

MLM for Experimental Data

PSYC 575

September 15, 2020 (updated: 7 October 2022)

Experimental Designs

- Within-subjects/mixed designs
 - Random assignment at level 1
- Between-subjects
 - Cluster-randomized trial
 - Treatment at level 2
 - Multisite trial
 - Treatment at level 1
 - See example in chapter 11 of Snijders and Bosker (2012)

Learning Objectives

- Identify the correct levels with experimental studies
- Describe designs with crossed random levels
- Assign variables to appropriate levels
 - And tell which variables can have random slopes at which levels
- Compute effect sizes (R^2) for experimental data

Changes in Driving Scenes

- Example 1 from Hoffman & Rovine (2007)
 - Originally from Hoffman & Atchley (2001)
- Flicker paradigm:
https://coglab.cengage.com/labs/change_detection.shtml



Filo gèn' / CC BY-SA (with changes)
(<https://creativecommons.org/licenses/by-sa/4.0>)



Filo gèn' / CC BY-SA (with changes)
(<https://creativecommons.org/licenses/by-sa/4.0>)

Changes in Driving Scenes

- 153 persons
 - Younger ($n = 96$), $M_{\text{age}} = 19.7$ years ($SD = 2.3$);
 - Older ($n = 57$), $M_{\text{age}} = 75.7$ years ($SD = 5.4$)
- 51 scenes/items
 - Meaningfulness (0-5): meaningfulness to driving of the change
 - Salience (0-5): how visually conspicuous the change was within the scene
- Original plan: 2 (age group) \times 2 (meaning) \times 2 (salience) split-plot ANOVA

Data

- While RM-ANOVA uses the wide format, MLM requires the long format
 - Each unique response is in its own row
- See RStudio

Issues With ANOVA

- Unbalanced data/Missing responses
 - NA when change not detected within 60s
 - ANOVA uses listwise deletion: individual record is discarded when an individual has 1+ missing response

Issues of ANOVA

- ANOVA may require discretization of meaning and salience
- E.g., 0-2 for low salience, 3-5 for high salience
 - May hurt statistical power

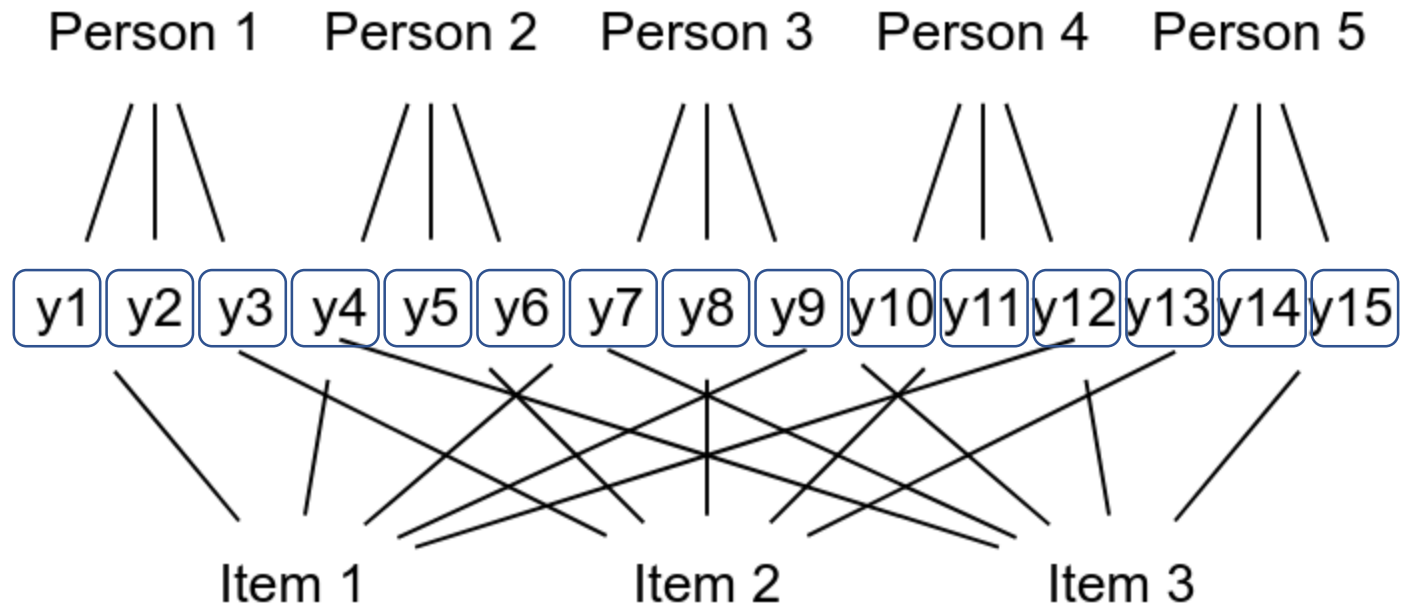
What is the Data Structure?

- Each response represents a person answering an item
- Responses nested within persons?
- Responses nested within items?

Crossed Levels

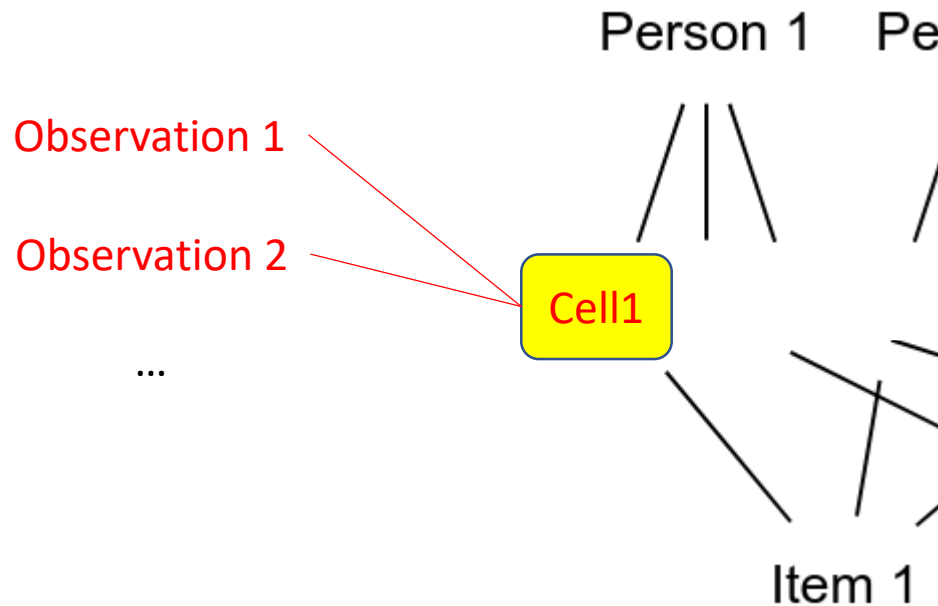
Crossed Levels

- Cross-classified model

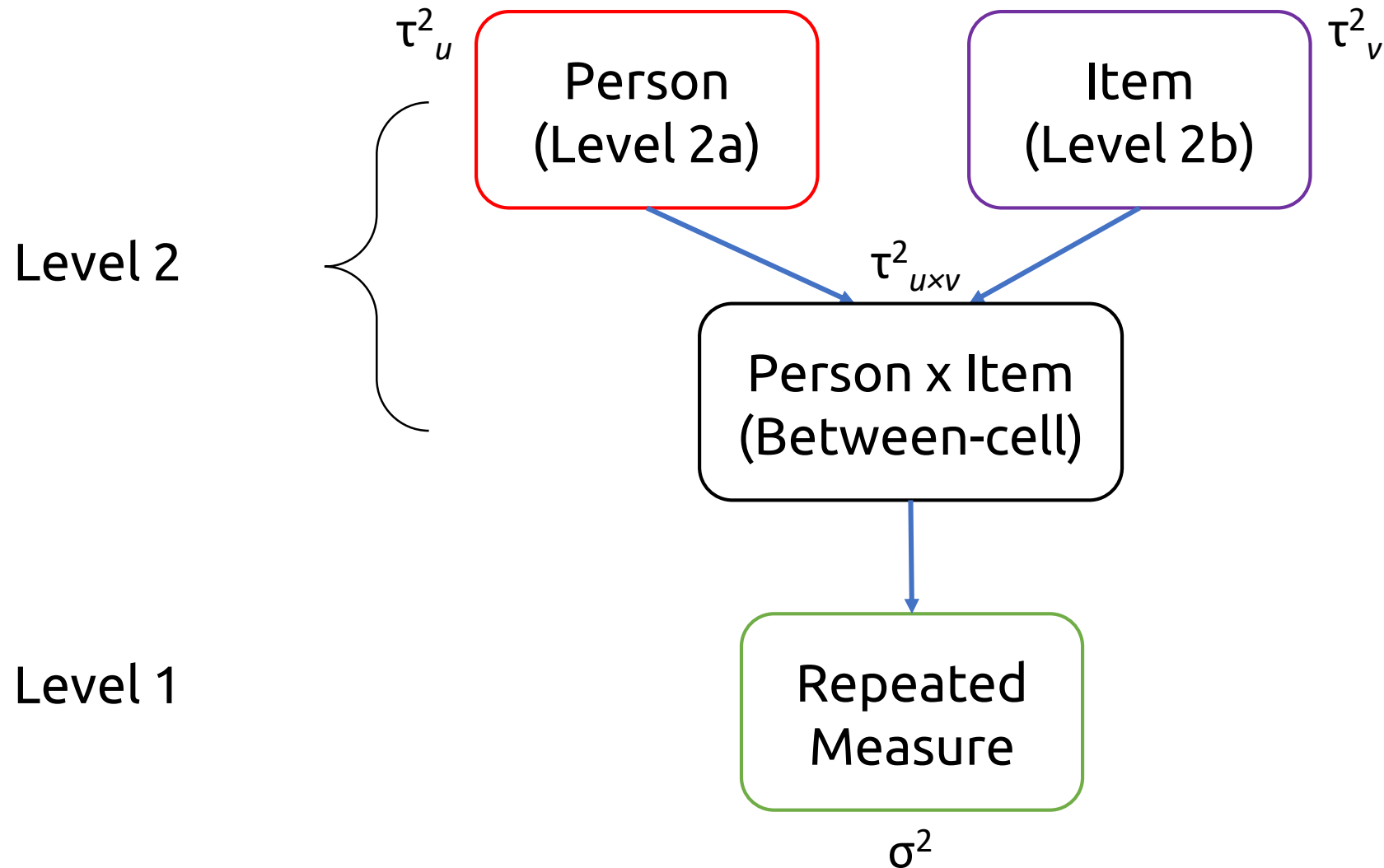


Crossed Levels With Interaction

- Multiple observations within each **cell**
 - E.g., Each person reacts to each stimulus in different conditions

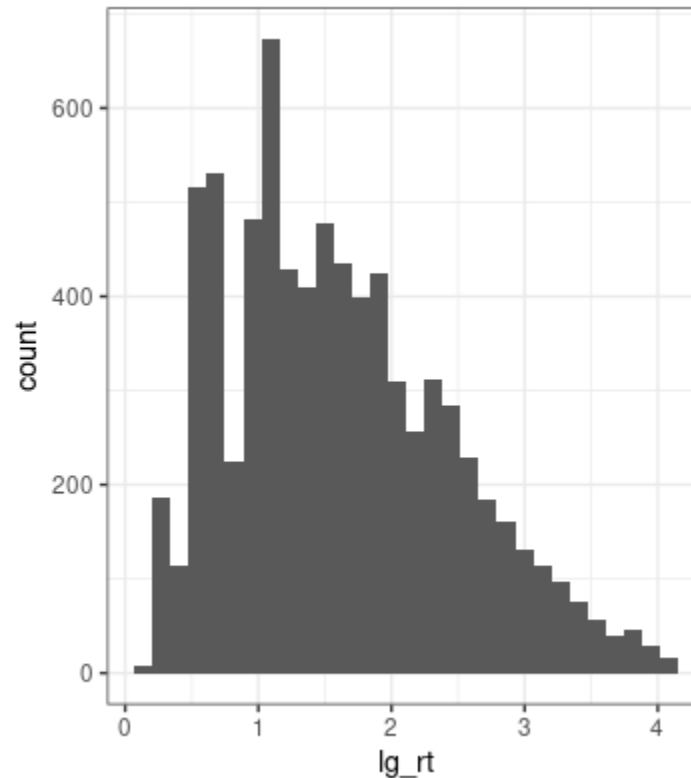
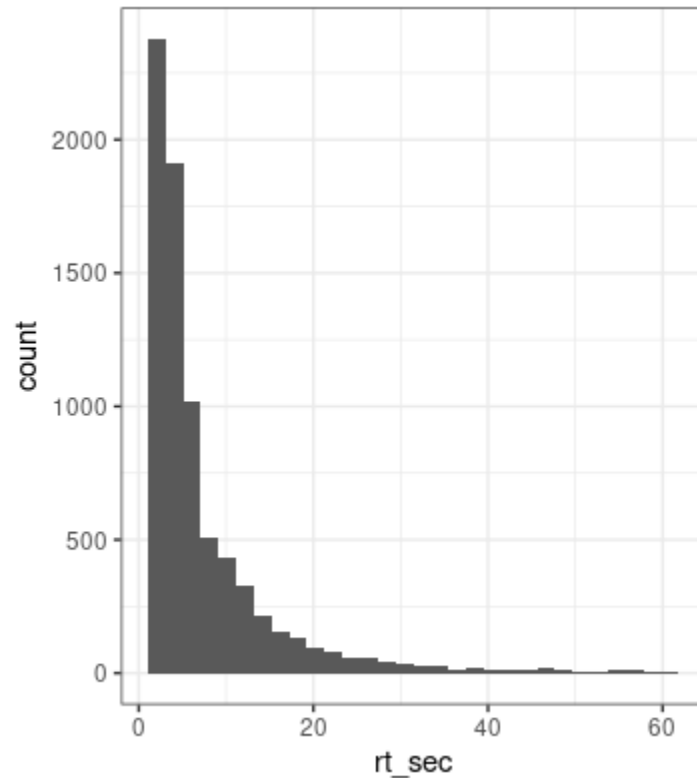


Crossed Levels at Level 2



Pre-Processing

- Log transformation of response time



Unconditional Model

- **Repeated Measure** (Within-cell) level (Lv 1)
 - $\lg_rt_{i(j,k)} = \beta_{0(j,k)} + e_{ijk}$
- **Between-cell** (**Person** × **Item**) level
 - $\beta_{0(j,k)} = \gamma_{00} + u_{0j} + v_{0k}$

In our example, each cell only has 1 observation

R Syntax

- If > 1 observations per cell, include interaction

`lg_rt ~ (1 | id) + (1 | Item) + (1 | id:Item)`

- If only 1 observation per cell, you get an error

> Error: number of levels of each grouping factor must be < number of observations (problems: id:Item)

- Otherwise, no interaction

`lg_rt ~ (1 | id) + (1 | Item)`

Intraclass Correlation

- **Person** level (Lv 2a) random effect: $u_{0j} \sim N(0, \tau_{u_0}^2)$

- $\text{ICC}(\text{person}) = \frac{\tau_{u_0}^2}{\tau_{u_0}^2 + \tau_{v_0}^2 + \sigma^2}$

- **Item** level (Lv 2b) random effect: $v_{0k} \sim N(0, \tau_{v_0}^2)$

- $\text{ICC}(\text{item}) = \frac{\tau_{v_0}^2}{\tau_{u_0}^2 + \tau_{v_0}^2 + \sigma^2}$

- $\text{ICC}(\text{person} + \text{item}) = \frac{\tau_{u_0}^2 + \tau_{v_0}^2}{\tau_{u_0}^2 + \tau_{v_0}^2 + \sigma^2}$

Intraclass Correlation

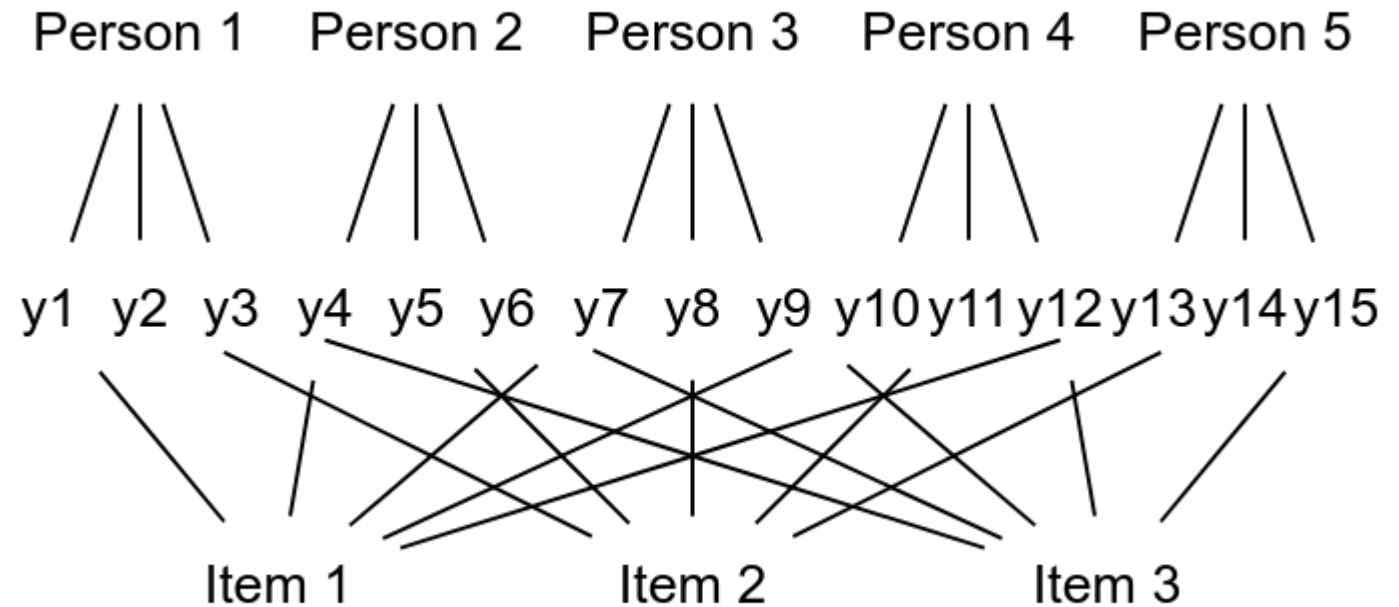
```
## Formula: lg_rt ~ (1 | id) + (1 | Item)
## Random effects:
##   Groups   Name          Variance Std.Dev.
##   id       (Intercept) 0.1803   0.4246
##   Item     (Intercept) 0.1259   0.3549
##   Residual                0.3899   0.6244
## Number of obs: 7646, groups:  id, 153; Item, 51
```

- $ICC(\text{person}) = 0.26$
- $ICC(\text{item}) = 0.18$
- $ICC(\text{person} + \text{item}) = 0.44$

Varying Slopes With Crossed Levels

Rule for Random Slopes

- A predictor can have random slopes at a level above or at a crossed level



Varying Slopes Across **Persons**

- Repeated Measure level (Lv 1)
 - $\lg_rt_{i(j, k)} = \beta_{0(j, k)} + e_{ijk}$
- Between-cell (**Person** × **Item**) level
 - $\beta_{0(j, k)} = \gamma_{00} + \beta_{1j} \text{meaning}_k + u_{0j} + v_{0k}$
- **Person** level (Lv 2a)
 - $\beta_{1j} = \gamma_{10} + u_{1j}$

Any predictors at the **repeated measure level** or at the **item level** can have random slopes across **persons**

Varying Slopes Across Items

- Repeated Measure level (Lv 1)
 - $\lg_rt_{i(j,k)} = \beta_{0(j,k)} + e_{ijk}$
- Between-cell (**Person** × **Item**) level
 - $\beta_{0(j,k)} = \gamma_{00} + \beta_{4k} \text{oldage}_j + u_{0j} + v_{0k}$
- **Item** level (Lv 2b)
 - $\beta_{4k} = \gamma_{40} + v_{4k}$

Any predictors at the repeated measure level or at the person level can have random slopes across items

Hypothesized Model

- Repeated Measure level (Lv 1)
 - $\lg_rt_{i(j,k)} = \beta_{0(j,k)} + e_{ijk}$
- Between-cell (**Person** × **Item**) level
 - $\beta_{0(j,k)} = \gamma_{00} + \beta_{1j} \text{meaning}_k + \beta_{2j} \text{salience}_k + \beta_{3j} \text{meaning}_k \times \text{salience}_k + \beta_{4k} \text{oldage}_j + u_{0j} + v_{0k}$
- **Person** level (Lv 2a) random slopes
 - $\beta_{1j} = \gamma_{10} + \gamma_{11} \text{oldage}_j + u_{1j}$
 - $\beta_{2j} = \gamma_{20} + \gamma_{21} \text{oldage}_j + u_{2j}$
 - $\beta_{3j} = \gamma_{30} + \gamma_{31} \text{oldage}_j + u_{3j}$

Hypothesized Model

- Item level (Lv 2b) random slopes
 - $\beta_{4k} = \gamma_{40} + v_{4k}$

Hypothesized Model

$$\begin{aligned}
 \bullet \lg_rt_{i(j, k)} = & \gamma_{00} && \text{Grand intercept} \\
 & + \gamma_{10} \text{ meaning}_k + \gamma_{20} \text{ salience}_k && \text{Item-level main and interaction effects} \\
 & + \gamma_{30} \text{ meaning}_k \times \text{salience}_k && \\
 & + \gamma_{40} \text{ oldage}_j && \text{Person-level main effect} \\
 & + \gamma_{11} \text{ meaning}_k \times \text{oldage}_j && \text{Person} \times \text{item cross-level interaction} \\
 & + \gamma_{21} \text{ salience}_k \times \text{oldage}_j && \\
 & + \gamma_{31} \text{ meaning}_k \times \text{salience}_k \times \text{oldage}_j && \\
 & + u_{0j} + u_{1j} \times \text{meaning}_k + u_{2j} \times \text{salience}_k && \text{Item-level random intercepts and slopes} \\
 & + u_{3j} \times \text{meaning}_k \times \text{salience}_k && \\
 & + v_{0k} + v_{4k} \times \text{oldage}_j && \text{Person-level random intercepts and slopes} \\
 & + e_{ijk} && \text{Within-cell deviation}
 \end{aligned}$$

Notes

- Because of counterbalancing
 - Person-level variables have no item-level variance
 - Item-level variables have no person-level variance
- Therefore, no need for cluster-mean centering

Formula: `c_sal ~ (1 | id)`

Data: `driving_dat`

REML criterion at convergence:
23545.51

Random effects:

Groups	Name	Std.Dev.
id	(Intercept)	0.000
Residual		1.094

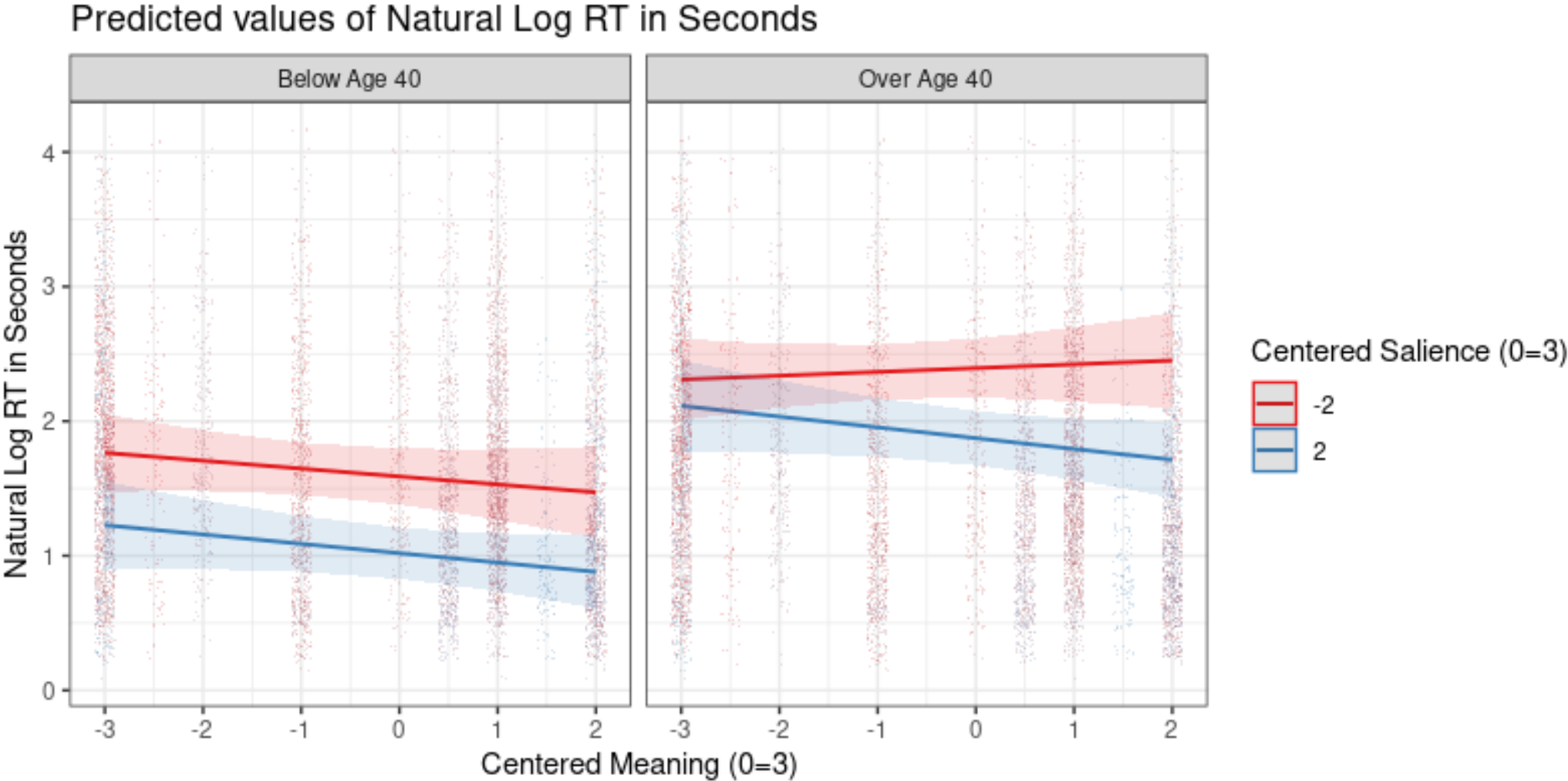
Notes

- To more easily interpret the interactions, we want to grand-mean center meaning and salience
 - They were centered to 3.0 in the data

Notes

- By testing random slopes one by one, the final model includes
 - Random slopes of `c_sal` (across persons)
 - Random slopes of `oldage` (across items)
- All two-way and three-way interactions were found not significant

Three-Way Interaction Plot



Effect Size

See R Code