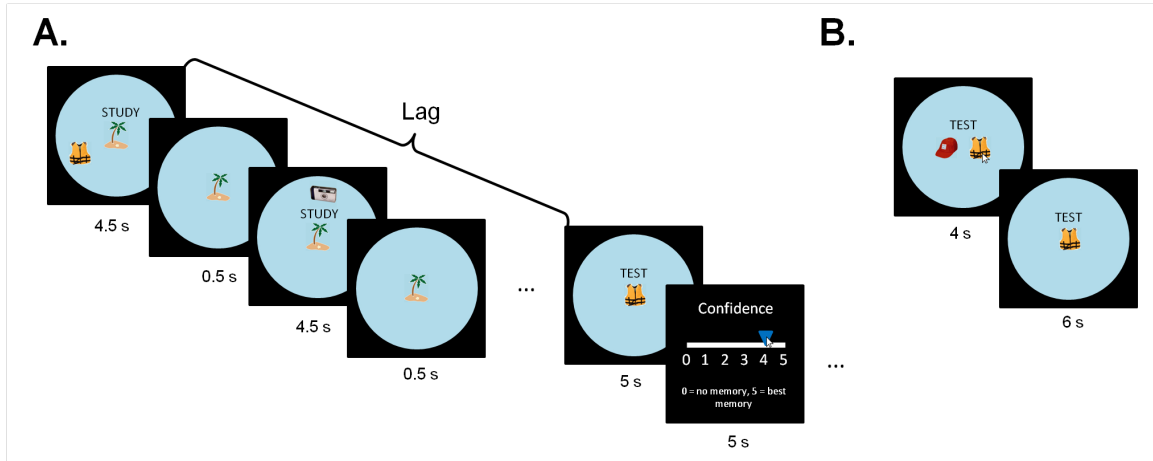
If age differences in recall following different study-test delays reflect a common limitation in working memory capacity ( Craik & Byrd, 1982; Hara & Naveh-Benjamin, 2015) or the quality of information encoded within the time given (Bartsch et al., 2019; S.-C. Li et al., 2000, 2005) we would expect a group difference in recall error following the shortest delay that does not increase with greater time between study and test. However, if additional age-related factors come into play over longer delay periods, such as a greater susceptibility to retroactive interference from intervening events (Hasher & Zacks, 1988; Hedden & Park, 2003) or difficulty in retrieval from long-term memory (Danckert & Craik, 2013; Healey & Kahana, 2016; Rhodes et al., 2019; Wahlheim & Huff, 2015), we should expect age differences to increase across the three study-test delays.

## Method

**Participants.** Thirty-three undergraduates (age range: 18-27 years,  $M=19.61$ ,  $SD=1.78$ ), enrolled in an introductory psychology class at the University of Missouri, and 32 older adults (age range: 65-80 years,  $M=73.59$ ,  $SD=3.96$ ) from the surrounding area took part in this experiment. One additional older adult was excluded for having more than 25% missing responses on the location recall task (criteria described below). All participants reported being in good physical health and the older group had completed significantly more years of education ( $M=14.72$ ,  $SD=2.07$ ) relative to the younger group ( $M=13.30$ ,  $SD=0.95$ ; Welch's  $t(43.28) = 3.53$ ,  $p < 0.05$ ). Participants completed a battery of neuropsychological tests (following Glisky, Polster, & Routhieaux, 1995; Peterson, Gargya, Kopeikin, & Naveh-Benjamin, 2017), the results of which are not discussed here (see supplement section 1 for group means).

Younger adults were given course credit for their participation, whereas the older adults were paid \$15. These experiments were conducted with approval from the University of Missouri's Institutional Review Board. Our inferential approach, which includes Bayesian estimation and reporting of uncertainty in parameter estimates, does not require the prespecification of a particular sampling approach (see Wagenmakers, 2007). Therefore, our sample size was mainly determined by time constraints, as data collection primarily took place during spring semester. However, we aimed for at least 30 participants in each group, which would result in a larger sample than other recent experiments that have found age-differences in fine-grained recall ( $N$ s of 20-24 per group; Korkki et al., 2020; Nilakantan et al., 2018).

**Stimuli.** Stimuli were presented on a LCD monitor with a resolution of  $1920 \times 1080$ . We used a similar cover story to that in Lew et al. (2016); participants were asked to imagine sailing around an island and were told that belongings had fallen overboard. The participant's task was to remember the locations of these lost belongings. Thus the images of objects were presented within a light blue ( $RGB = [173, 216, 230]$ ) circular area with a radius of 500 pixels (see Figure 1). Images of objects were taken from the set used in Brady, Konkle, Alvarez, and Oliva (2008) (available at <https://bradylab.ucsd.edu/stimuli.html>). For this experiment we selected 117 distinct objects ( $100 \times 100$  pixels). An image of an island ( $87 \times 100$  pixels) was presented throughout the experiment (except for during confidence ratings) to serve as an anchoring point (see Figure 1A). During the task the prompts "STUDY" and



*Figure 1.* (A.) Schematic of the general trial procedure used in the present experiments. Participants studied objects in different locations with recall probes intermixed. Lag is the number of events between study and test. In Experiment 1, at test participants saw an object in the center of the screen and used the mouse to move it to the remembered location before providing a confidence rating. (B.) In Experiment 2, participants first made a two-alternative forced choice between a previously studied object and a new lure before relocating the selected object. There was no confidence rating in this experiment. Note: items are not drawn to scale.

“TEST” were provided 60 pixels above the center of the screen with a text height of 20 pixels (see Figure 1).

**Procedure.** Figure 1A depicts the continuous recall task used in this experiment. During study events, participants were instructed to remember the location of the presented object. The object was presented for 4.5 seconds, followed by a 0.5 second inter-stimulus interval where only the central island was presented. During test, an image of a previously studied object was presented in the center of the screen and the participant was given 5 seconds to relocate the object by moving the mouse and clicking on the chosen location. Following this they were asked to rate their confidence on a scale from 0 (no memory at all) to 5 (best possible memory) (following Rademaker, Tredway, & Tong, 2012). Participants were also given 5 seconds to rate their confidence. These time restrictions on responding were to ensure that each “event” lasted a fixed time (5 s) with the test phase constituting two events (i.e., relocation and confidence).

Memory for a particular object was probed following one of nine different lags, which in turn were grouped into “short”, “medium”, and “long” delays (following Chen & Naveh-Benjamin, 2012).<sup>2</sup> The short delay condition consisted of 0, 1, or 2 events between study and test, where 0 events mean that the object was probed immediately after the inter-stimulus interval. The medium delay condition was lags 10, 11, and 12 (50, 55, and 60 seconds), and the long delay condition was lags 23, 24, and 25 (115, 120, 125 seconds).

<sup>2</sup>Note, our delay intervals correspond to the “short-term”, “short-long-term”, and “mid-long-term” intervals used in Chen and Naveh-Benjamin (2012).



Events were presented in two blocks to allow for a break in the middle of the session. Each block included 162 events with 54 relocation and confidence rating responses per block. This provided 108 responses total with 12 observations at each lag and, thus, 36 observations at the short, medium, and long delays per participant. To create the distribution of lags for each individual participant we prepared a set of nine lists with different orderings of study-test events at each lag. Each list contained 54 events divided equally across the nine lags and six of these lists were randomly sampled for each participant and concatenated into the two lists of 162 events. This ensured that the different lags were distributed across the entirety of each block. Prior to the main experiment, participants were given a practice block containing 27 events (9 study events). Participants were able to repeat this practice phase until they were comfortable with the task.

With the exception of the 9 images used in the practice phase, images were randomly assigned to study-test lags for each participant. The locations associated with each object were sampled pseudorandomly such that a sampled location could not be within 50 pixels of a previously studied location that had not yet been tested. Although the study area had a radius of 500 pixels, the distance from center of a sampled location could not exceed 429 pixels, so that the image did not “escape” the study area. Similarly, so that the images did not overlap with the central fixation image or instruction text, sampled locations had to be more than 141 pixels from the center of the screen. The paradigm was implemented using PsychoPy2 (Peirce et al., 2019) and the materials can be found at: <https://github.com/stephenrho/delay-precision>.

### Analysis.

**Recall Error.** The distribution of recall error, the Euclidean distance between presented and recalled locations in pixels, across age-groups and conditions is presented in Figure 2. To model this positively skewed data distribution, with a lower bound of 0, we used a generalized linear mixed effects model with a log normal response distribution. The mean of this log normal distribution is modeled as a function of fixed effects of delay and age-group and their interaction. There are additional random effects of participant and item (cue image) to account for these potential important sources of variation. Throughout the article we use sum-to-zero coding where one condition of a given factor is chosen as the reference level (coded -1 in all contrasts) and the remaining levels each get a contrast (in which they are coded 1 and other non-reference levels are coded 0). The result of this coding scheme is that the intercept term reflects the grand mean predicted value (on the transformed scale) and the contrasts associated with a given factor reflect the deviation from the grand mean associated with a particular level or condition. For the factor of study-test delay, the medium condition was chosen as the reference level.

As is the case for all of the main analyses reported here, we estimated this model under a hierarchical Bayesian framework and these mixed effects models were implemented with the R (R Development Core Team 3.0.1., 2013) package, **brms** (Bürkner, 2017, 2018).<sup>3</sup> The priors we used in our analysis were intended to be *weakly* informative; that is, they placed most prior density on reasonable parameter values and were broad enough to let the data drive inference (see supplement section 2 for details). All results reported here

<sup>3</sup>We also used the **papaja** (Aust & Barth, 2018) and **plyr** (Wickham, 2011) packages.

are based on a total of 8000 posterior samples from 4 separate chains following 4000 warm up samples (1000 per chain). Convergence was assessed with the  $\hat{R}$  statistic described in Gelman et al. (2014, pp. 284–286), which was always  $< 1.1$ .

We use the output of our Bayesian analysis (samples from the posterior distribution over our model parameters) to construct specific contrasts (see Kruschke, 2015 for an introduction). As our model parameters are often transformed to allow linear modeling we project our parameters back to the “manifest” space when reporting contrasts. This ensures that we have appropriate measures on uncertainty (i.e., 95% credible intervals) for contrasts on bounded scales (for example, recall error). Thus, specific contrasts on interpretable scales are presented throughout this article to test key hypotheses and the supplement (section 3) presents full model results on the transformed scale (tables of coefficients).

**Mixture Model.** To further probe the nature of age differences in performance on this task we fit a mixture model to the recall data. The approach we take is similar to the approach described in the Introduction, where responses are assumed to come from (imprecise) memory for target item on some proportion of trials and from a uniform (guess) distribution otherwise (e.g., W. Zhang & Luck, 2008). This general model can be extended to two dimensional data, such as our location responses (see Lew et al., 2016; Schneegans & Bays, 2016), by assuming that on some proportion of trials,  $m$ , the recalled  $(x, y)$  coordinates come from a bivariate normal distribution centered on the correct location with some imprecision,  $\sigma$ . On the remaining  $1 - m$  proportion of trials the observer is assumed to choose a location at random on the allowed area.

More concretely, the recalled coordinates ( $\mathbf{r}$ ) for individual  $i$ , in delay condition  $j$ , on trial  $k$  were assumed to be distributed as follows:

$$\mathbf{r}_{ijk} \sim m_{ij} \times \phi(\mathbf{p}_{ijk}, \sigma_{ij}^2) + (1 - m_{ij}) \frac{1}{A}$$

where  $\phi$  is a bivariate normal distribution centered on the presented location,  $\mathbf{p}$ , with standard deviation,  $\sigma$ , and  $A$  is the size of the area in which items could appear. The individual  $m_{ij}$  and  $\sigma_{ij}$  parameters were estimated with a hierarchical model, in which they were assumed to be drawn from separate multivariate normal distributions with different population level means for the three delay conditions (short, medium, long). As these parameters are bounded ( $m \in [0, 1]$  and  $\sigma \in [0, \infty)$ ) these distributions were placed on the logit transformation of  $m$  and the log transformation of  $\sigma$ . We sampled from the posterior distribution of this model, given the data, using **Stan** (Carpenter et al., 2017) via the **rstan** package (Stan Development Team, 2018) and the model was fit separately for the two age-groups. Our inference focuses on comparing the population level parameters between groups, transformed back to the more interpretable manifest scale.

## Results

As mentioned above, one older participant was excluded from further analysis for having less than 75% usable observations. In the remaining sample 4.25% of observations were lost; 3.62% were lost due to failure to make a relocation response within the 5 seconds

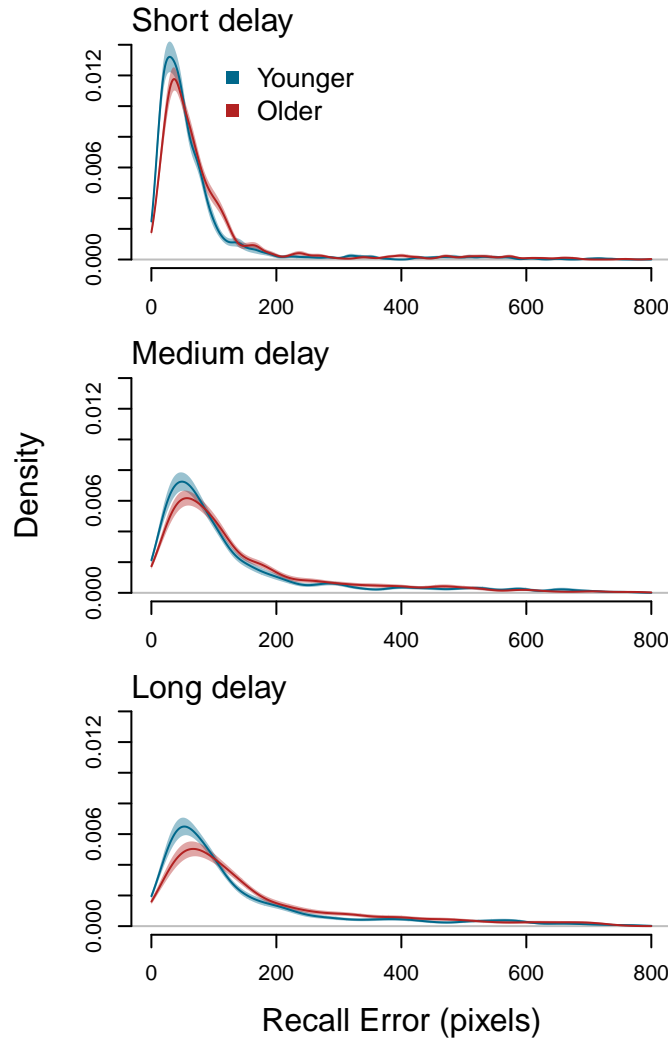


Figure 2. Distribution of recall error (kernel density estimate) by delay and age-group for Experiment 1. Shaded area represent bootstrapped standard errors (1000 replications, resampling both participants and trials).

given, and a further 0.63% were lost for clicking outside the allowed area. For the analysis of recall error we included trials where a confidence rating was not given within 5 seconds. In the analyses relating to confidence, a further 1.08% of observations were lost due to missing ratings.

**Recall error.** Figure 2 presents the distribution of recall error across age groups and delays and Figure 3 displays population level estimates from the linear mixed effects model (expected recall error). Of greatest interest to us is the effect of age and its interaction with delay. There was a main effect of age, with greater expected recall error for older adults (129.870 [89.845, 162.169] 95% credible interval) relative to younger adults (109.271 [73.489, 135.282]), with a difference of -20.599 [-26.887, -16.357] pixels. As shown in Figure 3, the magnitude of the age difference was fairly similar in the short (-16.260 [-30.639, -2.300])

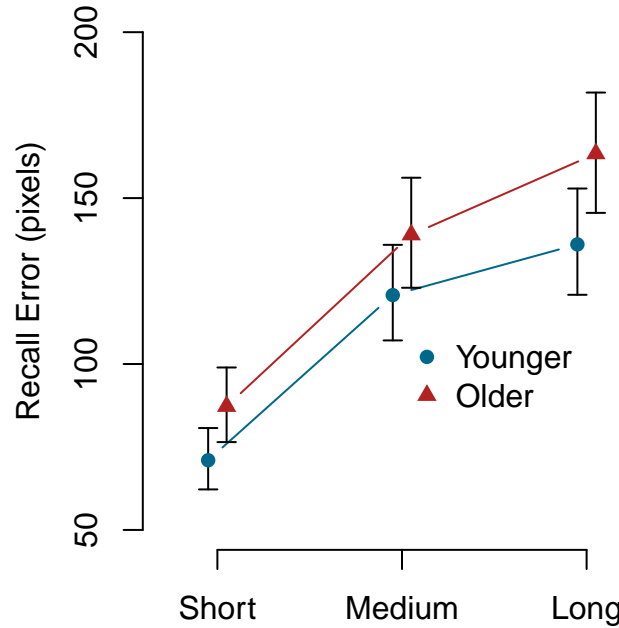


Figure 3. Experiment 1 recall error posterior means and 95% credible intervals by delay and age-group.

and medium conditions (-18.192 [-39.968, 3.154]) and increased somewhat in the long delay condition (-27.345 [-50.775, -3.811]). However, when we directly contrast age differences across the different delays, there is no convincing evidence that the group difference at the long delay is credibly different from that at the short (11.085 [-3.780, 25.568]) or medium delay (9.153 [-4.804, 23.117]).

**Mixture model.** Estimates of the probability of recall from memory,  $m$ , and the imprecision of memory recall,  $\sigma$ , are presented in Figure 4. First we consider the probability of recall from memory; averaging across delay condition,  $m$  for younger adults (0.819 [0.760, 0.869]) was only slightly larger than  $m$  for older adults (0.777 [0.715, 0.831]) and the difference was not credibly different from zero (0.042 [-0.040, 0.122]). As shown in Figure 4A the two age groups did not credibly differ at any of the delays (short: 0.025 [-0.021, 0.072], medium: 0.038 [-0.067, 0.142], long: 0.063 [-0.056, 0.178]). To assess whether age differences vary as a function of delay we can contrast the different delays directly. Age differences in the probability of successful recall did not significantly increase from short to medium delay (-0.013 [-0.098, 0.076]) or from short to long delay (-0.037 [-0.130, 0.055]).

For the imprecision of memory (Figure 4B) there is an overall age difference with younger adults having smaller estimates of  $\sigma$  (51.350 [48.152, 54.751]), implying more precise recall, relative to older adults (58.665 [55.638, 61.893]); this difference is credibly non-zero (-7.315 [-11.900, -2.759]). The magnitude of this age difference was consistent across the three delays (short: -6.692 [-10.693, -2.665], medium: -6.586 [-12.902, -0.348], long: -8.667 [-15.217, -2.124]). Further, age differences in recall imprecision did not credibly increase from short to medium delay (-0.106 [-5.750, 5.662]) or from short to long delay (1.975 [-3.680, 7.641]).

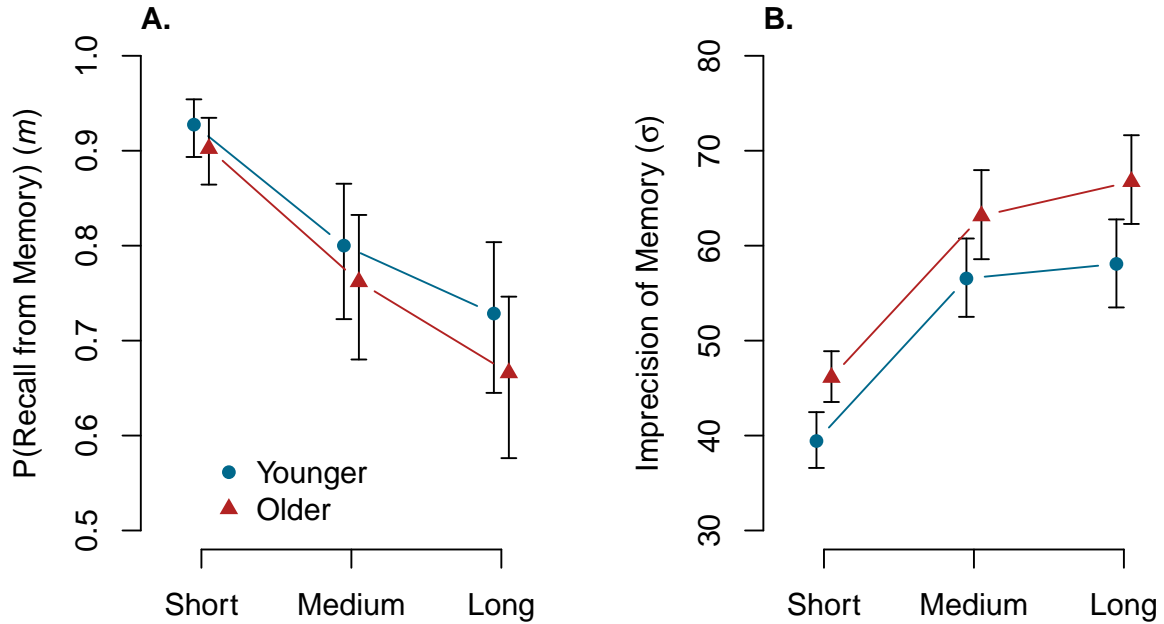


Figure 4. Experiment 1 mixture model estimates of (A.) the probability of successful recall and (B.) the imprecision of recall. Posterior means and 95% credible intervals.

As we mentioned in the Introduction, an additional potential source of error may be that participants erroneously recall the location associated with another studied object that is not the cue (see Bays et al., 2009; Lew et al., 2016; Pertsov et al., 2015). Given the continuous presentation of stimuli in this task, there are numerous potential non-target locations for mis-binding responses. However, as temporal order is a strong cue for recall (see, e.g., Polyn, Norman, & Kahana, 2009) it seems likely that mis-bindings would occur between pairs studied in close temporal proximity. Further, there is evidence that older adults are more likely to confuse proximal pairings (Campbell, Trelle, & Hasher, 2014). To address this we assessed the Euclidean distance between a particular recall response and the location studied before or after the cue presented. The resulting distributions of error suggest no tendency for younger or older adults to make mis-bindings in all delay conditions (see supplement, section 6).

**Confidence.** The analysis of confidence ratings and their relationship with both recall error and the mixture model parameters is presented in section 4 of the supplement. To summarize, confidence ratings drop from an average of 4.426 [4.298, 4.540] (recall, the scale goes from 0-5) at the short delay to 3.630 [3.435, 3.816] at the medium delay (short vs. medium: 0.796 [0.703, 0.891]). There is then a smaller drop in confidence going to the long delay 3.508 [3.304, 3.701] (medium vs. long: 0.122 [0.049, 0.194]). This change in confidence with delay is seen across both age groups. Further, confidence exhibits a clear relationship with recall error, probability of successful recall, and memory imprecision across both age groups.

## Discussion

Older adults' recall of spatial location was less accurate than that of younger adults and this appears to be primarily driven by less precise memory for the target location. There was no clear evidence of a difference in the rate of complete memory failure between the age groups. Thus, consistent with other recent reports (Korkki et al., 2020; Nilakantan et al., 2018), the rate of successful retrieval does not differ much with age but what is retrieved by older adults is somewhat less fine-grained. Both groups also appeared to have insight into the information available in memory, as confidence ratings were clearly associated with the degree of recall error as well as the parameters of the mixture model (see supplement, section 4).

While study-test delay had a very large effect on overall performance, it did little to modulate differences between the two age groups. This could implicate perceptual limitations or age differences in producing a fine grained motor response (i.e., using the mouse) rather than age differences in memory. However, the findings of Korkki et al. (2020) suggest that perceptual or motor limitations do not explain age differences in performance. Further supporting this, below we combine data from both experiments reported here and find that there are no clear age differences at lag 0 (i.e., immediate test).

Both groups showed a marked increase in memory imprecision with delay as well as a clear increase in the probability of recall failure (see Figure 4). Two obvious questions can be asked about the nature of this loss with delay; (1) Is change in performance with increasing study-test delay primarily driven by the amount of time that passes or the number of intervening events that may interfere with memory? (2) Where the mixture model analysis suggests that participants are guessing, what aspect of the stimulus have they failed to recall? They could forget the location itself (or its association with the cue object) or they could forget that they had seen the cue object at all during the session. These questions provided the motivation for Experiment 2.

## Experiment 2

To begin to address whether the decline in performance from short to long delay is primarily driven by the amount of time between study and test or the number of intervening events, in Experiment 2 we included a block of trials in which study and test were further spaced out by including a 2 second blank interval between them. Thus, in this block the amount of time between study and test for the long delay condition was, on average, 152 seconds (32 seconds longer than in the no interval condition). If time is a crucial determinant of the amount of information lost from memory in this task we would expect an interaction between the delay factor and this interval, such that the effect of delay is larger with an interval. On the other hand, if it is primarily the number of intervening events between study and test that cause the decline in performance, we would expect little difference between these two interval conditions.

As discussed above, Experiment 1 has indicated quite a large increase in the rate of retrieval failure, for both age groups, with increasing study-test delay. Given that our task

is cued recall, it is possible that part of this increase in the guessing rate is due to loss of memory for the cue object itself. To assess retention of the cue object we modified the test phase in Experiment 2 (see Figure 1B). Before relocating the object to its remembered location, we asked participants to distinguish a previously studied object from an unstudied lure (see Pertzov et al., 2015). If participants are failing to recall the cue object with increasing delay, we should see a clear drop in accuracy on this task. On the other hand, if memory for the cue object is good enough to discriminate studied from unstudied objects we would expect a high level of performance in both age groups (see S. M. Stark, Yassa, Lacy, & Stark, 2013), indicating that the age-related differences in Experiment 1 are specifically in the precision of memory for spatial location.

## Method

**Participants.** We again aimed for at least 30 participants in each age group. A new sample of 34 younger (age range 18-22,  $M=18.53$ ,  $SD=1.02$ ) and 31 older adults (aged 66-84,  $M=73.32$ ,  $SD=5.47$ ) were recruited from the same sources to participate in Experiment 2 and included in the analysis. One additional older participant was excluded from analysis as more than 25% of their observations were missing (due to slow responding). Again, the older group had completed more years of education ( $M=14.81$ ,  $SD=2.39$ ) relative to the younger group ( $M=12.32$ ,  $SD=0.88$ ; Welch's  $t(37.35) = 5.47$ ,  $p < 0.05$ ). Participants also completed the same battery of neuropsychological tests from Experiment 1 (see supplement, section 1).

**Stimuli and Procedure.** One change for Experiment 2 was to the test phase of the task. In this experiment, test began with a two-alternative forced choice task where two images were presented 75 pixels to the left and right of the center of the screen (see Figure 1B). One image had been studied previously, whereas the other image was an unstudied lure. The participant was given 4 seconds to select the studied image by clicking on it with the mouse. When an image was selected, the images disappeared and the program waited for the 4 seconds to pass. If a selection was not made in this time the message "PLEASE RESPOND FASTER" was presented in red at the center of the screen for 6 seconds. If a choice was made, the selected image was presented in the center of the screen and participants had 6 seconds to relocate the object to its remembered location. If a location was not selected in the time given the program moved on to the next event. The lags used were the same as Experiment 1, although now the lag refers to the number of events between study and the forced choice task (at which point the participant could also retrieve the location of a remembered object). The same 117 images were used with an additional 117 images, selected from the same database as Experiment 1, to serve as unstudied lures. The 216 non-practice images were randomly shuffled for each participant and could serve as either a studied or lure item in a particular session.

Another change was that, in one of the blocks of trials, study and test events were followed by a 2 second interval in which only the island fixation image was presented. This had the effect of spacing out events in time by about 26%. Thus, the average number of seconds between study and test in the 2 second interval condition was 6.33, 69.67, and 152 in the short, medium, and long delay conditions, respectively. In the no interval condition these values were 5, 55, and 120 seconds, respectively. If the amount of time between study

and test is crucial in driving the forgetting we saw in Experiment 1, then the rate of change in performance across the short, medium, and long delay conditions should be larger with the 2 second interval relative to no extra interval.

## Results

**Two-alternative forced choice accuracy.** Two-alternative forced choice accuracy was above 95% in all conditions for both groups (97.15% correct overall; 97.47% for young and 96.80% for older adults). This includes trials where the response period timed out before participants made a selection (scored as incorrect; 1.65% of observations). Trials on which the choice was incorrect were removed for the analysis of recall error.

In addition, a further 2.74% of observations were lost; 2.18% due to failure to recall within the 6 seconds given and 0.56% for clicking outside the study area. In all, 5.58% of observations were lost due to the reasons above, not including one older participant excluded for having over 25% missing data.

**Recall error.** Figure 5 presents the distribution of recall error across the delay and interval conditions. The same log normal linear mixed effects model from Experiment 1 was applied to this data, this time with additional fixed-effects associated with interval.

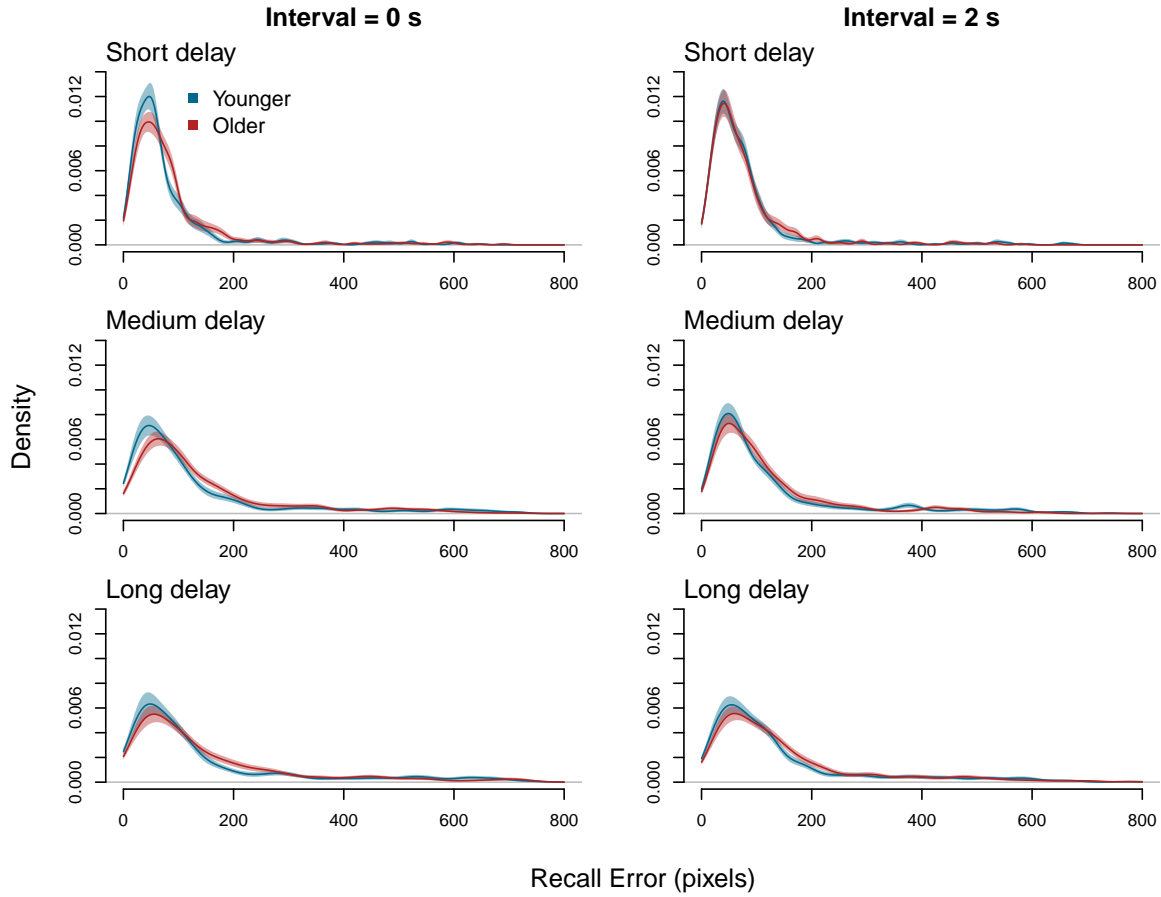
As in Experiment 1, there was a clear main effect of delay on recall error (see Figure 6). Expected recall error was lowest in the short delay condition (83.132 [76.009, 90.715]), followed by medium (122.261 [112.221, 132.593]), and recall error was greatest with a long delay (136.404 [125.238, 147.640]). The main effect of interval was not credibly different from zero. Estimated recall error was somewhat smaller when a 2 second interval was included following study and test (112.341 [103.160, 121.700]) relative to when no interval was inserted (115.524 [106.465, 125.004]), with a difference of 3.183 [-1.338, 7.850]. We return to this unexpected result in the discussion. Error was also numerically larger for older adults (119.333 [106.317, 133.553]) than for younger adults (108.532 [97.105, 120.929]), but the contrast for age was not credibly different from zero (-10.801 [-28.489, 7.285]).

Next, to assess the interaction of delay and interval, we contrast the two interval conditions at each delay. These contrasts are given such that positive values reflect larger error in the no interval condition and we see that the only credible difference is in the medium delay condition (11.715 [3.186, 20.120]). For the short (1.670 [-4.227, 7.558]) and long delay (-3.837 [-13.112, 5.207]) conditions the interval had no clear effect on recall error.

For the interaction of delay and age, in the no interval condition the posterior mean for the difference between age groups was consistent with that found in Experiment 1 (-15.646 [-34.376, 3.546]), although the 95% credible interval includes zero. On the other hand for the 2 second interval condition the difference between groups was slightly less pronounced (-5.956 [-24.092, 12.511]). Contrasting these two directly we find that the difference between groups was credibly smaller with an interval compared to without (-9.690 [-19.057, -0.678]).

However, as shown in Figure 6, this appears to be primarily driven by age differences in the medium condition. For younger adults there is no credible difference between the two interval conditions at any of the delays (short: -2.501 [-10.287, 5.380], medium: 1.809





*Figure 5.* Distribution of recall error (kernel density estimate) across the three delays and two interval conditions for younger and older adults in Experiment 2. Shaded area represents bootstrapped standard errors.

[-9.351, 12.869], long: -4.296 [-16.382, 7.628]) and, further, the delays themselves do not differ (short vs. medium: -4.310 [-18.039, 9.424], medium vs. long: 6.105 [-10.313, 23.125], short vs. long: 1.795 [-12.364, 16.137]). However, for the older group recall error was lower in the medium condition when there was a 2 second interval relative to no interval (21.620 [8.715, 34.607]). This was somewhat the case in the short delay condition, although the difference was not credibly different from zero (5.841 [-3.128, 14.785]), whereas at the long delay the two intervals did not differ (-3.378 [-17.428, 10.269]). The difference between the interval conditions appears to be specific to the medium delay condition when we compare the different delays directly (short vs. medium: -15.779 [-30.968, -0.487], medium vs. long: 24.998 [6.279, 44.244], short vs. long: 9.219 [-7.197, 25.808]). A final important test is to contrast the magnitude of difference between the no interval and 2 second interval conditions between the two age groups. For this contrast the two groups differed slightly, and not credibly, at the short delay (-8.342 [-20.212, 3.711]), whereas they clearly differ for the medium delay (-19.811 [-37.111, -2.859]). The difference between the interval conditions does not reliably vary across the two age groups at the long delay (-0.917 [-19.323, 17.306]).

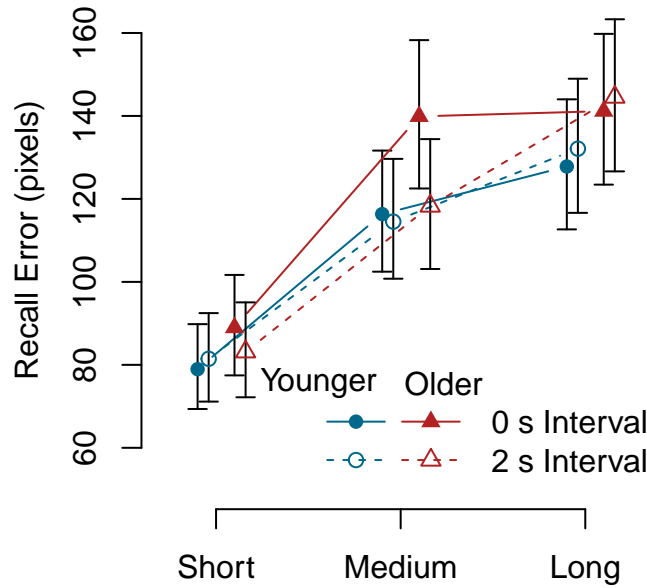


Figure 6. Experiment 2 recall error posterior means and 95% credible intervals by delay, interval, and age-group.

This aspect of the results was unexpected and there is no clear reason, we can think of, that separating events by a further 2 seconds would improve recall error *specifically* at a medium delay. To further assess the evidence for an interaction between age, delay, and interval we fit another model omitting this three-way interaction. Comparing these two models via the approximate leave-one-out cross validation (LOO) information criterion, which essentially penalizes the predictive utility of a model according to its flexibility (see Vehtari, Gelman, & Gabry, 2017), we find that including the three-way interaction does not improve predictive power (LOO full model: 13,322.34, LOO model without three-way interaction: 13,320.26, difference: 2.07, SE = 3.06).

**Mixture model.** The above analysis of the distribution of recall error suggests no consistent effect of interval on performance. Consequently, we fit the mixture model allowing the parameters to vary by delay only, as in Experiment 1. As there was some suggestion that including an interval may have improved the performance of older adults, we report a follow up analysis including the effect of interval below.

The probability of successful recall from memory ( $m$ ) is plotted in Figure 7A. As shown in this figure, there was very little age difference here. For younger adults, overall  $m$  was estimated as 0.841 [0.780, 0.893] and for older adults it was 0.826 [0.773, 0.874] (difference: 0.015 [-0.062, 0.089]). The difference between younger and older adults was not significant at all delays (short: 0.007 [-0.037, 0.050]; medium: 0.006 [-0.092, 0.103]; long: 0.032 [-0.078, 0.137]).

Figure 7B presents the imprecision of memory recall ( $\sigma$ ) by delay and group. As in Experiment 1, memory imprecision was lower for the younger group, 53.610 [50.793, 56.650], relative to the older group 60.886 [56.604, 65.551] (difference: -7.276 [-12.638, -2.070]). We next considered the age difference in  $\sigma$  as a function of delay. The 95% credible interval

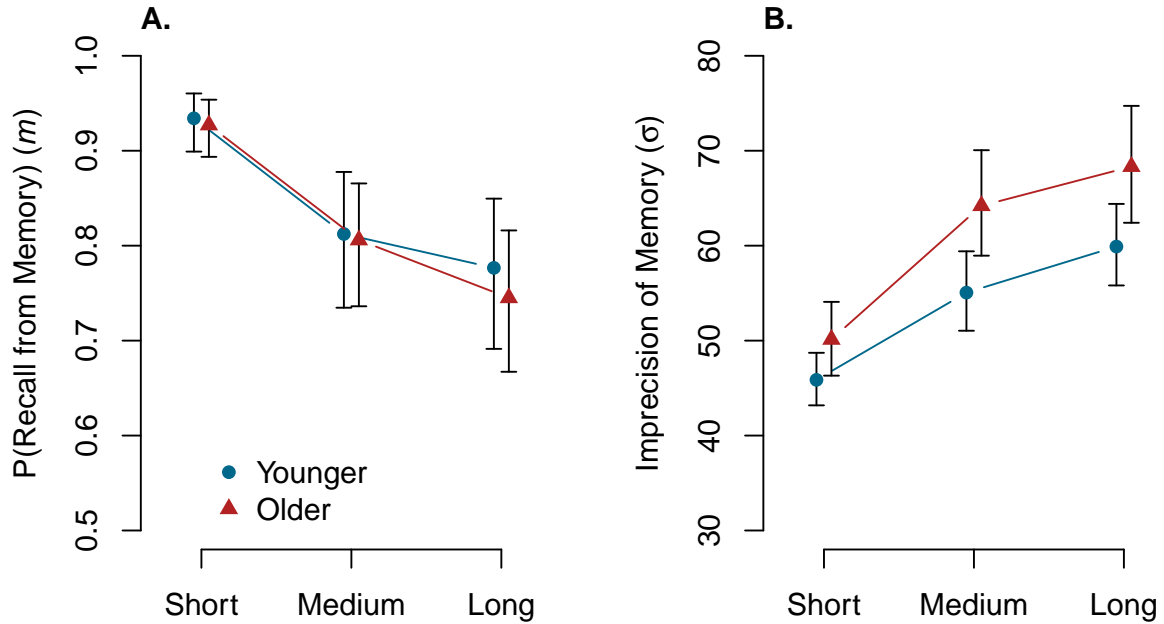


Figure 7. Experiment 2 mixture model estimates of (A.) the probability of successful recall and (B.) the imprecision of recall. Posterior means and 95% credible intervals.

overlapped with zero for the short condition ( $-4.260$  [ $-9.115, 0.391$ ]) but not the medium ( $-9.149$  [ $-16.307, -2.327$ ]) and long ( $-8.420$  [ $-15.981, -1.045$ ]) conditions. However, directly comparing the short delay condition to the others (vs medium:  $4.889$  [ $-0.786, 10.482$ ]; vs long:  $4.160$  [ $-1.821, 10.281$ ]) we do not have sufficient evidence to conclude that the age difference is smaller at the shortest delay relative to the longer delays. We return to this issue below when we combine the data from our two experiments.

The analysis of recall error above suggested that interval interacted with age such that overall age differences were less pronounced with a 2 second interval following study and test relative to no interval (although in both cases the age difference was not credibly different from zero). Thus, to further probe this possible interaction with age we fit an additional version of the mixture model with the added main effect of interval.

For the younger group the difference in parameter estimates between the two interval conditions was not different from zero ( $m$ :  $-0.011$  [ $-0.085, 0.058$ ];  $\sigma$ :  $-1.990$  [ $-4.759, 0.766$ ], where positive contrasts indicate higher rates without a delay). For the older group, the probability of responding from memory was reduced by  $-0.054$  [ $-0.101, -0.009$ ] when no interval was included between events. In other words, for older adults the probability of recalling from memory was around 5% higher with the 2 second interval relative to no interval. However, this is qualified by the fact that, when we contrast the age groups directly, there is no clear evidence that the change in  $m$  between the two interval conditions is larger for the older compared to younger group ( $-0.042$  [ $-0.126, 0.044$ ]). Finally, estimates of the imprecision of memory for the older group did not differ between the two interval conditions ( $0.874$  [ $-2.469, 4.228$ ]).

As in Experiment 1, there was no evidence that older or younger participants made

mis-binding errors between cue objects studied in close temporal proximity to the target (see supplement, section 6).

## Discussion

In this experiment we included a two-alternative forced choice task prior to relocation to assess whether failure of memory for the object cue itself played a role in increasing the guessing rate with delay. Further, in one block of trials we spaced out events to examine whether the time between study and test or the number of intervening events was the main factor determining loss of information from memory. Performance on the forced choice task was uniformly high and the pattern of results was similar to that in Experiment 1, even after excluding trials on which the incorrect object was selected. Thus, it seems clear that the increase in guessing rate is not due to failure to retain the cue object but rather the location itself, or the binding between location and object.

Considering recall error (Figures 5 and 6), while there was a tendency for older adults to relocate objects further away from their presented location, the age difference was not credibly different from zero (unlike Experiment 1). This could be due to the fact that, as mentioned above, we excluded trials on which the correct object was not identified. However, to assess whether the results of Experiments 1 and 2 are quantitatively different, it is important to compare the two directly. When we jointly analyze the data from Experiments 1 and 2 we find no evidence of an age group by experiment interaction for recall error. Rather this analysis suggests an overall age difference in recall error of -15.591 [-28.636, -2.183] pixels. The full results of this cross experiment analysis are presented in the supplement (section 5).

Extending the number of seconds between study and test, by including an extra blank interval between them, did not produce a greater degree of forgetting with delay. This is in line with the number of intervening events, rather than the amount of time, between study and test being the primary driver of declining performance with delay. It may be argued that, as the interval conditions were presented in separate blocks, the extra time in between events may have been used to further strengthen, elaborate on, or consolidate the object-location representation. This may have limited our ability to find increased forgetting with time. There was some evidence that the interval led to a slightly lower guessing rate for the older adults. However, the extra interval did not modulate the overall performance of younger adults nor the decline in performance with increasing study-test delay, so it is not clear that such a strengthening-counteracting-decay explanation works for this group. Of course, our conclusions regarding the role of time itself must be tempered with the fact that we did not test memory beyond a couple of minutes. Including probes of memory on timescales much greater than this is an important avenue for future work with tasks like this.

The results for the mixture model analysis are very similar across Experiments 1 and 2; there is no clear effect of age on the rate of memory failure but a clear age-related increase in the imprecision of recall from memory. There was, however, some ambiguity over whether age differences in  $\sigma$  increase with longer study-test delays, particularly when contrasting the

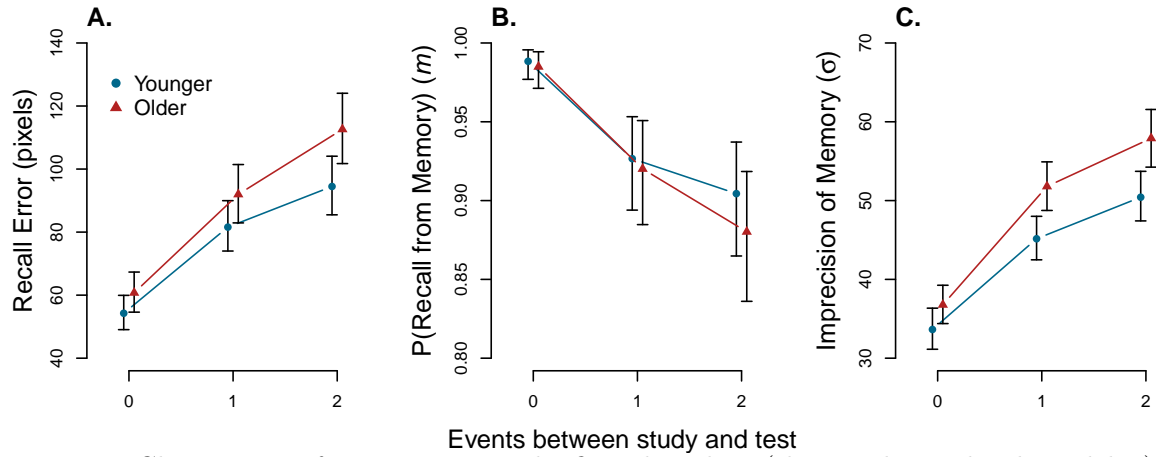


Figure 8. Change in performance across the first three lags (that made up the short delay). (A.) Recall error, (B.) probability of successful recall and (C.) recall imprecision. Data from Experiments 1 and 2 are combined.

shortest and longest delays. This is also true when we combine the data sets from the two experiments; age differences in imprecision at the short delay,  $-5.332$   $[-8.388, -2.173]$ , do not differ credibly from those in the medium,  $-7.890$   $[-12.370, -3.558]$ , and long conditions,  $-8.359$   $[-13.052, -3.803]$  (contrast short vs long:  $3.027$   $[-1.061, 7.200]$ ; see section 5 of the supplement for full analysis). In the broader literature, it is well established that most of the decline in performance in continuous tasks such as these is between the very early lags (i.e., in the time immediately following study; see, e.g., D. C. Rubin, Hinton, & Wenzel, 1999; Wixted, 1990) and some of the evidence discussed in the Introduction suggests that age differences might increase over the first few events following study (e.g., Maddox et al., 2011; Poon & Fozard, 1980). With this in mind we decided to conduct some more focused exploratory analyses of data from the very shortest lags.

### Change in performance within the short delay

Figure 8 presents estimated recall error and the mixture parameters for the three short lags (0, 1, or 2 events between study and test) for the combined data from Experiments 1 and 2. For recall error (Figure 8A), age differences are not significantly different from zero at lag 0 ( $-6.487$   $[-14.704, 1.543]$ ) or lag 1 ( $-10.389$   $[-22.719, 1.599]$ ), whereas at lag 2 the difference is non-zero ( $-18.132$   $[-32.676, -4.242]$ ). Looking at how age differences increase with lag by comparing these conditions we find an increase in the age difference between lags 0 and 2 ( $11.645$   $[0.384, 23.022]$ ), although the increase is not significant between the other lags (0 vs 1:  $3.902$   $[-5.436, 13.447]$ ; 1 vs 2:  $7.743$   $[-4.013, 19.704]$ ). Indeed, it was only within the short delay condition that the age difference significantly increased with increasing lag; comparing the different lags within the medium and long delay categories did not reveal any significant increases in the size of the age difference (see supplement, section 7).

Turning to the mixture model, there were no significant age differences in the probability of recall from memory during the first three lags (lag 0:  $0.003$   $[-0.011, 0.019]$ ; lag 1:  $0.006$

[-0.039, 0.051]; lag 2: 0.024 [-0.033, 0.080]; see Figure 8B). Further, the age contrast did not vary significantly across lags (0 vs 1: -0.003 [-0.048, 0.042]; 0 vs 2: -0.021 [-0.076, 0.035]; 1 vs 2: -0.018 [-0.074, 0.037]).

For memory imprecision (Figure 8C) the age difference at immediate test (lag 0) was small and not credibly different from zero (-3.132 [-6.737, 0.359]), but there were non-zero age differences at lag 1 (-6.644 [-10.706, -2.581]) and 2 (-7.478 [-12.296, -2.543]). However, in comparing the different lags directly, we find that the 95% credible intervals include zero in each case (0 vs 1: 3.512 [-0.892, 7.985]; 0 vs 2: 4.346 [-0.863, 9.316]; 1 vs 2: 0.834 [-4.373, 6.008]). Nevertheless, taking the mixture parameters together, the increasing age differences in recall error across the first few events following study appears to be more consistent with decline in the precision of memory rather than an increase in the failure of older adults' recall.

## General Discussion

Normal aging is associated with poorer performance on a range of memory measures, covering both short- and long-term memory. However, few studies have attempted to compare age differences in memory over the short- and long-term simultaneously. Older adults' performance may primarily be limited by poorer encoding of information (Bartsch et al., 2019; Hara & Naveh-Benjamin, 2015) leading to less precise representations (S.-C. Li et al., 2000, 2005) and worse recall at all delays. However, additional factors may come into play over longer delay intervals, such as a specific age-related deficit in search and retrieval from long-term memory (Healey & Kahana, 2016; Wahlheim & Huff, 2015) and increased susceptibility to interference from intervening events (Hasher & Zacks, 1988; Hedden & Park, 2003). To address these issues in the present experiments, we used a task in which participants recalled object location following a variable delay. Thus, in addition to being able to assess how age differences in recall error evolved with longer study-test delay, we were also able to investigate to what extent age differences in recall could be attributed to imprecise memory for the target location versus complete failure of recall.

To summarize our findings across both experiments, we find that older adults exhibit a greater degree of error in recalling location, relative to younger adults. Further, this appears to be primarily driven by more imprecise recall of object location rather than an age-related increase in the rate of complete retrieval failure, in line with other recent reports (Korkki et al., 2020; Nilakantan et al., 2018). Importantly, our task allowed us to probe memory following short and long delays with an identical study procedure. Further, as participants were unaware of how long it would be before a particular object-location pair was tested, differences between conditions can not be attributed to encoding or strategy differences between delays (see Hinrichs & Grunke, 1975). At the coarse level of short (0-2 intervening events), medium (around 10 intervening events) and long (around 20 intervening events) delays, there was no convincing evidence that age differences increased with delay. However, when focusing on changes over the first few events following study, we find that age differences appear with only one or two intervening events between study and test.

These findings align with some of those discussed in the introduction. In particular,

they are consistent with the findings of Maddox et al. (2011), who found that the benefits of repeated testing for paired associates were greatly diminished for older adults when there were two or more intervening events between tests. Also our findings are consistent with those of Poon and Fozard (1980), who found that age differences in continuous recognition performance increased considerably in the first few events following study then stayed relatively consistent (see also Wickelgren, 1975). On the other hand, Chen and Naveh-Benjamin (2012) found that older adults exhibited an associative recognition deficit when tested immediately. Importantly, this was driven by a higher rate of false alarms, whereas there were no clear age differences in hit rate. Thus, it appears that older adults were accurate at recognizing just seen information, consistent with the present recall findings.

Interestingly, there is a potential link to age differences in performance on the  $n$ -back task, a widely used working memory task in which participants view a continuous stream of stimuli and indicate whether each item was presented  $n$  positions back in the sequence or not. Bopp and Verhaeghen (2020) present a meta-analysis of age differences in this task and show that, for accuracy, differences are minimal at 1-back (analogous to our 0 lag) and increase considerably at 2-back. Furthermore, age differences in accuracy do not appear to increase further beyond 2-back.

So what do the present findings, along with the other work discussed here, tell us about age differences in memory performance? For sequentially presented items it seems that age differences in initial encoding and storage are minimal, as can be seen in the small, and not credibly different from zero, group difference at lag 0 (Figure 8). While age may influence stimulus perception and the ability to make the fine motor movements associated with this recall task, this does not appear to play a determining role in later recall (see Korkki et al., 2020). This seems consistent with Bopp and Verhaeghen’s suggestion that older adults are able to effectively retrieve a single item (or, in this case, association) from the focus of attention. However, once information is not in this highly accessible state, age differences in the retrieval of information begin to appear. Our findings suggest these age differences in the precision of recall do not increase considerably following the first few intervening events up to a delay of a few minutes. This suggests that, while interference from intervening events may contribute to the increase in recall error with delay seen in both groups, older adults are not more susceptible to retroactive interference in this task. While delays of several minutes are regularly used in studies of episodic long-term memory (e.g., Brady et al., 2013; Korkki et al., 2020; Nilakantan et al., 2018), it will be important for future studies to assess whether age differences in performance increase over even longer timescales.

The task used here allowed us to go beyond previous findings in probing the nature of age differences in retrieval following different delays. Using mixture modeling to disentangle sources of error, we found that older adults are not significantly more likely to fail to retrieve something from memory, and subsequently guess, relative to younger adults, but their recall is more imprecise (also Korkki et al., 2020; Nilakantan et al., 2018). Strikingly, this holds even at our longest delay, where age differences in the rate of retrieval failure would be expected, given the well established age related deficit for recall from long-term memory (Danckert & Craik, 2013; Rhodes et al., 2019). Why is this the case? A critical point to



note is that, unlike many other recall tasks in which a particular item must be recalled in response to a cue, participants may be able to use a more coarse form of representation that encodes the general location of an object (for example, that it was in the top-right quadrant). This is akin to the argument that older adults rely more on “familiarity” during recognition as opposed to precise recollection of an item or event (e.g., Howard, Bessette-Symons, Zhang, & Hoyer, 2006; Jennings & Jacoby, 1993, 1997). It is also in line with the finding that older adults make more gist-related errors in memory tasks, where they are more likely to endorse an unstudied item as studied if it is similar to an item they previously saw (Greene & Naveh-Benjamin, 2020; Koutstaal & Schacter, 1997; Paige, Cassidy, Schacter, & Gutchess, 2016; S. M. Stark et al., 2013; Trelle, Henson, Green, & Simons, 2017). Thus, it may be that older adults do not exhibit a higher rate of complete memory failure as they can recover coarse information about the location of an object, which results in somewhat greater imprecision.

In the working memory literature there is growing acknowledgement that participants exhibit reliable biases in precise recall tasks. For example, canonical color categories are over-represented in responses on a color wheel (Bae, Olkkonen, Allred, & Flombaum, 2015; Hardman, Vergauwe, & Ricker, 2017) and orientation responses are biased by the cardinal orientations (Pratte, Park, Rademaker, & Tong, 2017). This has resulted in models that distinguish representations of fine-grained, *continuous* information from representations of coarse *categorical* information (e.g., a prototypical blue). Crucially, for tasks like ours there are also well known “category effects”. Huttenlocher and colleagues (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, Corrigan, & Crawford, 2004) presented participants with a single dot at a time in a circular study area and had participants recall the location of the dot. Their work demonstrated that participants exhibit strong biases in this task that are consistent with participants implicitly dividing the study area into four quadrants, with boundaries at the vertical and horizontal axes. They found that participants are consistently biased away from these category boundaries and this bias increases following distraction (Huttenlocher et al., 1991).

In section 8 of the supplementary material we present additional analyses, focusing specifically on the angles of the locations recalled, to assess the influence of categorical representations in our task. Interestingly, we see clear over-representation of certain angles in responses, consistent with bias towards certain locations. Further, using a mixture model designed to distinguish the recall of continuous and categorical information (see Hardman et al., 2017), we find that recall of categorical information increases substantially in both age groups from the short to medium delays. We also find that the recall of categorical information increases further between the medium and long delays for the older group but not the younger group. Of course, we did not set out to examine this issue and these analyses are exploratory. Nevertheless, these results do suggest some age differences in the recall of categorical information and we think the categorical-continuous distinction could prove useful when assessing age differences in tasks such as the one used here. Further work examining this distinction may also forge links between the relatively new “precision” literature and the broader literature on memory and aging (e.g., the literature on gist mentioned above). Future work may also extend our approach to look at other stimulus dimensions, such as color (Biderman et al., 2019; Brady et al., 2013), orientation (Korkki et al., 2020), shape (A.



Y. Li, Liang, Lee, & Barense, 2020), and faces (Lorenc, Pratte, Angeloni, & Tong, 2014).

## Conclusion

Here we showed age-related decline in the recall of the precise location of objects and this age difference did not substantially increase from delays of seconds, used typically in studies of short-term or working memory, to those of minutes, used typically in studies of long-term episodic memory. Rather the clearest change in group differences with delay was immediately following study; age differences were largely absent when probed immediately to recall object location and emerged following only one or two intervening events. Our findings also build on a handful of fine-grained recall studies that have found no convincing age difference in the probability of successful recall from memory, but clear differences in precision. These age differences in the precision of successful recall emerge fairly quickly following study and remain relatively constant to delays typical in studies of episodic memory. Older adults appear to lose fine grained spatial information more quickly over the short-term.

## References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier.
- Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). Academic Press.
- Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, 144(4), 744–763. <https://doi.org/10.1037/xge0000076>
- Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does Differential Strategy Use Account for Age-Related Deficits in Working-Memory Performance? *Psychology and Aging*, 24(1), 82–92. <https://doi.org/10.1037/a0014078>
- Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging*, 4(1), 3–9. <https://doi.org/10.1037/0882-7974.4.1.3>
- Bartsch, L. M., Loaiza, V. M., & Oberauer, K. (2019). Does limited working memory capacity underlie age differences in associative long-term memory? *Psychology and Aging*, 34(2), 268–281. <https://doi.org/10.1037/pag0000317>
- Bastin, C., & Van Der Linden, M. (2005). The effects of aging on the recognition of different types of associations. *Experimental Aging Research*, 32(1), 61–77. <https://doi.org/10.1016/j.exger.2004.06.005>

- [//doi.org/10.1080/03610730500326291](https://doi.org/10.1080/03610730500326291)
- Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9(10), 7. <https://doi.org/10.1167/9.10.7>
- Benton, A. L., & Hamsher, K. deS. (1976). *Multilingual aphasia examination*. University of Iowa Press.
- Biderman, N., Luria, R., Teodorescu, A. R., Hajaj, R., & Goshen-Gottstein, Y. (2019). Working Memory Has Better Fidelity Than Long-Term Memory: The Fidelity Constraint Is Not a General Property of Memory After All. *Psychological Science*, 30(2), 223–237. <https://doi.org/10.1177/0956797618813538>
- Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, 60(5), P223—P233. <https://doi.org/10.1093/geronb/60.5.P223>
- Bopp, K. L., & Verhaeghen, P. (2020). Aging and n-back performance: A meta-analysis. *The Journals of Gerontology: Series B*, 75(2), 229–240. <https://doi.org/10.1093/geronb/gby024>
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, 105(38), 14325–14329. <https://doi.org/10.1073/pnas.0803390105>
- Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual Long-Term Memory Has the Same Limit on Fidelity as Visual Working Memory. *Psychological Science*, 24(6), 981–990. <https://doi.org/10.1177/0956797612465439>
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28. <https://doi.org/10.18637/jss.v080.i01>
- Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *R Journal*, 10(1), 395–411. <https://doi.org/10.32614/rj-2018-017>
- Bürkner, P.-C., & Charpentier, E. (2018). Monotonic effects: A principled approach for including ordinal predictors in regression models. *PsyArXiv Preprints*, 1–20. <https://doi.org/10.31234/osf.io/9qkhj>
- Bürkner, P.-C., & Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 2(1), 77–101. <https://doi.org/10.1177/2515245918823199>
- Campbell, K. L., Trelle, A. N., & Hasher, L. (2014). Hyper-binding across time: Age differences in the effect of temporal proximity on paired-associate learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, 40(1), 300–305. <https://doi.org/10.1037/a0034109>
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M.,

- ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1). <https://doi.org/10.18637/jss.v076.i01>
- Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory and Cognition*, 24(4), 403–416. <https://doi.org/10.3758/BF03200930>
- Chen, T., & Naveh-Benjamin, M. (2012). Assessing the associative deficit of older adults in long-term and short-term/working memory. *Psychology and Aging*, 27(3), 666–682. <https://doi.org/10.1037/a0026943>
- Cowan, N. (1988). Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Constraints Within the Human Information-Processing System. *Psychological Bulletin*, 104(2), 163–191. <https://doi.org/10.1037/0033-2909.104.2.163>
- Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin*, 145(8), 822–847. <https://doi.org/10.1037/bul0000199>
- Cowan, N., Johnson, T. D., & Saults, J. S. (2005). Capacity limits in list item recognition: Evidence from proactive interference. *Memory*, 13(3-4), 293–299. <https://doi.org/10.1080/09658210344000206>
- Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Treub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York: Plenum.
- Danckert, S. L., & Craik, F. I. M. (2013). Does aging affect recall more than recognition memory? *Psychology and Aging*, 28(4), 902–909. <https://doi.org/10.1037/a0033263>
- Ferris, S. H., Crook, T., Clark, E., McCarthy, M., & Rae, D. (1980). Facial recognition memory deficits in normal aging and senile dementia. *Journals of Gerontology*, 35(5), 707–714. <https://doi.org/10.1093/geronj/35.5.707>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 3). Chapman & Hall/CRC Boca Raton, FL, USA.
- Glisky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double Dissociation Between Item and Source Memory. *Neuropsychology*, 9(2), 229–235. <https://doi.org/10.1037/0894-4105.9.2.229>
- Greene, N. R., & Naveh-Benjamin, M. (2020). A specificity principle of memory: Evidence from aging and associative memory. *Psychological Science*, 31(3), 316–331. <https://doi.org/10.1177/0956797620901760>
- Hampstead, B. M., Towler, S., Stringer, A. Y., & Sathian, K. (2018). Continuous measurement of object location memory is sensitive to effects of age and mild cognitive impairment and related to medial temporal lobe volume. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, 10, 76–85. <https://doi.org/10.1016/j.dadm.2018.05.001>

[//doi.org/10.1016/j.dadm.2017.10.007](https://doi.org/10.1016/j.dadm.2017.10.007)

- Hara, Y., & Naveh-Benjamin, M. (2015). The role of reduced working memory storage and processing resources in the associative memory deficit of older adults: Simulation studies with younger adults. *Aging, Neuropsychology, and Cognition*, 22(2), 129–154. <https://doi.org/10.1080/13825585.2014.889650>
- Hardman, K. O. (2017). *CatContModel: Categorical and continuous working memory models for delayed-estimation tasks*.
- Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1), 30–54. <https://doi.org/10.1037/xhp0000290>
- Harlow, I. M., & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin and Review*, 20(2), 318–325. <https://doi.org/10.3758/s13423-012-0340-9>
- Harlow, I. M., & Yonelinas, A. P. (2016). Distinguishing between the success and precision of recollection. *Memory*, 24(1), 114–127. <https://doi.org/10.1080/09658211.2014.988162>
- Hasher, L., & Zacks, R. T. (1988). Working Memory, Comprehension, and Aging: A Review and a New View. *Psychology of Learning and Motivation - Advances in Research and Theory*, 22(C), 193–225. [https://doi.org/10.1016/S0079-7421\(08\)60041-9](https://doi.org/10.1016/S0079-7421(08)60041-9)
- Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, 123(1), 23–69. <https://doi.org/10.1037/rev0000015>
- Hedden, T., & Park, D. C. (2003). Contributions of Source and Inhibitory Mechanisms to Age-Related Retroactive Interference in Verbal Working Memory. *Journal of Experimental Psychology: General*, 132(1), 93–112. <https://doi.org/10.1037/0096-3445.132.1.93>
- Hinrichs, J. V., & Grunke, M. E. (1975). Control processes in short-term memory: Use of retention interval information. *Journal of Experimental Psychology: Human Learning and Memory*, 1(3), 229–237. <https://doi.org/10.1037/0278-7393.1.3.229>
- Howard, M. W., Bessette-Symons, B., Zhang, Y., & Hoyer, W. J. (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. <https://doi.org/10.1037/0882-7974.21.1.96>
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and Particulars: Prototype Effects in Estimating Spatial Location. *Psychological Review*, 98(3), 352–376. <https://doi.org/10.1037/0033-295X.98.3.352>
- Huttenlocher, J., Hedges, L. V., Corrigan, B., & Crawford, L. E. (2004). Spatial categories and the estimation of location. *Cognition*, 93(2), 75–97. <https://doi.org/10.1016/j>

[cognition.2003.10.006](#)

- Jaroslawska, A. J., & Rhodes, S. (2019). Adult age differences in the effects of processing on storage in working memory: A meta-analysis. *Psychology and Aging, 34*(4), 512–530. <https://doi.org/10.1037/pag0000358>
- Jennings, J. M., & Jacoby, L. L. (1993). Automatic Versus Intentional Uses of Memory: Aging, Attention, and Control. *Psychology and Aging, 8*(2), 283–293. <https://doi.org/10.1037/0882-7974.8.2.283>
- Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology and Aging, 12*(2), 352–361. <https://doi.org/10.1037/0882-7974.12.2.352>
- Korkki, S. M., Richter, F. R., Jeyarathnarajah, P., & Simons, J. S. (2020). Healthy Ageing Reduces the Precision of Episodic Memory Retrieval. *Psychology and Aging, 35*(1), 124. <https://doi.org/10.1037/pag0000432>
- Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37*(4), 555–583. <https://doi.org/10.1006/jmla.1997.2529>
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.
- Le Breck, D. B., & Baron, A. (1987). Age and practice effects in continuous recognition memory. *Journals of Gerontology, 42*(1), 89–91. <https://doi.org/10.1093/geronj/42.1.89>
- Lew, T. F., Pashler, H. E., & Vul, E. (2016). Fragile associations coexist with robust memories for precise details in long-term memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 42*(3), 379–393. <https://doi.org/10.1037/xlm0000178>
- Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100*(9), 1989–2001. <https://doi.org/10.1016/j.jmva.2009.04.008>
- Li, A. Y., Liang, J. C., Lee, A. C., & Barense, M. D. (2020). The validated circular shape space: Quantifying the visual similarity of shape. *Journal of Experimental Psychology: General, 149*(5), 949. <https://doi.org/10.1037/xge0000693>
- Li, S.-C., Lindenberger, U., & Frensch, P. A. (2000). Unifying cognitive aging: From neuromodulation to representation to cognition. *Neurocomputing, 32*, 879–890. [https://doi.org/10.1016/S0925-2312\(00\)00256-3](https://doi.org/10.1016/S0925-2312(00)00256-3)
- Li, S.-C., Naveh-Benjamin, M., & Lindenberger, U. (2005). Aging neuromodulation impairs associative binding: A neurocomputational account. *Psychological Science, 16*(6), 445–450. <https://doi.org/10.1111/j.0956-7976.2005.01555.x>
- Liddell, T. M., & Kruschke, J. K. (2018). Analyzing ordinal data with metric models: What could possibly go wrong? *Journal of Experimental Social Psychology, 79*, 328–348.

<https://doi.org/10.1016/j.jesp.2018.08.009>

- Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science*, 20(4), 240–245. <https://doi.org/10.1177/0963721411415340>
- Lorenc, E. S., Pratte, M. S., Angeloni, C. F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, and Psychophysics*, 76(7), 1975–1984. <https://doi.org/10.3758/s13414-014-0653-z>
- Maddox, G. B., Balota, D. A., Coane, J. H., & Duchek, J. M. (2011). The Role of Forgetting Rate in Producing a Benefit of Expanded Over Equal Spaced Retrieval in Young and Older Adults. *Psychology and Aging*, 26(3), 661–670. <https://doi.org/10.1037/a0022942>
- Mickes, L., Johnson, E. M., & Wixted, J. T. (2010). Continuous Recollection Versus Unitized Familiarity in Associative Recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, 36(4), 843–863. <https://doi.org/10.1037/a0019755>
- Mitchell, K. J., Johnson, M. K., Raye, C. L., Mather, M., & D’Esposito, M. (2000). Aging and reflective processes of working memory: Binding and test load deficits. *Psychology and Aging*, 15(3), 527–541. <https://doi.org/10.1037/0882-7974.15.3.527>
- Naveh-Benjamin, M. (1987). Coding of Spatial Location Information: An Automatic Process? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(4), 595–605. <https://doi.org/10.1037/0278-7393.13.4.595>
- Naveh-Benjamin, M. (1988). Recognition memory of spatial location information: Another failure to support automaticity. *Memory & Cognition*, 16(5), 437–445. <https://doi.org/10.3758/BF03214224>
- 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(5), 1170–1187. <https://doi.org/10.1037/0278-7393.26.5.1170>
- Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging*, 22(1), 202–208. <https://doi.org/10.1037/0882-7974.22.1.202>
- Nilakantan, A. S., Bridge, D. J., VanHaerents, S., & Voss, J. L. (2018). Distinguishing the precision of spatial recollection from its success: Evidence from healthy aging and unilateral mesial temporal lobe resection. *Neuropsychologia*, 119, 101–106. <https://doi.org/10.1016/j.neuropsychologia.2018.07.035>
- Noack, H., Lövdén, M., & Lindenberger, U. (2012). Normal aging increases discriminial dispersion in visuospatial short-term memory. *Psychology and Aging*, 27(3), 627–637. <https://doi.org/10.1037/a0027251>
- Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological*



- Bulletin*, 143(9), 992–1009. <https://doi.org/10.1037/bul0000108>
- Oberauer, K. (2002). Access to Information in Working Memory: Exploring the Focus of Attention. *Journal of Experimental Psychology: Learning Memory and Cognition*, 28(3), 411–421. <https://doi.org/10.1037/0278-7393.28.3.411>
- Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning Memory and Cognition*, 43(1), 1–22. <https://doi.org/10.1037/xlm0000302>
- Old, S. R., & 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. <https://doi.org/10.1037/0882-7974.23.1.104>
- Paige, L. E., Cassidy, B. S., Schacter, D. L., & Gutchess, A. H. (2016). Age differences in hippocampal activation during gist-based false recognition. *Neurobiology of Aging*, 46, 76–83. <https://doi.org/10.1016/j.neurobiolaging.2016.06.014>
- Peich, M. C., Husain, M., & Bays, P. M. (2013). Age-related decline of precision and binding in visual working memory. *Psychology and Aging*, 28(3), 729–743. <https://doi.org/10.1037/a0033236>
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., . . . Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- Pertsov, Y., Heider, M., Liang, Y., & Husain, M. (2015). Effects of healthy ageing on precision and binding of object location in visual short term memory. *Psychology and Aging*, 30(1), 26–35. <https://doi.org/10.1037/a0038396>
- Peterson, D. J., Gargya, S., Kopeikin, K. S., & Naveh-Benjamin, M. (2017). The impact of level of education on age-related deficits in associative memory: Behavioral and neuropsychological perspectives. *Cortex*, 91, 9–24. <https://doi.org/10.1016/j.cortex.2016.12.020>
- Peterson, D. J., Schmidt, N. E., & Naveh-Benjamin, M. (2017). The role of schematic support in age-related associative deficits in short-term and long-term memory. *Journal of Memory and Language*, 92, 79–97. <https://doi.org/10.1016/j.jml.2016.05.007>
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A Context Maintenance and Retrieval Model of Organizational Processes in Free Recall. *Psychological Review*, 116(1), 129–156. <https://doi.org/10.1037/a0014420>
- Poon, L. W., & Fozard, J. L. (1980). Age and word frequency effects in continuous recognition memory. *Journals of Gerontology*, 35(1), 77–86. <https://doi.org/10.1093/geronj/35.1.77>
- Pratte, M. S., Park, Y. E., Rademaker, R. L., & Tong, F. (2017). Accounting for stimulus-specific variation in precision reveals a discrete capacity limit in visual working

- memory. *Journal of Experimental Psychology: Human Perception and Performance*, 43(1), 6–17. <https://doi.org/10.1037/xhp0000302>
- R Development Core Team 3.0.1. (2013). *A Language and Environment for Statistical Computing* (Vol. 2, pp. <https://www.R-project.org>). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org>
- Raaijmakers, J. G. W., & Shiffrin, R. M. (2002). Models of memory. In H. Pashler & D. Medin (Eds.), *Steven's handbook of experimental psychology: Memory and cognitive processes* (pp. 43–76–195). John Wiley & Sons Inc.
- Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, 12(13), 21. <https://doi.org/10.1167/12.13.21>
- Rhodes, S., Greene, N. R., & Naveh-Benjamin, M. (2019). Age-related differences in recall and recognition: a meta-analysis. *Psychonomic Bulletin and Review*, 26(5), 1529–1547. <https://doi.org/10.3758/s13423-019-01649-y>
- Richter, F. R., Cooper, R. A., Bays, P. M., & Simons, J. S. (2016). Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *eLife*, 5(OCTOBER2016), e18260. <https://doi.org/10.7554/eLife.18260>
- Rouder, J. N., Thiele, J. E., Province, J. M., & Cusumano, M. (2014). The evidence for a guessing state in working-memory judgments. *Unpublished Manuscript*, Retrieved from: <http://pcl.missouri.edu/node/134>.
- Rubin, D. C., Hinton, S., & Wenzel, A. (1999). The Precise Time Course of Retention. *Journal of Experimental Psychology: Learning Memory and Cognition*, 25(5), 1161–1176. <https://doi.org/10.1037/0278-7393.25.5.1161>
- Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences*, 26(6), 709–728. <https://doi.org/10.1017/s0140525x03000165>
- Schneegans, S., & Bays, P. M. (2016). No fixed item limit in visuospatial working memory. *Cortex*, 83, 181–193. <https://doi.org/10.1016/j.cortex.2016.07.021>
- Sharps, M. J., & Gollin, E. S. (1988). Aging and free recall for objects located in space. *Journals of Gerontology*, 43(1), P8—P11. <https://doi.org/10.1093/geronj/43.1.P8>
- Souza, A. S. (2016). No age deficits in the ability to use attention to improve visual working memory. *Psychology and Aging*, 31(5), 456–470. <https://doi.org/10.1037/pag0000107>
- Spencer, W. D., & Raz, N. (1995). Differential Effects of Aging on Memory for Content and Context: A Meta-Analysis. *Psychology and Aging*, 10(4), 527–539. <https://doi.org/10.1037/0882-7974.10.4.527>
- Stan Development Team. (2018). RStan: the R interface to Stan. Retrieved from <http://mc-stan.org/>
- Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. (2013). A task to as-



- sess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, 51(12), 2442–2449. <https://doi.org/10.1016/j.neuropsychologia.2012.12.014>
- Trelle, A. N., Henson, R. N., Green, D. A., & Simons, J. S. (2017). Declines in representational quality and strategic retrieval processes contribute to age-related increases in false recognition. *Journal of Experimental Psychology: Learning Memory and Cognition*, 43(12), 1883–1897. <https://doi.org/10.1037/xlm0000412>
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, 114(1), 104–132. <https://doi.org/10.1037/0033-295X.114.1.104>
- Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, 27(5), 1413–1432. <https://doi.org/10.1007/s11222-016-9696-4>
- Verhaeghen, P., Marcoen, A., & Goossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journals of Gerontology*, 48(4), P157—P171. <https://doi.org/10.1093/geronj/48.4.P157>
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin and Review*, 14(5), 779–804. <https://doi.org/10.3758/BF03194105>
- Wahlheim, C. N., & Huff, M. J. (2015). Age differences in the focus of retrieval: Evidence from dual-list free recall. *Psychology and Aging*, 30(4), 768–780. <https://doi.org/10.1037/pag0000049>
- Wechsler, D. (1987). *WMS-R: Wechsler memory scale-revised*. Psychological Corporation.
- Wechsler, D. (1997). *Wechsler memory scale-(WMS-III)*. Psychological Corporation.
- Wechsler, D. (n.d.). *WAIS-R manual: Wechsler adult intelligence scale-revised*. Psychological Corporation.
- Wickelgren, W. A. (1975). Age and storage dynamics in continuous recognition memory. *Developmental Psychology*, 11(2), 165–169. <https://doi.org/10.1037/h0076457>
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1), 1–29. <https://doi.org/10.18637/jss.v040.i01>
- Wixted, J. T. (1990). Analyzing the Empirical Course of Forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(5), 927–935. <https://doi.org/10.1037/0278-7393.16.5.927>
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441–517. <https://doi.org/10.1006/jmla.2002.2864>
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192), 233–235. <https://doi.org/10.1038/nature06860>