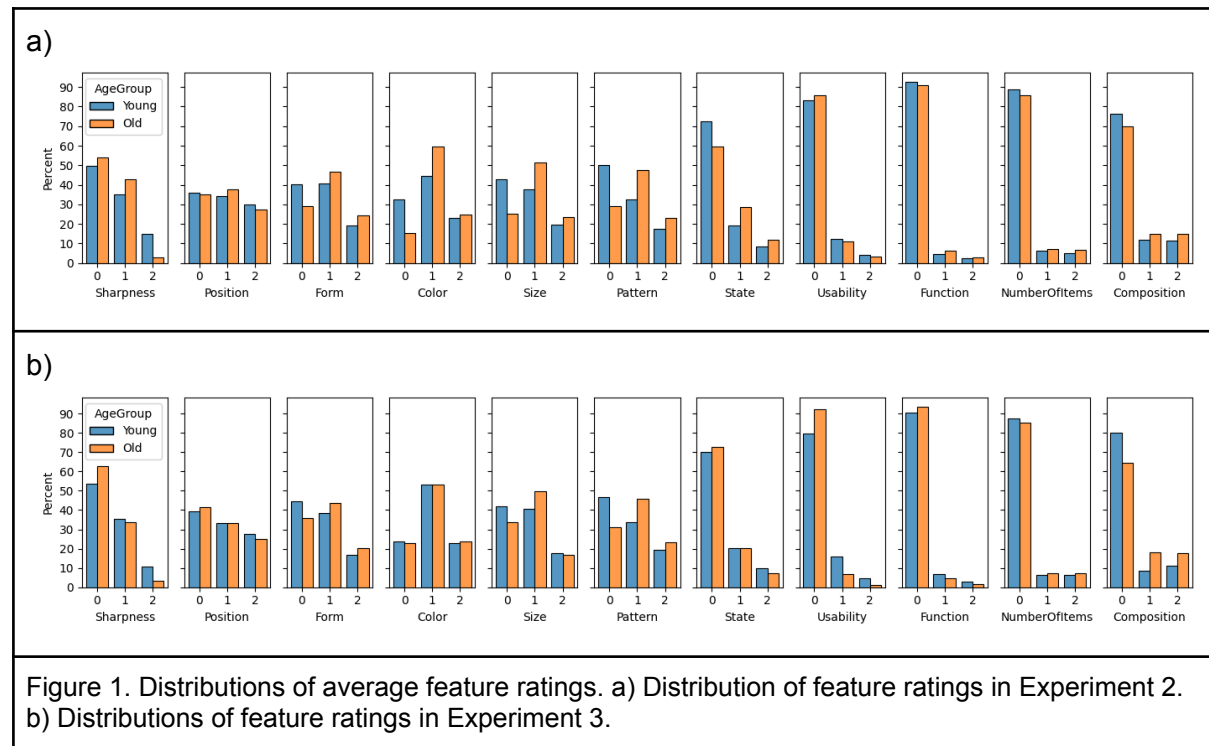


Different object features shape memory discrimination in younger and older adults

Supplementary Results.....	2
The distribution of feature difference ratings.....	2
Correlations between average feature ratings in Experiment 2 and Experiment 3.....	2
Correlations between average feature ratings, the general dissimilarity judgement, and the similarity bins.....	3
Delay between the study and the test phase in the MST.....	3
Linear mixed effects models with item intercepts.....	4
Supplementary Materials.....	6
Feature definitions and demonstration.....	6
Sensitivity Power Analysis and Sample Size Estimation for Future Research.....	11
References.....	12

Supplementary Results

The distribution of feature difference ratings



Correlations between average feature ratings in Experiment 2 and Experiment 3

	Cross-Age Group		Older Adults		Younger Adults	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Sharpness	0.94	< .001	0.89	< .001	0.91	< .001
Position	0.99	< .001	0.98	< .001	0.98	< .001
Form	0.97	< .001	0.96	< .001	0.96	< .001
Color	0.98	< .001	0.95	< .001	0.95	< .001
Size	0.97	< .001	0.96	< .001	0.95	< .001
Pattern	0.98	< .001	0.94	< .001	0.96	< .001
State	0.92	< .001	0.87	< .001	0.87	< .001
Usability	0.89	< .001	0.8	< .001	0.79	< .001
Function	0.82	< .001	0.77	< .001	0.8	< .001
Number Of Items	0.99	< .001	0.99	< .001	0.97	< .001
Composition	0.98	< .001	0.95	< .001	0.98	< .001
Diff	0.97	< .001	0.93	< .001	0.96	< .001

Correlations between average feature ratings, the general dissimilarity judgement, and the similarity bins

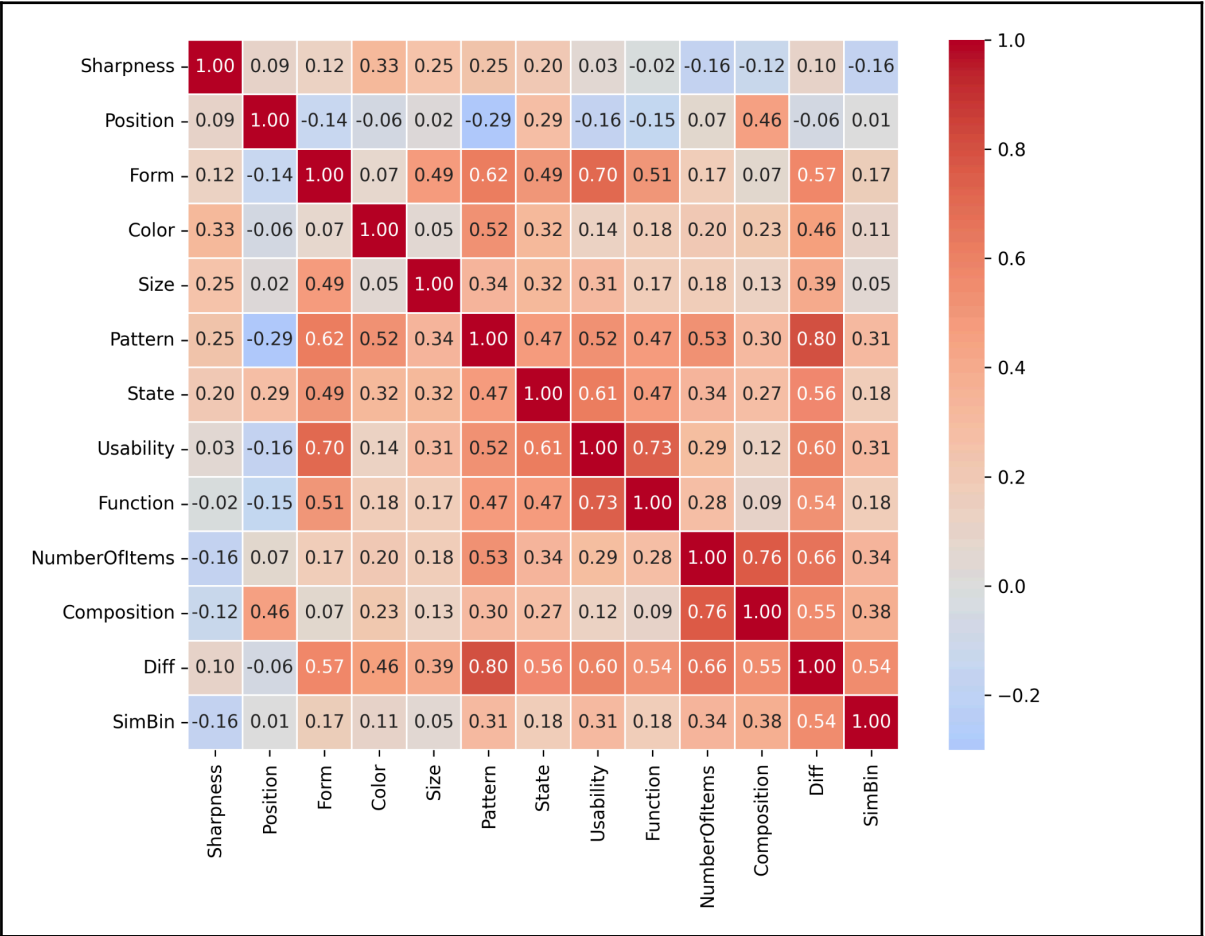


Figure 2. Correlation matrix of aggregated feature ratings, the general dissimilarity judgement, and the similarity bins (Lacy et al., 2011; S. M. Stark et al., 2019)

Delay between the study and the test phase in the MST

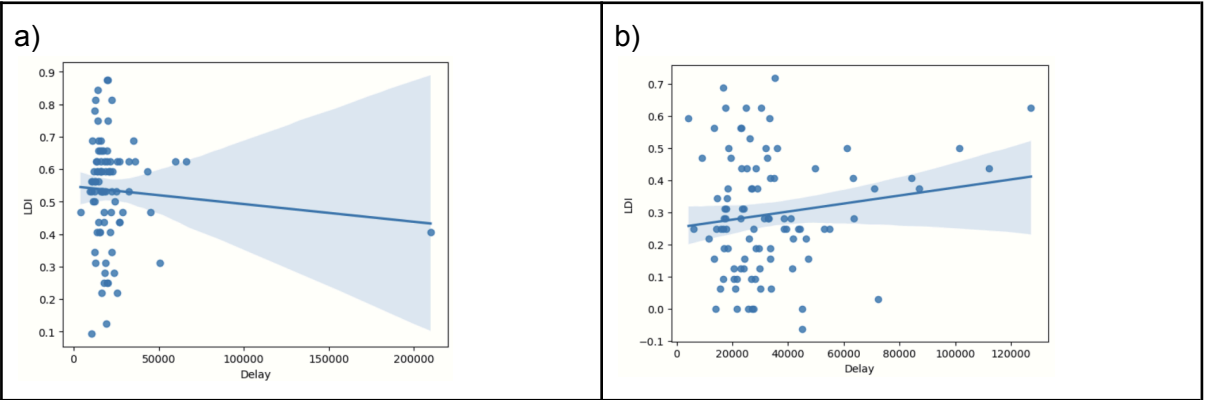
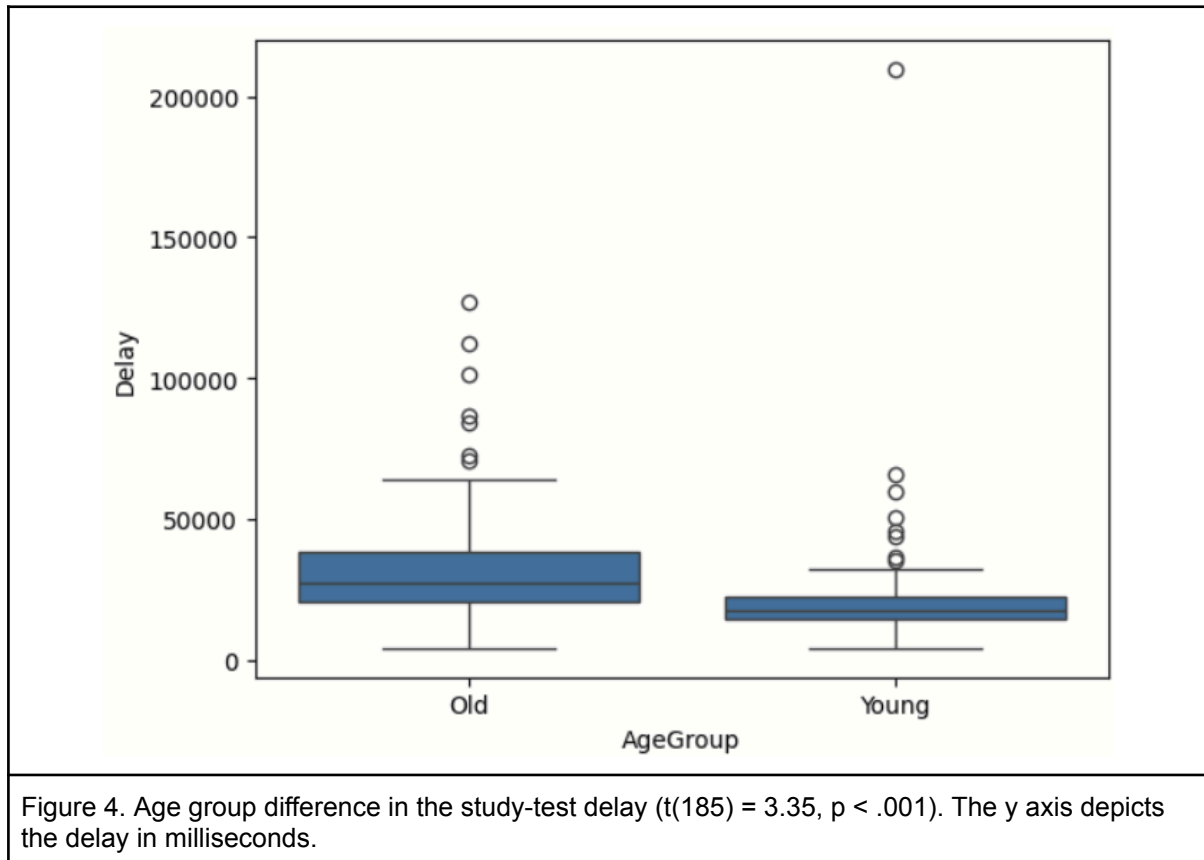


Figure 3. Non-significant correlations between the study-test delay (in ms) and LDI a) Scatter plot showing younger participants ($r = -0.08$, $p = 0.47$). b) Scatter plot showing older participants ($r = 0.15$, $p = 0.15$).



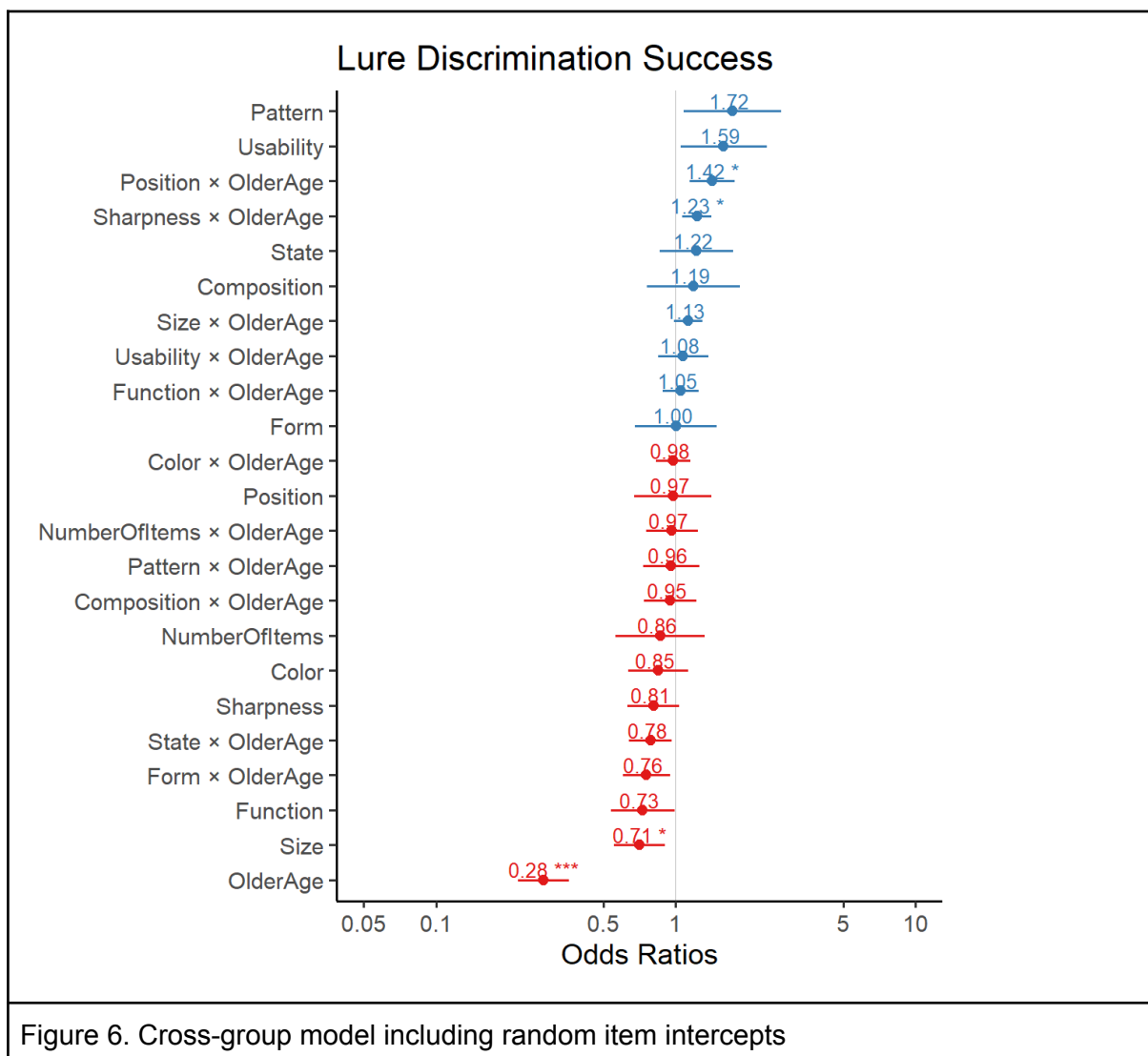
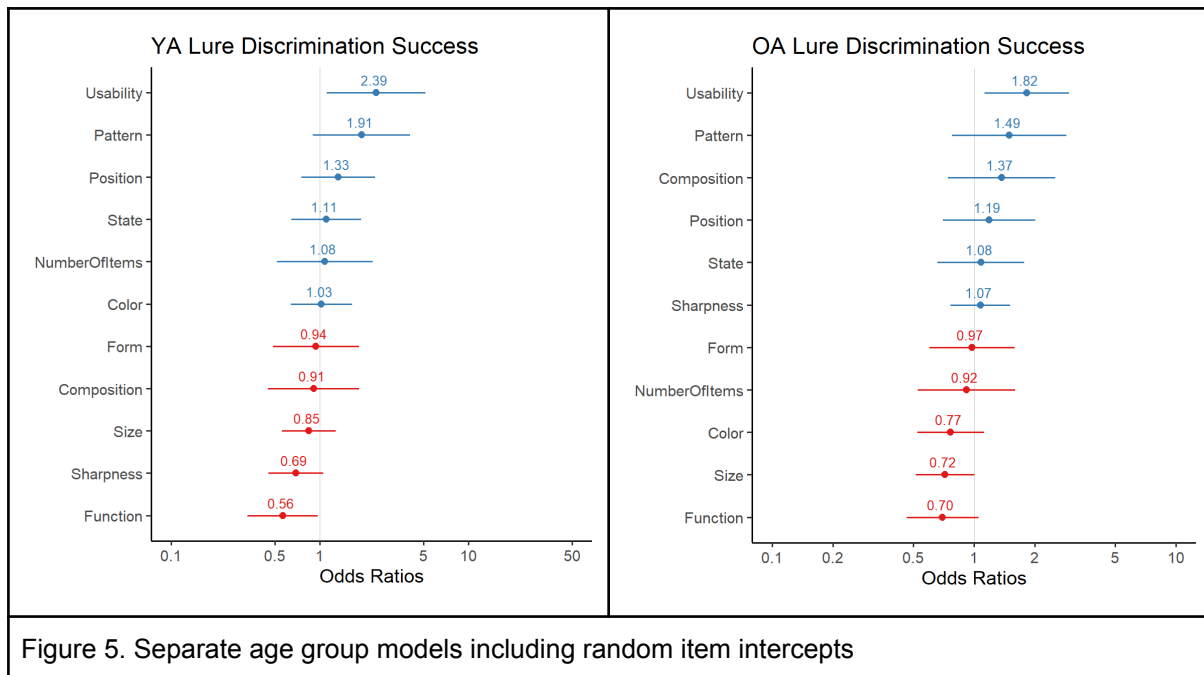
Linear mixed effects models with item intercepts

Model specification for separate age group models:

*Success ~ 1 + Sharpness + Position + Form + Color + Size + Pattern + State + Usability +
Function + NumberOfItems + Composition + (1 | Participant) + (1 | Stimulus)*

Model specification for the model with cross-group aggregated feature ratings as predictors:

*Success ~ 1 + Sharpness + Position + Form + Color + Size + Pattern + State + Usability +
Function + NumberOfItems + Composition + Sharpness:OlderAge + Position:OlderAge +
Form:OlderAge + Color:OlderAge + Size:OlderAge + Pattern:OlderAge + State:OlderAge +
Usability:OlderAge + Function:OlderAge + NumberOfItems:OlderAge +
Composition:OlderAge + OlderAge + (1 | Participant) + (1 | Stimulus)*



Supplementary Materials

Feature definitions and demonstration

The text below is translated from Hungarian. The image stimuli were selected from the MST stimuli set, courtesy of Craig Stark, University of California, Irvine (Stark et al., 2019).



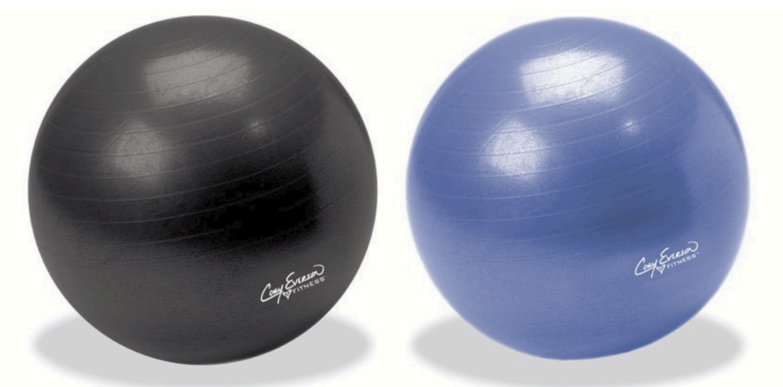
Sharpness: This feature refers to the image quality, the sharpness of the images.



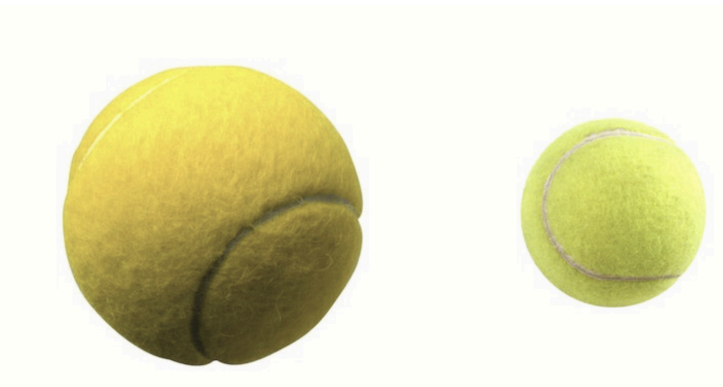
The above images differ in their **Position**.



The images above differ in **Form**.



The objects above differ in **Color**.



The images above differ in the following feature: **Size**.



The images above differ in **Within-Item Pattern**.

The **State** feature refers to the status relevant to the function. For example, the object is currently open or closed, switched on or off, whole or sliced (e.g., bread, fruit).



The objects differ in **State**, because one is open and the other is closed.



The objects differ in **State**. The green light means *GO*, and the red light means *STOP*.



The objects above differ in State because one is empty and the other is full.

The **Usability** feature allows you to indicate that you find one object more suitable for use than another.

For example, if you had to choose between objects and you think that one object fulfills its function better, is more practical, or is more pleasant or advantageous for any other reason, you can indicate this with the **Usability** property. **Usability** is a subjective decision.



If one backpack seems more comfortable or practical (e.g., due to its material or pocket placement), you can mark it with the **Usability** attribute.



If one of the dishes seems tastier, you can mark it with the Usability property.

The **Function** feature allows you to specify whether there are differences in the services or capabilities provided by the two objects.



For example, one of the above refrigerators also has a display, so it can provide more information, which is why it differs in the **Function** feature.



The images above differ in their **Function** feature, because one is an air freshener and the other is a cleaning agent.



The images above differ in the following feature: the **Number of Objects**. One image shows a small container and glitter, while the other shows only glitter.



Select the **Number of objects** feature even if the image contains the same objects, but in different numbers.

Composition. Select this feature as different if both images contain multiple objects but in different arrangements.



The images differ in **Composition**.



The images differ in **Composition**.

Sensitivity Power Analysis and Sample Size Estimation for Future Research

Sensitivity power analysis

We conducted a sensitivity power analysis with the software G*Power (Faul et al., 2007, 2009). Using sensitivity power analysis, we can calculate the required effect size given alpha, power, and the sample size.

First, we examined the sensitivity of our study for detecting the mean difference in mnemonic discrimination between age groups. With $\alpha = 0.05$, $power = .8$, and a sample size of $n = 93$ in the younger adult and $n = 103$ in the older adult group, we estimated that an effect size of $d = 0.4$ would be required to detect the effect with a two-tailed test. Previous studies (C. E. L. Stark et al., 2023) indicate that the age group difference in the LDI is medium to large ($d > 0.5$). Therefore, we conclude that Experiment 1 was adequately powered to detect the age difference in LDI.

Second, we examined the interaction between age group and visual discrimination on mnemonic discrimination. We chose the test “linear bivariate regression for two groups, differences between slopes” from the t-test family. As input parameters, we

defined the test type (two-tailed), the alpha error probability ($\alpha = 0.05$), the power ($1 - \beta = 0.8$), the sample size in the younger group ($n = 93$) and older group ($n = 103$), the residual standard deviation of the full linear regression model ($\sigma = 0.18$), and the standard deviation of VDR in the younger adult group ($\sigma = 0.14$) and the older adult group ($\sigma = 0.19$). The sensitivity power analysis indicated that with the given parameters, we would need a true difference in slopes of 0.46 to reliably detect the difference.

Calculating the detected difference in slopes

As a comparison to the effect size obtained by the sensitivity analysis, we performed a post-hoc calculation of the observed difference in slopes in Experiment 1, also using the software G*Power. We determined the difference in slopes with the following inputs: standard deviation of VDR in younger ($\sigma = 0.14$) and older adults ($\sigma = 0.19$), the standard deviation of LDI in younger ($\sigma = 0.17$) and older adults ($\sigma = 0.19$), the correlation between LDI and VDR in younger ($r = .19$) and older adults ($r = .09$), and the allocation ratio of groups ($n_2/n_1 = 1.11$). This analysis showed that the difference between slopes in our study was 0.14.

Sample size estimation for future investigations

We followed up on the results of the previous analyses by estimating the sample size needed for future studies investigating a potential interaction effect between visual discrimination and age group on mnemonic discrimination. Specifically, we estimated what sample size would yield 80% power to detect the smallest difference in slopes that could be of interest ($|\Delta \text{slope}| = 0.1$). We selected a conservative smallest effect size of interest of 0.1, corresponding to a small interaction effect, given the absence of established benchmarks for meaningful differences in slopes in this literature. Using the 'A priori' power analysis option in G*Power with the previously described inputs (allocation ratio, residual standard deviation, and the standard deviation of the VDR), we found that at $\alpha = .05$, a total sample size of $n = 4072$ would be needed to detect a .1 difference in slopes with 80% power.

References

- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. L. (2011). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning & Memory*, 18(1), 15–18. <https://doi.org/10.1101/lm.197111>
- Stark, C. E. L., Noche, J. A., Ebersberger, J. R., Mayer, L., & Stark, S. M. (2023). Optimizing the mnemonic similarity task for efficient, widespread use. *Frontiers in Behavioral Neuroscience*, 17. <https://www.frontiersin.org/articles/10.3389/fnbeh.2023.1080366>
- Stark, S. M., Kirwan, C. B., & Stark, C. E. L. (2019). Mnemonic Similarity Task: A Tool for Assessing Hippocampal Integrity. *Trends in Cognitive Sciences*, 23(11), 938–951. <https://doi.org/10.1016/j.tics.2019.08.003>