# Recognition Task: Measures of signal detection and criterion shift in groups of younger and older adults

During peer review of this paper, we were asked to examine age group differences in more classic measures of memory performance, for example the signal detection measure *d’* and the criterion shift measure *c*. *d’* is a measure for recognition discrimination, as it takes into account response bias (e.g., Fraundorf et al., 2019). *c* compares criterion shifts by also taking into account response bias. These results of these analyses aligned with our main analysis and presented below.

We calculated *d’* as a difference score between z-transformed hits and false alarms, using the formula *qnorm(HR) - qnorm(FAR)* in R.[[1]](#footnote-1) *c* was computed using the formula *– ( (qnorm(HR)+qnorm(FAR)) /2;* see Fraundorf et al., 2019; Kantner & Lindsay, 2012; Stanislav and Todorov, 1999*)*. We computed these measures by only taking into account new items, not lures, as is common in the false memory literature (e.g., Brainerd et al., 2008; Chang & Brainerd, 2022; Ghetti et al., 2002). False alarm and hit rates equal to zero or one were replaced by 0.01 and 0.99, respectively.

Table S1 shows hits and false alarms, as well as *d’* and *c* in all subjects, split out by low- and high- confidence judgments. in low-confidence judgments, younger and older adults had comparable *d’* and *c* scores, Welch’s *t*(61.90) = -0.64, *p* = .52, suggesting that both age groups discriminated equally successfully between old and new words. However, in high-confidence judgments, *d’* and *c* scores were much lower for older compared to younger adults, Welch’s *t*(110.84) = 2.69, *p* < .01. Hence, in high-confidence judgments, older adults discriminated less successfully between old and new items. In line, the criterion shift measure *c* was, on average, lower for older (*M* = -0.44) than younger adults (*M* = 0.01), Welch’s *t*(82.76) = -7.65, *p* < .001, indicating that older adults showed a more liberal response criterion to endorse new, unseen, items as old.

**Table S1**

*Average Group Performance in the Recognition Memory Task*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Hits | False Alarms | *d’* | *c* |
|  | Low Confidence | | | |
| Psychology Students | 0.50 | 0.29 | 0.63 | -0.31 |
| Prolific Sample | 0.52 | 0.26 | 0.74 | -0.37 |
| **All Younger Adults** | **0.51** | **0.27** | **0.68** | **-0.34** |
| **Older Adults** | **0.65** | **0.38** | **0.78** | **-0.39** |
|  | High Confidence | | | |
| Psychology Students | 0.88 | 0.28 | 2.09 | -1.05 |
| Prolific Sample | 0.86 | 0.19 | 2.27 | -1.13 |
| **All Younger Adults** | **0.87** | **0.23** | **2.18** | **-1.09** |
| **Older Adults** | **0.92** | **0.41** | **1.80** | **-0.90** |

*Note.* We report the by-group performance for ease of comparison between the samples; all statistical analyses were run on the combined data from all groups using a scaled continuous variable for age.

# Recognition Task: Proportions of Responses Allocated by Response Bin

Table S2 shows count values and global proportions of responses allocated by response bin to lure, old, and new items in younger and older adults. Proportions were initially calculated separately for each participant, by dividing, for each subject, their number of responses allocated per condition-response-bin by their total number of responses. For Table S2, the resulting values were aggregated.

**Table S2**

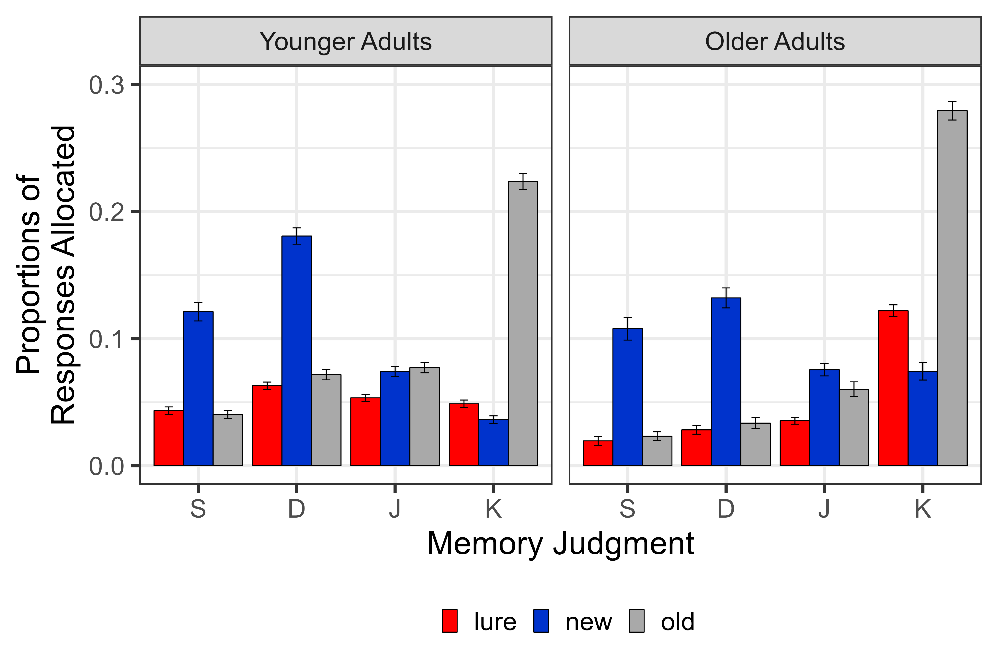
*Counts and Proportions of Responses Issued by Groups of Younger and Older Adults per Confidence Bin*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Psychology  Students | Prolific  Sample | All Younger  Adults | Older  Adults |
| High  Confidence  New | lure | 246 (0.05) | 195 (0.04) | 441 (0.04) | 66 (0.01) |
| new | 692 (0.13) | 583 (0.11) | 1275 (0.12) | 583 (0.11) |
| old | 178 (0.03) | 180 (0.03) | 358 (0.03) | 122 (0.02) |
| Low  Confidence  New | lure | 310 (0.06) | 344 (0.06) | 654 (0.06) | 127 (0.02) |
| new | 895 (0.17) | 1039 (0.19) | 1934 (0.18) | 684 (0.13) |
| old | 350 (0.06) | 408 (0.08) | 758 (0.07) | 167 (0.03) |
| Low  Confidence  Old | lure | 219 (0.04) | 320 (0.06) | 539 (0.05) | 181 (0.04) |
| new | 371 (0.07) | 385 (0.07) | 756 (0.07) | 399 (0.08) |
| old | 357 (0.07) | 447 (0.08) | 804 (0.07) | 307 (0.06) |
| High  Confidence  Old | lure | 296 (0.05) | 209 (0.04) | 505 (0.05) | 656 (0.13) |
| new | 212 (0.04) | 130 (0.02) | 342 (0.03) | 401 (0.08) |
| old | 1289 (0.24) | 1106 (0.21) | 2395 (0.22) | 1469 (0.28) |
| **Sum** |  | **5415 (1.00)** | **5346 (1.00)** | **10761 (1.00)** | **5162 (1.00)** |

In order to examine potential age group differences in proportions of memory responses, we submitted the by-subject proportional responses to a 2 (younger, older) \* 3 (lure, new, old) \* 4 (sure new, maybe new, maybe old, sure old) ANOVA (see Figure S1 below).

**Figure S1**

Proportions of responses allocated per response bin



*Note*. S = Sure New, D = Maybe New, J = Maybe Old, K = Sure Old.

The ANOVA showed a significant three-way interaction between age group, condition and response bin, *F*(3, 315) = 4.93, *p* < .01. Follow-up ANOVAs that split the data by response bins showed significant interactions between condition and age group only for “maybe old”, *F*(2, 268) = 6.76, *p* < .01, and “sure old”, *F*(2, 272) = 7.52, *p* < .01 responses. We focus on these two here.

For “maybe old” responses, follow-up t-tests using the Bonferroni correction showed a significant age effect for lures, *p* < .01, but not for new or old items (*p* = .32 and *p* = .09, respectively), indicating that older adults allocated fewer “maybe old” responses to lures than younger adults did (EMMs of 0.03 and 0.05, respectively; see Figure S1). The age effect for lures also came out in an analysis that controlled for response bias, i.e., when we subtracted proportions of “maybe old” responses allocated to new items from those allocated to lures: Older adults showed significantly reduced difference scores, *p* < .001, indicating that they allocated fewer low-confidence “maybe old” responses to lures than younger adults did.

For “sure old” responses, follow-up t-tests demonstrated significant age effects for all three conditions, i.e., lures, new items, and old items (EMMs of 0.11, 0.06 and 0.28 in older adults, and 0.05, 0.03 and 0.22 in younger adults, all *p*’s < .001), suggesting that older adults allocated significantly more high-confidence “old” responses across the board. However, when corrected for response bias (i.e., when subtracting by-subject proportions of “sure old” responses to new items), only the age group effect for lures prevailed: Older allocated more high-confidence “old” judgments to lures than younger adults, *p* < .001, but they did not differ from younger adults in how much high-confidence “old” judgments they allocated to previously seen, old, items (*p* = .11).

Taken together, this analysis shows an age-related increase in subjective confidence for false memory judgments: In low-confidence judgments, older adults allocated *fewer* “old” judgments to lures than younger adults. In high-confidence responses, older adults allocated *more* “old” judgments to lures. These findings confirm and extend our main analysis by showing that older adults are more prone to high-confidence memory intrusions (e.g., Dodson et al., 2007).

# Age Differences in Reading and Recognition Rates Based on Education

Our main analysis controlled for individual differences in education. Here, we also investigate whether there were direct interactions between age and educational attainment in reading and recognition rates.

*Age Differences in Reading Rates Based on Education*. We ran models identical to those reported in the section on self-paced reading, except that we additionally included a fixed effect for the interaction between continuous age and education.[[2]](#footnote-2) Education was included as a factorial predictor, with three levels: No higher-level education, high school degree, or university degree (see Table 1). Education was Helmert-coded and included two contrasts: *High-School Degree vs No Higher* *Degree*, and *University Degree vs High School Degree*. The interaction between age and education remained non-significant in all models (all *p*’s > .80); neither did any of the models show a significant three-way interaction between predictability, age, and education (all *p*’s > .50) Hence, the age effects in our analysis were unlikely driven by individual differences in education.

*Age Differences in Recognition Rates Based on Education*. To investigate the effects of education and aging on recognition memory, we ran follow-up models that specified the four-way interaction between word type, age, confidence, and education, including all two- and three-way interactions between the predictors and all main effects. However, there were persistent convergence issues with the four-way interaction model, even when we simplified the random structure and included only a random intercept for subjects. To resolve these issues, we ran three follow-up models on subsets of our data that aimed to investigate if the findings reported in the main analysis held up when we split participants based on their education background. In other words, we ran three separate models for participants with no higher-level degree, participants with a high-school degree and those with a university degree. Much like in our main analysis, each model included the three-way interaction between word type, age, and confidence.

The results in all three models were identical to those reported in the main analysis, that is, across subject groups, we substantiated the three-way interactions for true and false recognition memory, with one exception: The model for participants with no higher-level education (*n* = 22 subjects) showed no three-way interactions. However, that same model did show all relevant two-way interactions for the contrasts *lure vs new* and *old vs new*. Therefore, we believe these results are more quantitative than qualitative in nature and may result from low power. Based on these analysis, we have no evidence to believe that the findings reported in our main analysis were modulated by educational attainment.

1. The qnorm() function is the R-equivalent to the NORMSINV function in MS Excel. It transforms values into p values and looks up their z value in the normal distribution. [↑](#footnote-ref-1)
2. We also ran models on all words in the critical region that included the three-way interaction between predictability, age, and education. [↑](#footnote-ref-2)