HW 8

Zoey Zixi Lyu & Neshat Darvishi 2024-12-03

Warning: package 'ggplot2' was built under R version 4.3.1

Warning: package 'tidyr' was built under R version 4.3.1

Warning: package 'readr' was built under R version 4.3.1

Warning: package 'dplyr' was built under R version 4.3.1

Warning: package 'stringr' was built under R version 4.3.1

Warning: package 'lubridate' was built under R version 4.3.1

Warning: package 'modelsummary' was built under R version 4.3.1

Warning: package 'posterior' was built under R version 4.3.1

Warning: package 'bayesplot' was built under R version 4.3.1

Warning: package 'brms' was built under R version 4.3.1

Warning: package 'Rcpp' was built under R version 4.3.1

Research Question

> Can the acquisition of non-adjacent dependencies be facilitated by pairing them with either matching or non-matching acoustic cues?

Variables

- `id` : ID for participant
- `testSlider.response` : respond from participants ranking their familiarity level from 1 (not familiar at all) to 5 (very familiar)
- `stim` : stimuli type, rule: stimuli following NAD rule; class: stimuli is positional; catch: stimuli for attention check
- `trial_number` : order number of test trial
- `study_condition` : 5 study conditions: silent; wave-sound; speech; wave-sound that mismatch the NAD rule; speech that mismatch the NAD rule

Data Import

The data is in raw format and needs pre-processing.

Read Raw Data

Combine Data

Filter Data

To date, a total of 318 subjects have been tested. Data from 188 subjects could be used (speech = 47, tone = 46, speech_mismatch = 18, tone_mismatch = 22, silent = 55). 130 subjects have to be excluded because they did not pass the attention check.

Variable Summary

Table Table 1 shows the summary statistics of evaluation ratings by groups.

Table 1: Descriptive statistics by groups

		silent / positional	silent / NAD	tone / positional	tone / NAD	spe posi
Response	N	990	990	828	828	
	Mean	3.26	3.28	3.17	3.35	
	SD	1.22	1.21	1.26	1.26	
	Histogram					

Model

Let Y = `testSlider.response` , C = `study_condition` , S = `stim` , T = `trial_number` , ID = `id`

Since we are using a cumulative logit link model, we have 4 cuts $c_1 < c_2 < c_3 < c_4$ to separate Y into 5 levels

$$P(\boldsymbol{Y} \leq k) = \text{logit}^{-1}(\boldsymbol{\eta} - c_k), \quad k = 1, 2, 3, 4$$

The η is the linear predictor:

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}$$
$$\boldsymbol{b_i} \sim N\left(\mathbf{0}, \text{diag}\left(\boldsymbol{\tau}\right)\boldsymbol{\Omega}\text{diag}\left(\boldsymbol{\tau}\right)\right)$$

where β is the coefficients for fixed effect and b_i is the coefficients for random effect for participant i

Prior: Since we have 4 levels for cutting threshold c_k, we do not need a separate intercept in the linear part. For β, we have 4 for `study_condition` , 1 for `stim` , 1 for `trial_number` , 4 for `study_condition:stim` , 4 for `study_condition:trial_number` , 1 for `stim:trial_number` , 4 for `study_condition:stim:trial_number` , a total count of 19. For τ_l, we have the SD for `Intercept` , `stim` , `trial_number` , and `stim:trial_number` . For correlation we use LKJ prior.

$$c_k \sim N(0, 2),\ k = 1, 2, 3, 4$$
$$\beta_j \sim N(0, 2),\ j = 1, 2, \ldots, 19$$
$$\tau_l \sim Gamma(2, 2),\ l = 1, 2, 3, 4$$
$$\boldsymbol{\Omega} \sim \boldsymbol{lkj}(1)$$

Analysis

We used 4 chains, each with 4,000 iterations (first 2,000 as warm-ups).

Results

As shown in the rank histogram in Figure 1 below, the chains mixed well.

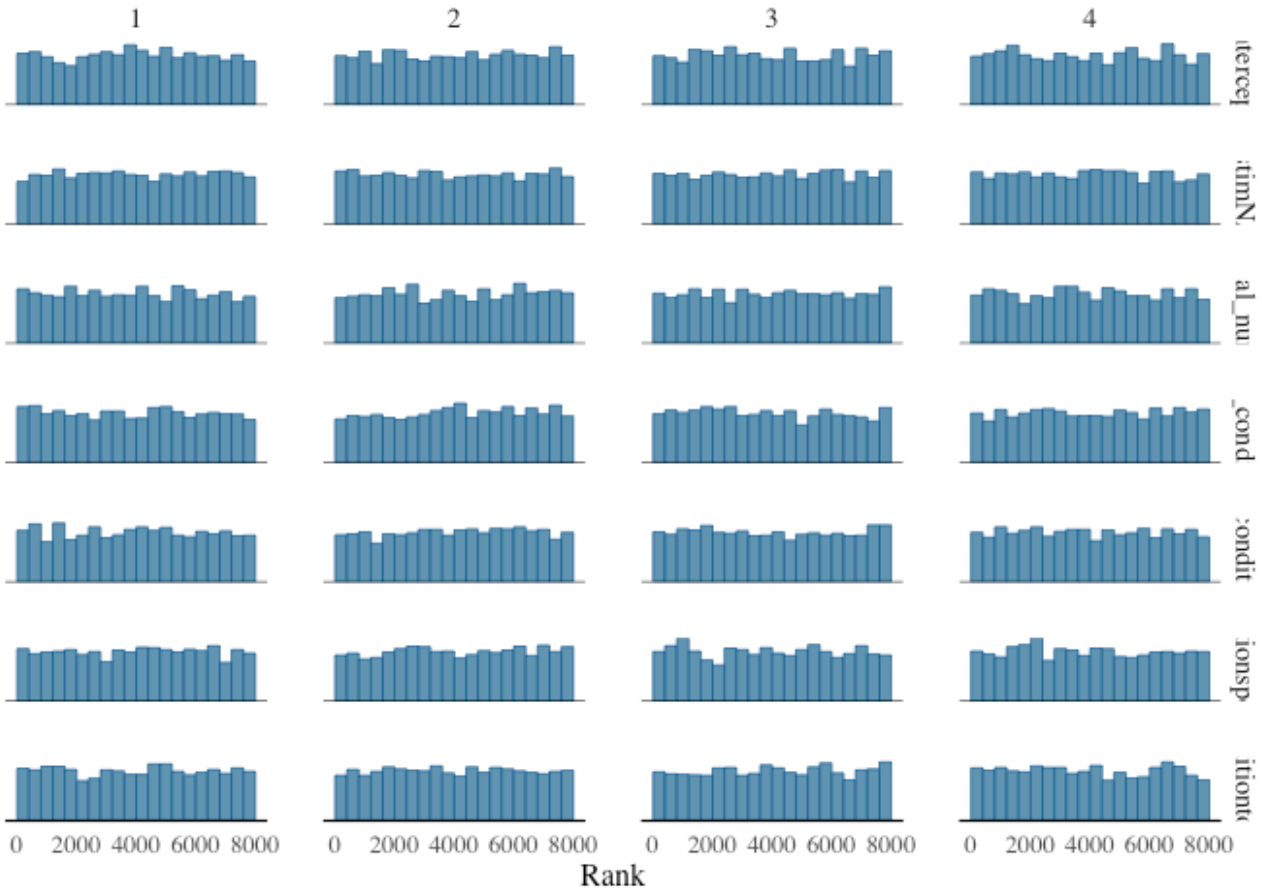

Figure 1: Rank histogram of the posterior distributions of selected model parameters.

Table 2 shows the posterior distributions of
$b(stimNAD) + b(study\ condition\ tone : stimNAD)$,
$b(stimNAD) + b(study\ condition\ speech : stimNAD)$,
$b(stimNAD) + b(study\ condition\ tone\ mismatch : stimNAD)$,
$b(stimNAD) + b(study\ condition\ speech\ mismatch : stimNAD)$, which indicate how much higher score for NAD stimuli vs. positional stimuli, or if the participant learns the NAD pattern correctly.

Table 2: Posterior summary of the model parameters.

variable	mean	median	sd	mad	q5	q95	rhat	ess_bulk
NAD silent	0.10	0.10	0.17	0.18	-0.18	0.38	1	3644.49
NAD tone	0.45	0.45	0.19	0.20	0.14	0.76	1	10728.30
NAD speech	0.33	0.33	0.19	0.19	0.02	0.64	1	10005.78
NAD tone mismatch	0.35	0.35	0.28	0.29	-0.12	0.81	1	9703.62
NAD speech mismatch	-0.08	-0.08	0.30	0.30	-0.58	0.42	1	8835.60

The analysis revealed that participants failed to learn the NAD pattern in the silent condition and the two mismatch conditions, as they were unable to distinguish between NAD stimuli and positional stimuli. In contrast, participants successfully learned the NAD pattern in the tone and speech conditions, as indicated by the posterior mean of 0.45 and 0.33 separately, and a 90% CI of [0.14, 0.76] and [0.02, 0.64] separately.