SUPPLEMENTARY MATERIAL for DOIXXXXXXXXXX

Gregor Kachel1, Daniel Haun2, & Manuel Bohn1

1 Leuphana University

2 Max-Planck-Institute for Evolutionary Anthropology

Author note

Correspondence concerning this article should be addressed to Gregor Kachel, Universitätsallee 1, C1.008a, 21335 Lüneburg. E-mail: gregor.kachel@leuphana.de

SUPPLEMENTARY MATERIAL for DOIXXXXXXXXXX

# Additional Analyses including item-level random effects

In addition to our main analyses, we preregistered an additional exploratory analyses for evaluating item-level effects separately in each study. For this we added a crossed random effect for items, allowing the relationship between age and performance to vary across items (correct ~ condition\*z.age + z.trial + sex + (z.trial |subid)). Due to the lower number of individual items within a task, we expect this model to be less diagnostic with regard to our main research question. However, it adds detail about the equivalence of items within conditions.

## Study 1

An exploratory model with additional item-level random effects showed that the variability of intercepts at the item level was small ( = 0.16, 95% CrI [0.01,0.40). Likewise there was only modest variation for the effect of age across items ( = 0.14, 95% CrI [0.01,0.34).The correlation between item difficulty and item-specific age slopes was highly uncertain (ρ = 0.35, 95% CrI [-0.84,0.98). Model weights based on the WAIC suggested that the full model was more likely to provide accurate predictions (weight = 81.52% ) compared to the item model (weight = 18.48%).The estimated difference in expected log predictive density (elpd) between models, however, was not much larger than the difference itself (ELPD diff = -0.82), indicating considerable uncertainty as to whether either model was preferable. For on overview of logg odds for all items across conditions see figure 1. For a full overview of coefficients and the random effects structure of the item model see table 1.

## Study 2

An exploratory model including item-level random effects indicated modest variability ( = 0.25, 95% CrI [0.02,0.54). The variability of age effects across items was small ( = 0.10, 95% CrI [0.00, 0.29), suggesting that the effect of age was fairly consistent across items. The correlation between item intercepts and age slopes was near zero and highly uncertain (ρ = 0.02, 95% CrI [-0.93,0.94), providing no reliable evidence that slightly easier or difficult items varied in age-related gains. Comparing the models using WAIC weights showed that including item-level random effects in Study 2 did not provide a clear advantage in out-of-sample predictive accuracy (full model 52.11%; item model 47.89%). Comparing the models corroborates this (ELPD WAIC; full model = -889.73; item model = -889.65). The difference in expected log predictive density was negligible (ELPD diff = 0), and within the range of sampling uncertainty (SE =0). For an overview of item level variability across conditions see figure 1. For a full table of coefficients and random effects of the exploratory model see table 3.

## Study 3

To evaluate the equivalence of items within conditions, item-level random effects were included in an exploratory model showing that item variation at the intercept was moderate ( = 0.21, 95% CrI [0.02,0.45) and that the effect of age varied slightly across items ( = 0.23, 95% CrI [0.04,0.45). Again, the correlation between item difficulty and item-specific age slopes was moderately positive but uncertain (ρ = 0.55, 95% CrI [-0.52,0.99) providing weak evidence that easier items had stronger age-related gains. Comparing the models using WAIC weights showed that the item model had higher predictive power (full model 26.07%; item model 73.93%). Comparing the models’ WAIC via ELPDs corroborates this (full model = -941.74; item model = -938.99). The difference in expected log predictive density favored the item model (ELPD diff = -2.75), though the uncertainty around this estimate suggests that the evidence is not decisive (SE =3.27). Graphically exploring item-level variability via logg odds indicates that items are still very similar with their estimated deviation from the intercept still including zero (cf. figure 1). For an overview of coefficients and random effects see table 2 in appendix 3.

## Discussion

In studies 1 and 2, the inclusion of item-level random effects did not meaningfully improve model fit. The main analyses reported in the manuscript and the exploratory models reported here account for the data equally well. For study 3, including item-level random effects yielded moderate variability. Here the analysis provides modest evidence for variability due to item level effects. As in all studies reported above, these were small compared to the participant random effects. None of the model comparisons provided substantial evidence for a better fit of the exploratory models. Hence, the additional analyses provided here provide no grounds for deterring from the preregistered analysis reported in the manuscript.

Table 1: Study 1 - Posterior Estimates for Exploratory Model with Item-Level Random Effects.

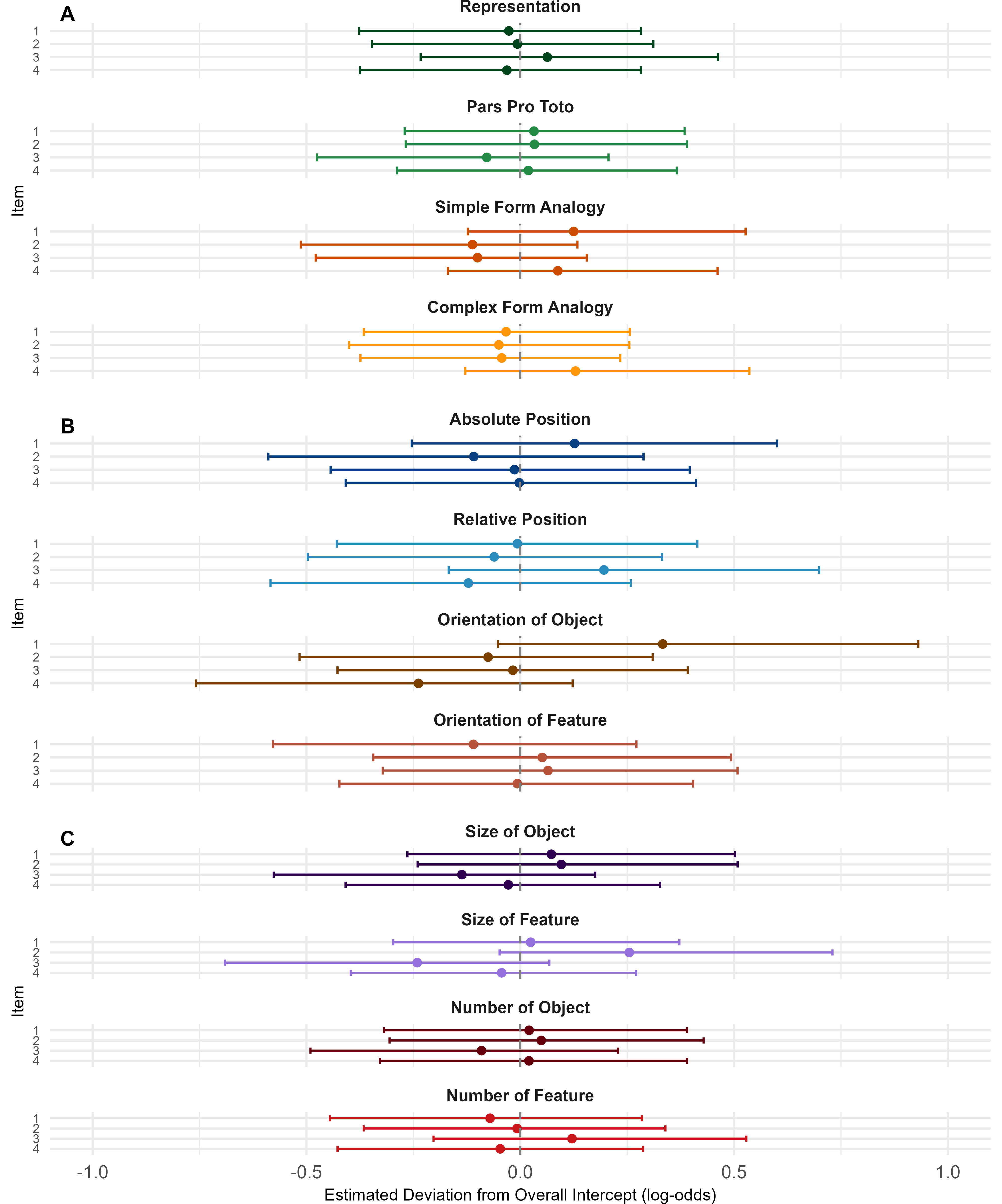
| Predictor | Estimate | SD | MAD | 95% CrI | Bulk ESS | Tail ESS |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 2.04 | 0.23 | 0.22 | [1.61, 2.49] | 5,822 | 6,699 |
| Pars Pro Toto | -0.14 | 0.25 | 0.24 | [-0.62, 0.34] | 7,100 | 6,942 |
| Simple Form Analogy | -1.40 | 0.23 | 0.22 | [-1.84, -0.95] | 7,235 | 6,211 |
| Complex Form Analogy | -1.37 | 0.23 | 0.22 | [-1.83, -0.92] | 6,472 | 6,144 |
| Age\* | 0.43 | 0.17 | 0.17 | [0.10, 0.76] | 4,841 | 5,254 |
| Trial\* | -0.01 | 0.07 | 0.07 | [-0.15, 0.12] | 13,286 | 7,953 |
| Sex (Male) | -0.29 | 0.21 | 0.21 | [-0.70, 0.11] | 5,499 | 6,735 |
| Pars Pro Toto × Age\* | 0.30 | 0.22 | 0.21 | [-0.12, 0.74] | 5,737 | 7,038 |
| Simple Form Analogy × Age\* | -0.20 | 0.20 | 0.19 | [-0.59, 0.20] | 4,929 | 5,546 |
| Complex Form Analogy × Age\* | -0.08 | 0.20 | 0.19 | [-0.47, 0.33] | 5,700 | 5,382 |
| Random Effects |  |  |  |  |  |  |
| Item Intercept (SD) | 0.16 | 0.10 | 0.10 | [0.01, 0.40] | 3,265 | 4,204 |
| Item × Age Slope (SD) | 0.14 | 0.09 | 0.09 | [0.01, 0.34] | 3,690 | 4,263 |
| Subject Intercept (SD) | 0.85 | 0.10 | 0.10 | [0.66, 1.06] | 3,964 | 6,536 |
| Subject × Trial slope (SD) | 0.17 | 0.10 | 0.11 | [0.01, 0.38] | 3,398 | 4,291 |
| Correlation: Item Intercept & Age Slope | 0.35 | 0.52 | 0.52 | [-0.84, 0.98] | 5,633 | 6,845 |
| Correlation: Subject Intercept & Trial Slope | -0.48 | 0.41 | 0.37 | [-0.98, 0.61] | 7,159 | 6,151 |
| Note. Estimates represent posterior means with 95% equal-tailed credible intervals (CrIs). MAD indicates Median Absolute Deviation. ESS refers to effective sample size. R̂ values omitted as they are all ~1 indicating convergence. \* = variables were standardized. | | | | | | |

Table 2: Study 3 - Posterior Estimates for Exploratory Model with Item-Level Random Effects.

| Predictor | Estimate | SD | MAD | 95% CrI | Bulk ESS | Tail ESS |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 1.06 | 0.21 | 0.21 | [0.64, 1.48] | 7,756 | 6,720 |
| Size of Feature | -0.72 | 0.24 | 0.23 | [-1.21, -0.24] | 7,616 | 6,724 |
| Number of Object | -0.08 | 0.25 | 0.24 | [-0.58, 0.41] | 7,955 | 6,891 |
| Number of Feature | -0.31 | 0.25 | 0.23 | [-0.83, 0.17] | 7,900 | 7,050 |
| Age\* | 0.72 | 0.19 | 0.18 | [0.35, 1.09] | 6,326 | 5,977 |
| Trial\* | 0.21 | 0.08 | 0.08 | [0.06, 0.37] | 12,910 | 8,132 |
| Sex (Male) | -0.18 | 0.19 | 0.19 | [-0.55, 0.19] | 9,583 | 8,379 |
| Size of Feature × Age\* | -0.51 | 0.25 | 0.23 | [-0.99, -0.03] | 6,989 | 6,449 |
| Number of Object × Age\* | 0.00 | 0.25 | 0.23 | [-0.49, 0.48] | 7,253 | 6,504 |
| Number of Feature × Age\* | -0.28 | 0.24 | 0.23 | [-0.77, 0.20] | 7,217 | 6,510 |
| Random Effects |  |  |  |  |  |  |
| Item Intercept (SD) | 0.21 | 0.11 | 0.10 | [0.02, 0.45] | 3,598 | 3,875 |
| Item × Age Slope (SD) | 0.23 | 0.10 | 0.09 | [0.04, 0.45] | 3,986 | 4,416 |
| Subject Intercept (SD) | 0.73 | 0.10 | 0.10 | [0.54, 0.94] | 3,766 | 6,193 |
| Subject × Trial slope (SD) | 0.42 | 0.11 | 0.11 | [0.18, 0.63] | 2,383 | 2,708 |
| Correlation: Item Intercept & Age Slope | 0.55 | 0.40 | 0.33 | [-0.52, 0.99] | 4,184 | 4,347 |
| Correlation: Subject Intercept & Trial Slope | 0.16 | 0.26 | 0.26 | [-0.34, 0.64] | 5,910 | 5,705 |
| Note. Estimates represent posterior means with 95% equal-tailed credible intervals (CrIs). MAD indicates Median Absolute Deviation. ESS refers to effective sample size. R̂ values omitted as they are all ~1 indicating convergence. \* = variables were standardized. | | | | | | |

Table 3: Study 2 - Posterior Estimates for Exploratory Model with Item-Level Random Effects.

| Predictor | Estimate | SD | MAD | 95% CrI | Bulk ESS | Tail ESS |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 1.09 | 0.25 | 0.25 | [0.60, 1.60] | 4,675 | 5,053 |
| Relative Position | -0.28 | 0.27 | 0.25 | [-0.83, 0.26] | 6,426 | 6,063 |
| Orientation of Object | -0.09 | 0.27 | 0.25 | [-0.64, 0.46] | 6,683 | 5,555 |
| Orientation of Feature | -0.12 | 0.27 | 0.25 | [-0.67, 0.42] | 6,297 | 5,765 |
| Age\* | 0.41 | 0.15 | 0.15 | [0.11, 0.71] | 5,036 | 6,300 |
| Trial\* | 0.26 | 0.09 | 0.09 | [0.10, 0.44] | 5,804 | 5,332 |
| Sex (Male) | -0.01 | 0.23 | 0.23 | [-0.46, 0.45] | 4,634 | 6,138 |
| Relative Position × Age\* | -0.06 | 0.18 | 0.18 | [-0.43, 0.30] | 6,852 | 6,539 |
| Orientation of Object × Age\* | 0.20 | 0.19 | 0.18 | [-0.17, 0.56] | 7,730 | 6,926 |
| Orientation of Feature × Age\* | -0.00 | 0.18 | 0.18 | [-0.37, 0.35] | 7,137 | 7,234 |
| Random Effects |  |  |  |  |  |  |
| Item Intercept (SD) | 0.25 | 0.13 | 0.12 | [0.02, 0.54] | 2,647 | 3,313 |
| Item × Age Slope (SD) | 0.10 | 0.08 | 0.08 | [0.00, 0.29] | 4,221 | 4,813 |
| Subject Intercept (SD) | 0.98 | 0.12 | 0.12 | [0.76, 1.25] | 2,948 | 6,009 |
| Subject × Trial slope (SD) | 0.47 | 0.12 | 0.12 | [0.24, 0.71] | 2,905 | 3,707 |
| Correlation: Item Intercept & Age Slope | 0.02 | 0.55 | 0.68 | [-0.93, 0.94] | 9,138 | 6,529 |
| Correlation: Subject Intercept & Trial Slope | 0.63 | 0.20 | 0.19 | [0.19, 0.95] | 3,470 | 4,339 |
| Note. Estimates represent posterior means with 95% equal-tailed credible intervals (CrIs). MAD indicates Median Absolute Deviation. ESS refers to effective sample size. R̂ values omitted as they are all ~1 indicating convergence. \* = variables were standardized. | | | | | | |



*Figure* 1. *Item-Level Random Intercepts by Condition.*