Supplement to 'Age differences in the precision of memory at short and long delays'

Stephen Rhodes[1], Emily E. Abbene[2], Ashley M. Meierhofer[2], & Moshe Naveh-Benjamin[2]

[1] Rotman Research Institute
[2] University of Missouri

Abstract

This document provides supplementary information to "Age differences in the precision of memory at short and long delays". It presents (1) a summary of the neuropsychological battery completed by participants and a table of scores, (2) extra detail on the analyses reported in the main manuscript; (3) tables of model coefficients on transformed scale for Experiments 1 and 2 (see paper for contrasts on the 'manifest' scale); (4) analysis of confidence ratings from Experiment 1; (5) the results of combined analyses of the data from Experiments 1 and 2; (6) an assessment of the possibility that participants commit mis-binding errors to other locations studied in close temporal proximity to the target location; (7) results of analyses assessing recall error and mixture parameters as a function of study-test lag (as opposed to the 'short', 'medium', and 'long' delay conditions); (8) an analysis assessing the possible role of (gist-like) categorical representations in the recall of spatial location.

1 Neuropsychological Battery

Participants completed a battery of several neuropsychological tasks (following Glisky, Polster, & Routhieaux, 1995; Peterson, Gargya, Kopeikin, & Naveh-Benjamin, 2017). The tasks used were (1) controlled oral word association (Benton & Hamsher, 1976), (2) mental arithmetic from the Wechsler Adult Intelligence Scale-Revised (WAIS-R: Wechsler, n.d.), (3) mental control from the Wechsler Memory Scale-Revised (WMS-R: Wechsler, 1987), (4) digit span also from the WMS-R, (5) logical memory I and II (immediate and delayed), and (6) family pictures I and II from the Wechsler Memory Scale-Third Edition (WMS-III: Wechsler,

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

Table 1
Background information for samples in the present experiments. Mean (and standard deviation) raw scores are given for the neuropsychological measures. An asterisk signifies that older performance was significantly lower than that of younger adults according to a one sided Welch t-test ($p < 0.05$). The test for age differences in years of education was two sided. Due to an error, data for logical memory I is missing for 3 younger participants in Experiment 2.

	Experiment 1		Experiment 2	
	Younger	Older	Younger	Older
N	33	32	34	31
Age	19.61 (1.78)	73.59 (3.96)	18.53 (1.02)	73.32 (5.47)
Years of Education	13.30 (0.95)	14.72 (2.07)*	12.32 (0.88)	14.81 (2.39)*
Logical Memory I	38.52 (8.61)	40.59 (8.98)	45.74 (8.96)	41.90 (10.54)
Family Pictures I	44.06 (10.14)	36.72 (11.07)*	53.03 (5.26)	39.03 (10.63)*
Logical Memory II	23.67 (6.54)	22.53 (8.38)	30.06 (7.69)	24.06 (8.18)*
Family Pictures II	44.97 (10.03)	35.94 (11.31)*	52.38 (9.16)	39.23 (11.01)*
Oral Word Association	35.52 (9.34)	38.38 (9.18)	34.62 (7.36)	36.29 (8.12)
Mental Arithmetic	8.18 (2.65)	9.38 (2.99)	9.09 (2.63)	9.94 (2.73)
Mental Control	22.48 (5.19)	20.41 (4.30)*	24.71 (5.25)	22.00 (6.01)*
Digit Span	7.88 (2.27)	7.44 (2.35)	7.53 (1.44)	6.45 (1.82)*

1997). The first four tasks and the immediate logical memory and family pictures tasks were completed at the beginning of the session and the delayed tasks were completed after the main continuous recall task described below (a delay of about 30 minutes). Neuropsychological scores for both experiments can be found in Table 1.

2 Modeling Details

2.1 Recall Error

As noted in the main manuscript, we analyzed recall error with a generalized linear mixed effects model in which error was assumed to be log-normally distributed. The model included fixed effects of delay, age group, and inter-event interval in Experiment 2 (see the main manuscript for how these variables were coded). Recall error in pixels was divided by 100 prior to analysis and priors were chosen to be minimally informative on this scale. For the intercept and fixed effect coefficients we used a Cauchy(0, 2.5) prior distribution. The model also included a random effect term for participants, allowing the mean of the log-normal distribution to vary across individuals, and an independent random effect for the imaged used to cue location recall. The standard deviation for these random effects was given a folded Cauchy(0, 2.5) prior. The default prior used by `brms` was used as the prior for the standard deviation of the log-normal distribution, which is a half Student's t distribution with 3 degrees of freedom, centered on zero with scale 10.

2.2 Mixture Model

As discussed in the main manuscript, we modeled our recall data as a mixture of responses from memory, with some imprecision, and random guesses due to a failure of recall. The recalled location, $\mathbf{r}$, for individual i, in delay condition j, on trial k was assumed to be distributed as follows:

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

This is a mixture of a bivariate normal distribution centered on the presented location, $\mathbf{p}$, with standard deviation σ_{ij}, and a uniform distribution on the study area, A. The former are assumed to reflect responses from (noisy) memory and the latter are assumed to be guesses. The parameter m_{ij} reflects the proportion of responses from memory for individual i in condition j.

The parameters were modeled as follows: we modeled transformations of the parameters that map them onto the real line and, therefore, allow the parameters to be modeled with a linear model. Let $\mathbf{X}$ be a matrix of contrast codes with J rows, where J is the number of conditions. We used the following sum-to-zero coding scheme for the factor of delay:

```
##        [,1] [,2] [,3]
## short     1    1    0
## medium    1   -1   -1
## long      1    0    1
```

We used the same contrast matrix for both m and σ so we estimate three coefficients for per mixture parameter for each participant. The first coefficient is the grand mean of the parameter (on the transformed latent scale), the second coefficient reflects the deviation from this grand mean associated with the short condition, and the third coefficient reflects the deviation associated with the long condition. The medium condition was set as the reference, and its deviation from the grand mean can be found by multiplying coefficients 2 and 3 by -1 and taking the sum.

Parameter vectors for each individual participant, $\mathbf{b}_i$, were sampled from a multivariate normal distribution with population mean vector β and covariance matrix $\mathbf{\Sigma}$. The m and σ parameters were modeled with independent population distributions. Therefore, m for individual i in condition j was determined as follows:

$$\begin{aligned}\text{logit}(m_{ij}) &= \mathbf{b}_i^{(m)} \cdot \mathbf{X}_j \\ \mathbf{b}_i^{(m)} &\sim MVN(\beta^{(m)}, \mathbf{\Sigma}^{(m)})\end{aligned}$$

Analogously, for σ:

$$\begin{aligned}\log(\sigma_{ij}) &= \mathbf{b}_i^{(\sigma)} \cdot \mathbf{X}_j \\ \mathbf{b}_i^{(\sigma)} &\sim MVN(\beta^{(\sigma)}, \mathbf{\Sigma}^{(\sigma)})\end{aligned}$$

Table 2
Experiment 1 results of analysis of recall error.

	Estimate	Q2.5	Q97.5
Intercept	-0.234	-0.323	-0.143
Short	-0.410	-0.442	-0.379
Medium	0.124	0.093	0.156
Long	0.286	0.254	0.319
Age group	-0.097	-0.183	-0.012
Age × S	-0.010	-0.041	0.022
Age × M	0.019	-0.014	0.050
Age × L	-0.009	-0.041	0.023

Priors are needed for the population level parameters β and $\mathbf{\Sigma}$. We chose Normal(0, 5) priors for the $\beta^{(m)}$ and $\beta^{(\sigma)}$ parameters. Note that these priors a placed on transformations of these parameters and are quite broad on the transformed scales. For example, for the intercept of the σ parameter approximately 68% of the prior should fall between exp(-5) = 0.01 and exp(5) = 148.41.

The variance-covariance matrix is split into a correlation matrix and standard deviations associated with each coefficient. We use the LKJ prior (Lewandowski, Kurowicka, & Joe, 2009) in `Stan` with the shape parameter set to 2. This places greater prior density on smaller correlations but is broad enough to not rule out strong correlations (a shape parameter of 1 gives a uniform correlation matrix). For the standard deviation parameters, we use a half Cauchy(0, 2.5) prior distribution. The model was implemented in `Stan` using the `rstan` package. The implementation of the model differs slightly from that presented here in an effort to speed up posterior sampling (see `Stan` user manual for details).

3 Tables of Model Coefficients

3.1 Experiment 1

Table 2 presents estimates of fixed effects from the log normal linear mixed effects model fit to the data from Experiment 1. Recall that 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). For the factor of study-test delay, the medium condition was chosen as the reference level. Therefore, in Table 2 the coefficients associated with the medium condition are the sum of the other coefficients each multiplied by -1.

Table 3 presents the results of the cumulative mixed effects model, which models ratings as coming from a underlying distribution with estimated thresholds for determining the rating given (see Bürkner & Vuorre, 2019). The table presented the 5 estimated thresholds and the fixed effects of age and delay that shift the underlying logistic distribution to higher or lower ratings.

Table 3
Experiment 1 results of analysis of confidence ratings.

	Estimate	Q2.5	Q97.5
Threshold[1]	-4.050	-4.355	-3.743
Threshold[2]	-2.945	-3.237	-2.653
Threshold[3]	-2.060	-2.346	-1.771
Threshold[4]	-0.941	-1.223	-0.654
Threshold[5]	0.549	0.269	0.837
Short	1.039	0.968	1.113
Medium	-0.430	-0.495	-0.366
Long	-0.609	-0.675	-0.544
Age group	-0.029	-0.303	0.249
Age × S	0.104	0.034	0.173
Age × M	-0.014	-0.078	0.051
Age × L	-0.090	-0.154	-0.023

Table 4
Experiment 1 results of analysis of recall error as a function of confidence rating. Confidence was inluded as a monotonic predictor and the coefficient below can be interpreted as the expected average difference between adjacent categories.

	Estimate	Q2.5	Q97.5
Intercept	-0.714	-0.793	-0.635
Short	-0.226	-0.268	-0.185
Medium	0.079	0.037	0.124
Long	0.147	0.108	0.188
Age group	-0.084	-0.160	-0.008
Confidence rating	0.357	0.329	0.385
Age × S	0.017	-0.017	0.052
Age × M	0.007	-0.035	0.053
Age × L	-0.024	-0.072	0.018
Confidence × S	0.028	0.004	0.056
Confidence × M	-0.025	-0.053	-0.001
Confidence × L	-0.003	-0.026	0.019
Age × Confidence × S	-0.041	-0.071	-0.014
Age × Confidence × M	0.029	0.004	0.058
Age × Confidence × L	0.012	-0.011	0.035

Table 4 presents coefficients for the analysis of recall error as a function of confidence rating (and age and delay).

Table 5
Experiment 2 results of analysis of recall error.

	Estimate	Q2.5	Q97.5
Intercept	-0.246	-0.332	-0.162
Short	-0.311	-0.341	-0.281
Medium	0.094	0.064	0.125
Long	0.216	0.185	0.248
Age group	-0.049	-0.133	0.037
Inter-event interval	0.015	-0.007	0.036
Age $\times$ S	0.013	-0.018	0.043
Age $\times$ M	-0.010	-0.041	0.022
Age $\times$ L	-0.003	-0.034	0.028
S $\times$ IEI	-0.005	-0.035	0.024
M $\times$ IEI	0.036	0.005	0.065
L $\times$ IEI	-0.030	-0.061	-0.001
Age $\times$ IEI	-0.023	-0.045	-0.003
S $\times$ Age $\times$ IEI	-0.002	-0.032	0.028
M $\times$ Age $\times$ IEI	-0.019	-0.050	0.012
L $\times$ Age $\times$ IEI	0.021	-0.009	0.052

3.2 Experiment 2

Table 5 presents the model coefficients for the analysis of recall error in Experiment 2. This included the additional factor of inter-event interval and in the model no-interval was coded 1 and the 2 second interval condition was coded -1. Thus the coefficients associated with this factor reflect the change from grand mean (Intercept) associated with the no-interval condition (the analagous value for the 2 second condition can be found by multiplying the coefficient by -1).

4 Analysis of Confidence - Experiment 1

4.1 Confidence ratings

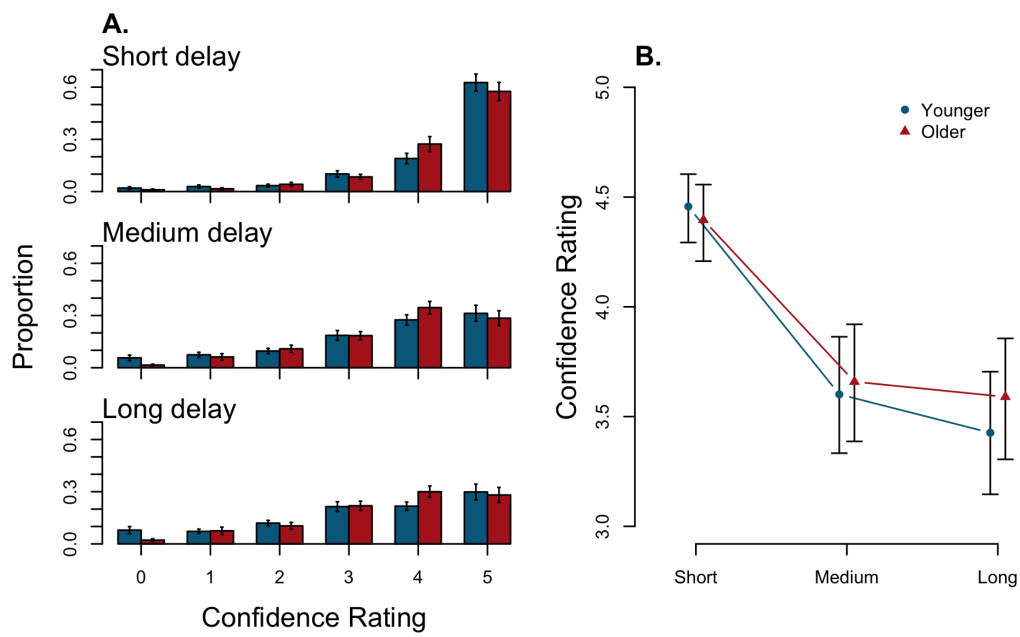

Figure 1. Experiment 1 (A.) distribution of confidence ratings by delay and age-group (error bars are bootstrapped standard errors). (B.) Posterior mean (and 95% CIs) confidence ratings by delay and group.

Confidence ratings were given on a scale from 0 (no memory at all) to 5 (best possible memory). To model confidence ratings as a function of age group and study-test delay, we used ordinal (cumulative) mixed effects regression. Specifically, we assume that confidence ratings arise due to an underlying latent distribution and response thresholds (Bürkner & Vuorre, 2019). This is more appropriate than models that assume a metric outcome variable and support better inferences regarding underlying effects (Liddell & Kruschke, 2018). The area in between adjacent thresholds (or the area above or below the extreme thresholds) determines the probability of producing a particular rating. Group or experimental factors influence ratings by shifting the underlying distribution towards higher or lower values. We assume a logistic latent distribution and use the `cumulative` family in `brms` to estimate this model (see Bürkner & Vuorre, 2019 for a tutorial).

The distribution of confidence ratings by group and condition are presented in Figure 1A. As described in the analysis section above, we modeled confidence using a cumulative mixed effects model (see Bürkner & Vuorre, 2019), which forms the basis of our contrasts. As shown in Figure 1B, the short delay was associated with higher ratings of confidence (4.426 [4.298, 4.540]), whereas ratings were considerably lower for the medium (3.630 [3.435, 3.816]) and long delays (3.508 [3.304, 3.701]). Contrasting the delays we find that confidence ratings clearly drop between the short and medium delays (0.796 [0.703, 0.891]) and the medium and long delays (0.122 [0.049, 0.194]), although the former drop is clearly learger (see Figure 1B).

Confidence ratings, on average, did not differ greatly between younger (3.828 [3.599, 4.053]) and older adults (3.881 [3.639, 4.106]; difference: -0.053 [-0.374, 0.278]). However, there was evidence of interaction between age and delay. As shown in Figure 1B, the change in confidence between the short delay and long delay was slightly steeper for the younger group (1.031 [0.879, 1.179]) relative to the older group (0.806 [0.674, 0.938]; difference: 0.225 [0.028, 0.422]). Nevertheless, the confidence ratings given in each delay condition did not clearly differ between the two groups (age difference for short: 0.061 [-0.170, 0.300]; medium: -0.058 [-0.430, 0.322]; long: -0.163 [-0.553, 0.229]).

4.2 Confidence and recall error

Next, to assess the relationship between confidence and recall error, we extended the mixed effects analysis of recall error, presented above, to include confidence rating. Specifically, we included confidence as a monotonic effect (see Bürkner & Charpentier, 2018), which allows us to model the magnitude of relationship between confidence and error without assuming equal change between each consecutive confidence rating. The results of this analysis are presented in Figure 2. Comparing the two figures we see that confidence clearly relates to expected recall error for both groups. There is some evidence of a three way interaction between age, delay, and confidence and this is attributable to somewhat greater spread of recall error by confidence ratings for older adults in the short and long conditions (see Figure 2B). However, the take away from the analysis of confidence is (1) that younger and older adults exhibit similar changes in confidence ratings across the three delay conditions and (2) there is a clear relationship between confidence and recall error for both groups, suggesting that both groups are able to gauge the accuracy of their memory for object location. We probe this further below with our mixture model, which attempts to distinguish guess responses from recall from memory.

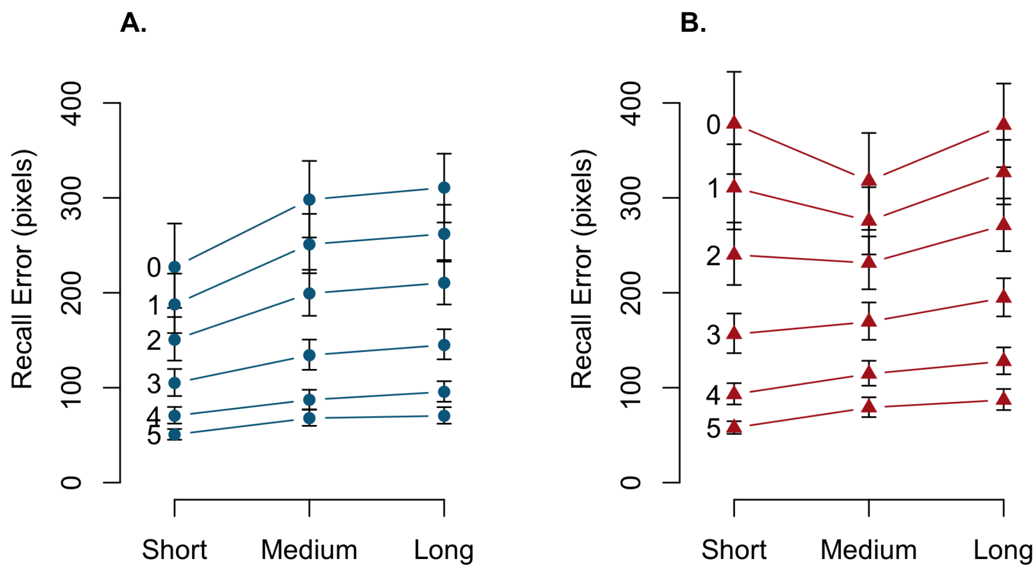

Figure 2. Recall error for different confidence ratings for (A.) younger and (B.) older adults. Posterior means and 95% credible intervals.

4.3 Confidence and the mixture parameters

We next considered how confidence relates to estimates of the m and σ parameters of our mixture model. Given the relatively low number of responses for the lowest confidence ratings of 0 and 1 (see Figure 1A), we decided to group confidence ratings into "low" (ratings 0-2) and "high" (3 and above) categories. We then included this factor as a predictor in the mixture model to allow the parameters to vary.

Estimates of m clearly differed for high and low confidence ratings for both younger (high: 0.951 [0.921, 0.974], low: 0.317 [0.218, 0.422]) and older adults (high: 0.866 [0.816, 0.907], low: 0.281 [0.179, 0.397]). Indeed, contrasting low and high confidence directly, we find that the difference is quite similar across groups (younger: 0.634 [0.533, 0.733], older: 0.585 [0.488, 0.673]). The precision of memory responses also varied reliably with confidence for both groups (for the younger group, high: 51.125 [47.844, 54.503], low: 66.460 [57.141, 76.526], difference: -15.335 [-25.565, -5.773] and for older adults, high: 58.638 [55.541, 61.947], low: 73.756 [62.838, 86.203], difference: -15.119 [-28.251, -3.538]).

When we consider the role of delay there does not appear to be any systematic change in the difference between high and low confidence estimates for m (for younger adults, short: 0.579 [0.399, 0.743], medium: 0.621 [0.473, 0.755], long: 0.703 [0.589, 0.804]; and for older adults, short: 0.643 [0.443, 0.808], medium: 0.501 [0.351, 0.641], long: 0.612 [0.500, 0.710]) or memory imprecision, σ (for younger adults, short: -7.326 [-23.645, 5.840], medium: -25.017 [-41.091, -9.754], long: -13.662 [-32.112, 3.319]; and for older adults, short: -20.886 [-45.951, -1.945], medium: -16.461 [-35.802, 0.884], long: -8.009 [-29.503, 10.805]). However, we should note that, at this level of analysis, the 95% credible intervals are rather broad and do not support particularly strong conclusions on the interaction between delay and confidence in relation to the mixture parameters. Nevertheless, the analysis of the mixture parameters as a function of confidence, averaged across delay, complements the analysis of recall error above; participants' confidence ratings carry information about the quality of information they are able to retrieve, predicts the likelihood of guessing, and, crucially, this appears to be equally true for the younger and older adults we assessed.

5 Combined Analysis of Experiments 1 and 2

As noted in the main manuscript, there was an important difference between our two experiments in the results for the difference in recall error between groups. Thus we decided to directly compare the results of the two experiments (1) to more formally assess how consistent findings were and (2) hopefully obtain more precise estimates for our crucial comparisons. Table 6 presents the results of this combined analysis for recall error. As can be seen in this table there is no main effect of experiment and, importantly, the age by experiment interaction contrast is not different from zero. In pixels, the age difference in the combined analysis amounts to -15.591 [-28.636, -2.183]. The model coefficients suggest that age differences do not vary systematically with delay and this is true for both experiments.

The only non-zero interaction contrast involving experiment was with delay. Recall error did not differ between the two experiments in the short (experiment 1 minus 2: -6.441 [-16.745, 3.833]), medium (4.599 [-10.632, 19.770]), or long delay conditions (9.954 [-6.574, 26.534]). The interaction arises as recall error is slightly lower in Experiment 1 at the short delay but slightly larger at the medium and long delays. Thus when we contrast the difference between experiments in the short delay condition with that in the medium (-11.039 [-19.670, -2.492]) and long (-16.395 [-26.208, -6.784]) conditions we find a difference. The difference between the experiments does not itself significantly differ between the medium and long delay conditions (-5.356 [-14.697, 3.980]). Despite this overall difference in the increase in recall error between the short and medium delays, the results seem generally consistent between experiments. Note that the tendency for Experiment 1 to have a slightly steeper change across delay is completely unexpected, given that in Experiment 2 half of the observations were spaced out such that more time elapsed between study and test.

Fitting the mixture model to the combined data sets we find that the overall probability of responding from memory is 0.831 [0.793, 0.866] for younger adults and 0.801 [0.761, 0.838] for older adults (difference: 0.030 [-0.023, 0.083]). Further, the difference between age groups does not differ greatly with delay (short: 0.017 [-0.013, 0.048]; medium: 0.024 [-0.043, 0.092]; long: 0.049 [-0.029, 0.125]; contrast short vs long: -0.031 [-0.092, 0.031]). In contrast there was a clear age difference in the imprecision of memory (younger: 52.451 [50.312, 54.711]; older: 59.645 [57.179, 62.268]; difference: -7.193 [-10.559, -3.922]). Further, the age groups differed at all delays (short: -5.332 [-8.388, -2.173]; medium: -7.890 [-12.370, -3.558]; long: -8.359 [-13.052, -3.803]; contrast short vs long: 3.027 [-1.061, 7.200]). Thus, even in this combined analysis there is no strong evidence that age differences in the imprecision of memory increase with delay, at least across the coarse levels of "short", "medium", and "long".

Table 6
Results of cross-experiment analysis of recall error.

	Estimate	Q2.5	Q97.5
Intercept	-0.240	-0.302	-0.178
Short	-0.360	-0.381	-0.339
Medium	0.109	0.087	0.130
Long	0.251	0.229	0.273
Age group	-0.073	-0.134	-0.010
Experiment	0.005	-0.057	0.067
Age × S	0.001	-0.021	0.023
Age × M	0.005	-0.017	0.027
Age × L	-0.006	-0.028	0.016
S × Exp	-0.048	-0.070	-0.027
M × Exp	0.015	-0.007	0.037
L × Exp	0.033	0.011	0.055
Age × Exp	-0.023	-0.083	0.037
S × Age × Exp	-0.012	-0.034	0.010
M × Age × Exp	0.014	-0.008	0.036
L × Age × Exp	-0.002	-0.025	0.020

6 Mis-binding Responses

The mixture modeling approach here attempts to separate recall responses that are centered on the presented value from responses that are essentially random. However, it is possible that some responses that appear random with respect to the target value actually reflect a *mis-binding* error, in which the value associated with another cue item is recalled erroneously. For example, having studied a tennis ball to the top-right of the study area and a lamp to the bottom-right, a participant might wrongly place the tennis ball near the bottom-right position that should be associated with the lamp. This mis-binding tendency might be especially strong when items are studied in close proximity to each other, as temporal order is a strong cue for recall (see, e.g., Polyn, Norman, & Kahana, 2009) and may lead to confusions between proximal items, especially for older adults (Campbell, Trelle, & Hasher, 2014).

Thus if such mis-binding responses are present they would be categorized by our mixture model as guesses (i.e., complete memory failure). To explore the possible influence of mis-binding responses we calculated recall error relative to items studied immediately prior to and after the target item. We then compared the distribution of recall error o that expected if the target response was random with respect to the preceding or succeding non-target location. However, in doing this it is important to note that the distribution Euclidean distance between independently sampled locations is not uniform. The fact that our locations were sampled on an annulus complicates things further. Therefore, we simuated 1 million differences between independently sampled locations on the annulus used in the present experiments. The distribution is shown in Figure 3.

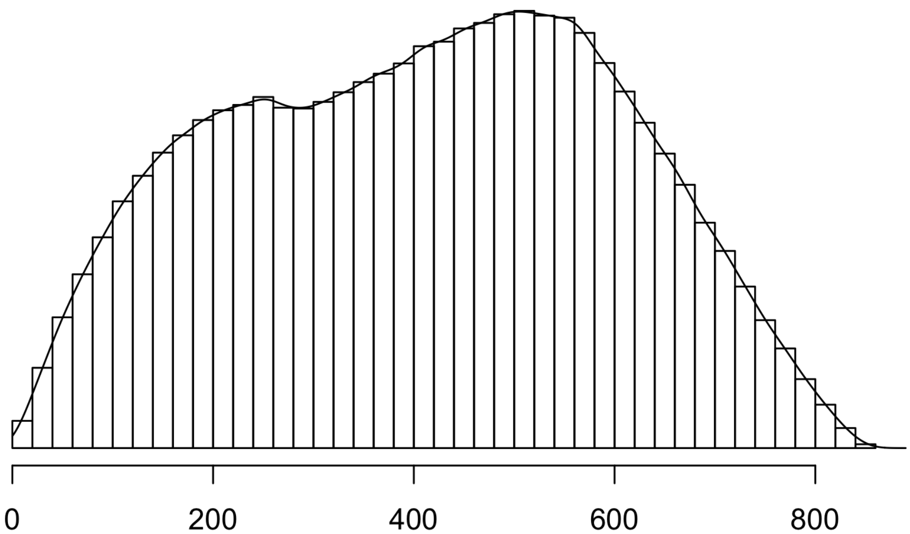

Figure 3. Distribution of Euclidean distance between 1 million randomly sampled location pairs using the same annulus dimensions as those used in the present experiments.

Figure 4 displays recall error relative to the item studied immediately before or after the target image for a given response. Both histograms and qq-plots show that these distributions are what we would expect if the response to a target is unrelated to these non-target items studied nearby. In Figure 4 the data are combined across Experiments 1 and 2 but the figures look very similar when the data is split by experiment. Splitting the non-targets into those appearing only before or those appearing only after the target item also does not change things. We also looked at non-target items studied immediately before the target item was tested and there was no evidence of deviation from the random histogram in Figure 3.

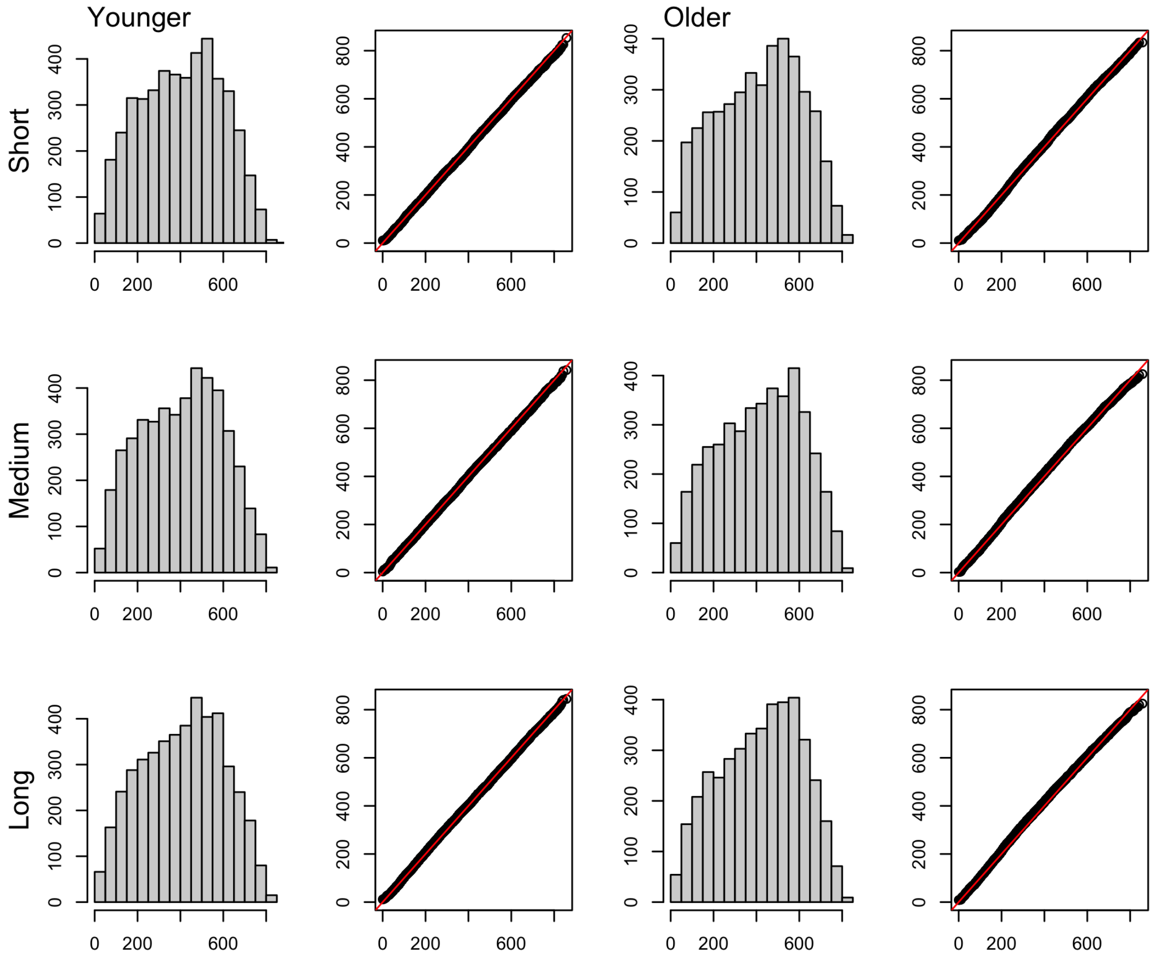

Figure 4. Histograms of recall error relative to the item studied immediately before or immediately after the relevant target item and qq-plots comparing the quantiles of the observed distribution to the simulated distribution above. Data from both experiments are combined but similar patterns are observed in both experiments separately.

7 Full Lag Analysis

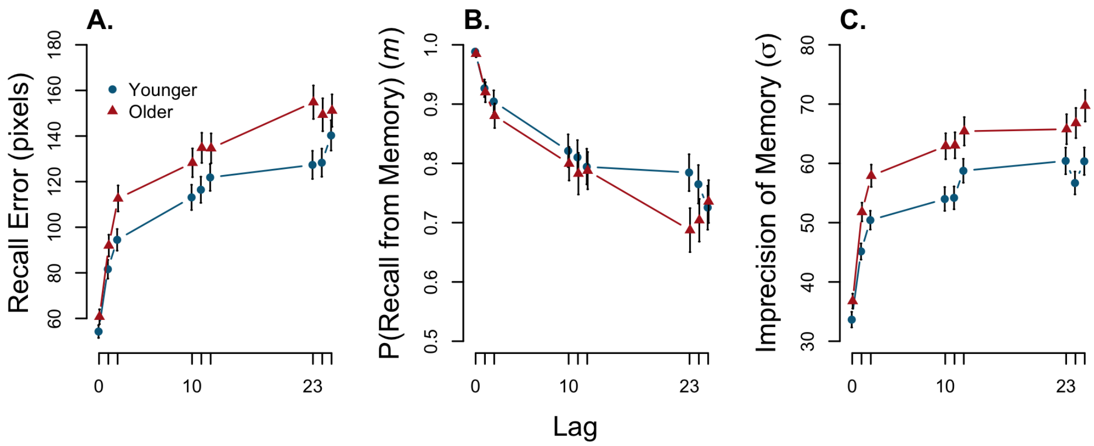

Figure 5. Results across both experiments split by study-test lag for (A.) recall error, (B.) the probability that recall was from memory, relative to a random guess, and (C.) the standard deviation, or imprecision, of recall.

In the main manuscript we presented the results for the first three study-test lags, which were grouped into the short delay. Figure 5 presents recall error, probability of recall from memory, and the imprecision of memory as a function of lag.

Table 7 presents the results of the linear mixed effects model for recall error. Contrasts for the first three lags are presented in the main manuscript. The magnitude of the difference between younger and older adults did not differ much between lags 10 (-15.177 [-31.334, 0.817]), 11 (-18.387 [-35.472, -1.513]), and 12 (-12.789 [-29.927, 3.921]), although the broad confidence intervals when splitting the data by lag means that sometimes the differences were not credibly different from zero (contrasting lags directly, 10 vs 11: 3.210 [-11.327, 17.610]; 10 vs 12: -2.388 [-17.389, 12.368]; 11 vs 12: -5.598 [-20.537, 9.456]).

For the longest lags, the estimated age difference was numerically larger for lag 23 (-27.488 [-46.324, -9.149]), relative to lag 24 (-21.029 [-39.087, -2.980]) and 25 (-10.899 [-30.108, 7.609]). However, again due to the noisy estimates when splitting the data by lag, non of these differences were credibly non-zero (23 vs 24: -6.460 [-22.719, 9.658]; 23 vs 25: -16.590 [-33.399, 0.539]; 24 vs 25: -10.130 [-26.945, 6.395]).

Figure 5B shows probability of recall from memory by lag. There was no evidence of a difference between groups at lags 10 (0.022 [-0.059, 0.101]), 11 (0.028 [-0.063, 0.119]), and 12 (0.007 [-0.081, 0.094]) and there were no differences between these (10 vs 11: -0.006 [-0.091, 0.079]; 10 vs 12: 0.015 [-0.067, 0.100]; 11 vs 12: 0.021 [-0.069, 0.112]). At the longest study-test delay there was a non-zero age difference at lag 23 (0.097 [0.001, 0.194]), but not lags 24 (0.061 [-0.039, 0.157]) or 25 (-0.010 [-0.113, 0.094]). Contrasting these age differences we find that lags 23 and 25 differ (0.107 [0.010, 0.204]), whereas the credible intervals for the other contrasts include zero (23 vs 24: 0.037 [-0.057, 0.130]; 24 vs 25: 0.071 [-0.029,

Table 7
Results of fine grained analysis of recall error by lag

	Estimate	Q2.5	Q97.5
Intercept	-0.241	-0.302	-0.180
Lag 0	-0.699	-0.742	-0.657
Lag 1	-0.278	-0.321	-0.236
Lag 2	-0.095	-0.138	-0.051
Lag 10	0.071	0.029	0.115
Lag 11	0.114	0.071	0.159
Lag 12	0.138	0.095	0.182
Lag 23	0.242	0.198	0.286
Lag 24	0.225	0.182	0.269
Lag 25	0.281	0.237	0.325
Age group	-0.073	-0.134	-0.015
Age by Lag 0	0.015	-0.026	0.057
Age by Lag 1	0.010	-0.031	0.053
Age by Lag 2	-0.020	-0.064	0.023
Age by Lag 10	0.004	-0.040	0.048
Age by Lag 11	-0.008	-0.051	0.034
Age by Lag 12	0.018	-0.026	0.061
Age by Lag 23	-0.037	-0.080	0.006
Age by Lag 24	-0.012	-0.057	0.030
Age by Lag 25	0.030	-0.014	0.074

0.167]). As this difference is related to the difference between the groups reducing with a longer study-test delay (see Figure 5B) we suspect this is attributable to measurement noise and do not attempt to interpret this.

The imprecision of memory as a function of lag is show in Figure 5C. The size of the age difference in this parameter was generally consistent across the different lags (lag 10: -8.934 [-14.809, -3.051]; lag 11: -8.840 [-14.721, -3.127]; lag 12: -6.667 [-12.923, -0.469]; lag 23: -5.354 [-12.134, 1.485]; lag 24: -10.112 [-16.592, -3.701]; lag 25: -9.380 [-16.393, -2.452]). There are no differences between the lags that were grouped into the medium delay (10 vs 11: -0.095 [-7.455, 7.099]; 10 vs 12: -2.267 [-9.686, 5.324]; 11 vs 12: -2.172 [-9.647, 5.432]) nor between those grouped into the long delay (23 vs 24: 4.758 [-3.342, 13.083]; 24 vs 25: -0.732 [-9.026, 7.609]).

8 Categorical Responding

In delayed estimation tasks, such as those requiring recall on a circular space (e.g., color wheel, orientation), it is well established that participants show categorical biases in their responding (Bae, Olkkonen, Allred, & Flombaum, 2015; Hardman, Vergauwe, & Ricker, 2017; Pratte, Park, Rademaker, & Tong, 2017). For example, certain prototypical colors are over-represented in participants responses, suggesting that coarse category representations contribute to task performance along with more precise, continuous representations. A similar pattern has been noted in tasks where participants are required to recall location (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, Corrigan, & Crawford, 2004). In particular participants in these experiments showed a bias towards diagonal locations (e.g., top-left, bottom-right).

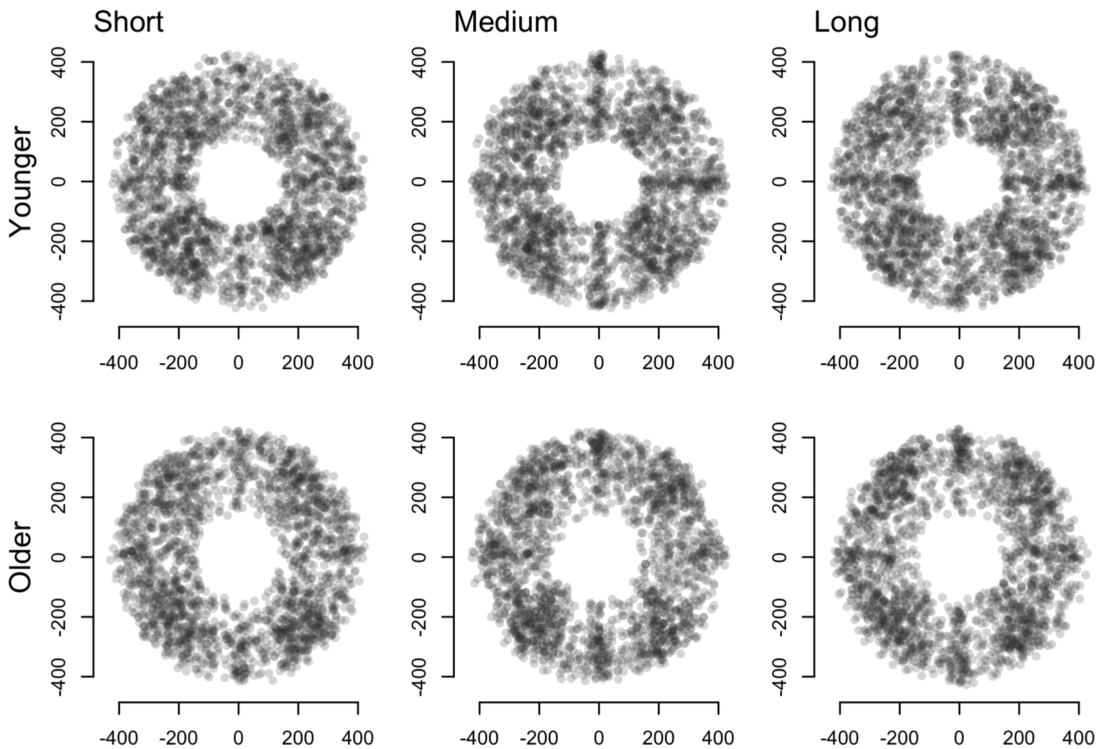

Figure 6. Recalled (x,y) pixel coordinates by group and delay across both experiments. The presented locations were uniformly distributed on the annulus.

Three plots suggest that participants are using categorical representations in the present task. Figure 6 plots the recalled locations across both experiments. The presented locations were uniformly distributed on the annulus, therefore, darker areas in this plot suggest these locations were utilized more often during recall relative to other locations.

The bias exhibited in Figure 6 appears to be primarily a angular one, where vertical, horizontal, and diagonal locations are over-represented. This is seen more clearly in Figure 7, which plots histograms of the angle of the recalled location. Figure 8 presents the scatterplot of the recalled by presented location angle. The use of categorical representations can be seen in the clustering along the diagonal of this plot. There is also some indication of categorical guessing (i.e., responses that show no relation to the presented angle) in the form of the "bands" of points parallel to the x-axis (see Rouder, Thiele, Province, & Cusumano, 2014).

To try and disentangle the contribution of categorical and continuous representations

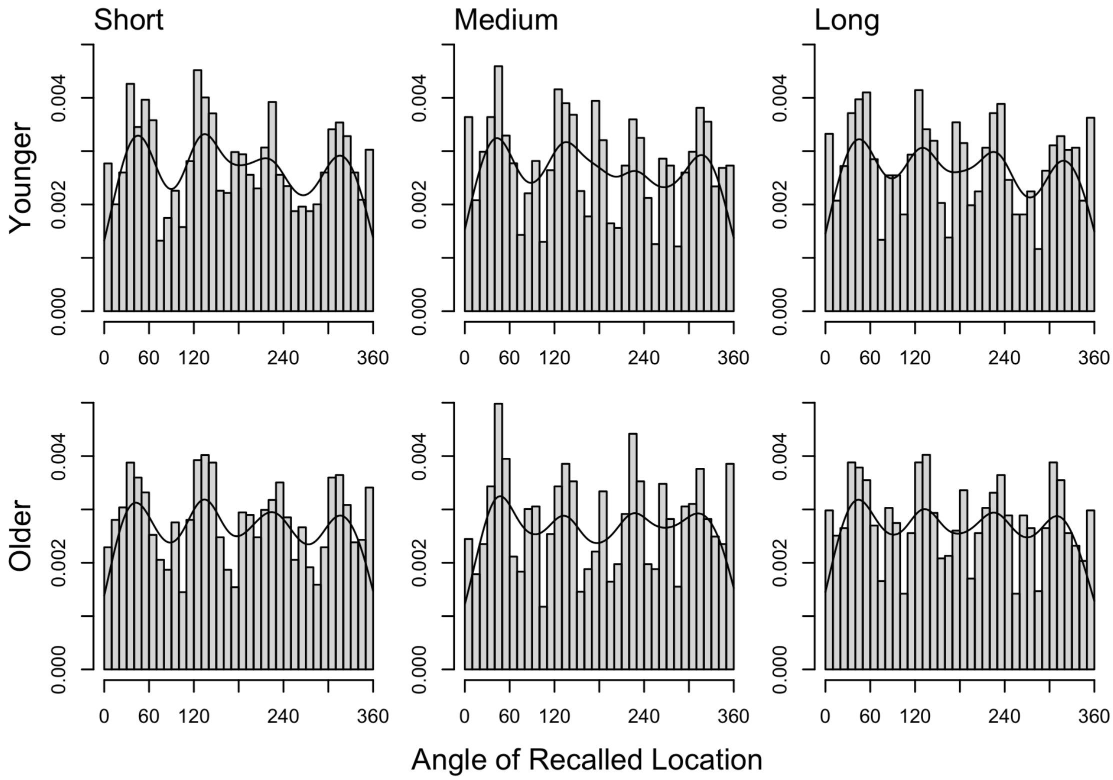

Figure 7. Histograms of recalled angle by group and delay across both experiments.

to the performance of our task we used the categorical-continuous model of Hardman et al. (2017; as implemented in the `CatContModel` package for `R`; Hardman, 2017). This model has three main parameters: P^M the probability of recall from memory, P^O the probability that what is recalled from memory is a continuous representation (versus a categorical representation), and σ^O, which is the imprecision of continuous memory (the standard deviation of a Von Mises distribution). The model has other parameters which are involved in estimating the number and location of categories (see Hardman et al., 2017 for a full description) but our primary interest is on these three parameters. We fit the model separately for younger and older groups using the `CatContModel` package with the default priors. The parameter estimates below are based on 10000 post-warm-up posterior samples per group. All age differences in parameter estimates below are given such that positive differences reflect larger parameter values for the younger group.

There was no overall age difference for P^M (-0.008 [-0.048, 0.032]) and differences were small and not different from zero in all three delay conditions (short: 0.003 [-0.019, 0.026]; medium: -0.015 [-0.065, 0.033]; long: -0.012 [-0.075, 0.050]; see Figure 9A).

For continuous versus categorical responding, P^O, there was a difference of 0.056 [-0.011, 0.122], which was not significantly different from zero. Age differences in continuous recall increase across the three delays (short: -0.006 [-0.054, 0.040]; medium: 0.041 [-0.058, 0.141]; long: 0.133 [0.017, 0.244]). As shown in Figure 9B both groups show an increase in

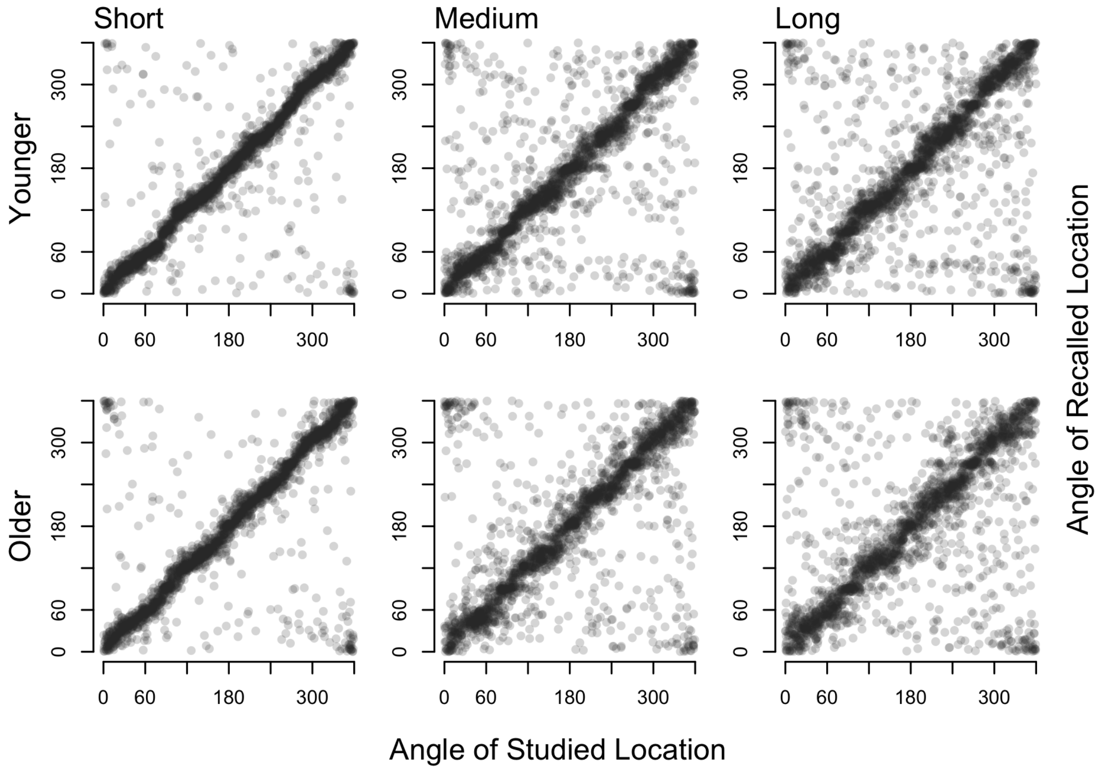

Figure 8. Presented and recalled location angles by group and delay across both experiments.

categorical responding between the short and medium delays where as older adults show a further increase between the medium and long delay.

Finally, for the imprecision of continuous recall, σ^O, older adults were slightly more imprecise overall but this difference was not credibly greater than zero (-0.611 [-1.519, 0.301]). As shown in Figure 9C the two groups differ slightly in the short (-0.926 [-1.720, -0.151]) and medium (-0.928 [-2.192, 0.313]) conditions and are then essentially identical in the long condition (0.022 [-1.500, 1.484]).

In summary, this analysis suggests that there is quite a rapid increase in the likelihood of recalling the general location of an object, relative to its precise location, with increasing delay. For younger adults the increase in categorical recall is mainly seen between the short and medium delays, whereas for older adults this continued on to the long delay. Thus it is likely that some of the age differences in precision seen in the full mixture analysis of recall location, presented in the main manuscript, is due to reliance on more categorical representations (the “gist” of where an object was presented), relative to a fine grained representation of location. Nevertheless, a limitation of this analysis is that we disgarded potentially useful information in converting recalled locations into angles. Thus a focus for further work could be developing mixture models, which incorporate categorical and continuous responses for two-dimensional responses.

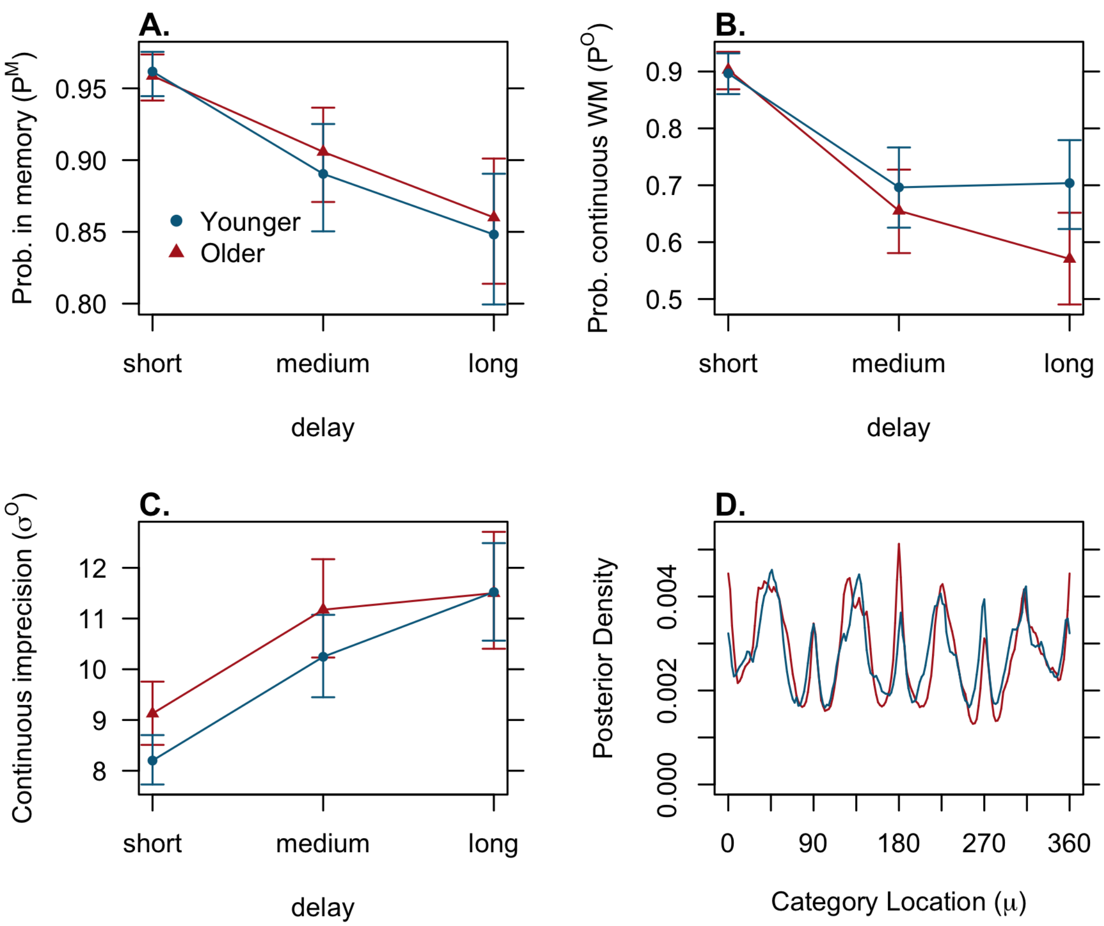

Figure 9. Results of the categorical-continuous model fit to angle of recalled location (posterior means and 95% credible intervals). (A) Probability of recall from memory, (B) imprecision of continuous memory representations, (C) probability that memory recall is continuous vs categorical, (D) estimated locations of categories.

9 References

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

Benton, A. L., & Hamsher, K. deS. (1976). *Multilingual aphasia examination*. University of Iowa Press.

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

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

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

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

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

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

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

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

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

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.

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.