Appendix: Pilot Experiment

2022-04-26

A pilot study was conducted by one laboratory in March, 2021 to verify the feasibility of the experiment's design, and to test the planned procedures. Given the small size of the pilot, we do not expect to find significant effects, but we nonetheless also carried out the full planned analysis.

Participants

Participants were 19 children 8-12 months of age (10 females) recruited from (BLINDED FOR SUBMISSION) (7 8-month-olds, 2 9-month-olds, 2 10-month-olds, 4 11-month-olds, and 4 12-month-olds).

Materials and Design

The piloting phase adopted the same material and design as planned for the main study (see full description in the main text).

Procedure

The piloting phase adopted the same procedure as planned for the main study (see full description in the main text).

Results

We carried out the preregistered regression, but without per-lab random effects and procedure fixed effects since these variables have no variation in our sample. Thus, the mixed-effects linear regression predicted infants' log(looking time) the following fixed effects: familiarized rule (ABB vs. ABA), test trial type (same or different as familiarized rule), age (in months) and trial number (1-12), as well as 2-way interactions of trial type and each of 1) familiarized rule, 2) age, and 3) trial number. The model's random effects structure included per-subject intercepts with slopes by test trial number and type. The R syntax for the complete model was thus: log(looking time) ~ 1 + familiarization order * trial_type + age * trial_type + trial_num * trial_type + (trial_num*trial_type | subject)). However, this model's fit was singular (even when trial and age were centered and scaled), indicating that the random effects structure is too complex for the dataset. Thus, we pruned the random effects structure following standard procedures: removing either random slopes by trial number or by trial type allowed the model to converge. The model with slopes varying by trial number achieved better fit (AIC=409.6 vs. AIC=413.9 with slopes by trial type). Table 1 shows the regression coefficients from the preferred model, and Figure 1 shows these coefficients plotted with 95% confidence intervals. Only the effect of trial was significant ($\beta = -0.04$, $p = .01$), showing that looking time declined across trials. There was a marginal interaction of trial type and trial number ($\beta = 0.04$, $p = .05$), suggesting that participants' looking time may not decline across trials on trials exemplifying the same rule they were familiarized with during training. Although age and trial type did not show any significant effects or interaction in this small pilot, Figure 2 shows the log(looking time) as a function of these factors, as we may expect some effects in the full sample.

Table 1: Regression coefficients.

	Estimate	Std. Error	df	t value	Pr(>|t|)
(Intercept)	8.19	0.61	24.46	13.46	0.00
fam_conditionABB	0.32	0.18	22.31	1.75	0.09
trial_typesame	-0.36	0.47	188.68	-0.77	0.44
age_months	0.01	0.06	22.42	0.19	0.85
trial	-0.04	0.02	45.62	-2.54	0.01
fam_conditionABB:trial_typesame	-0.10	0.14	187.88	-0.71	0.48
trial_typesame:age_months	0.03	0.05	188.43	0.66	0.51
trial_typesame:trial	0.04	0.02	195.51	1.94	0.05

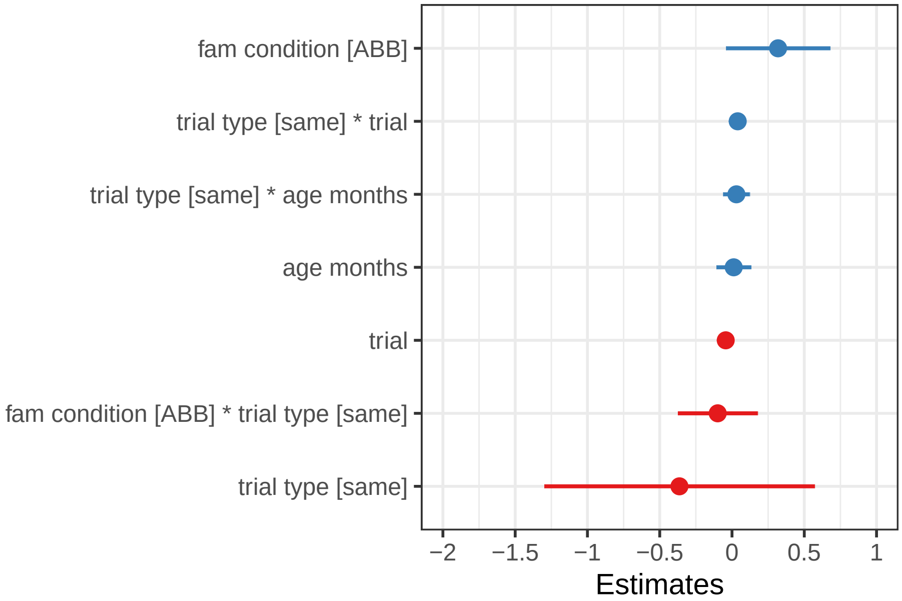

Figure 1: Regression coefficients with 95% confidence intervals.

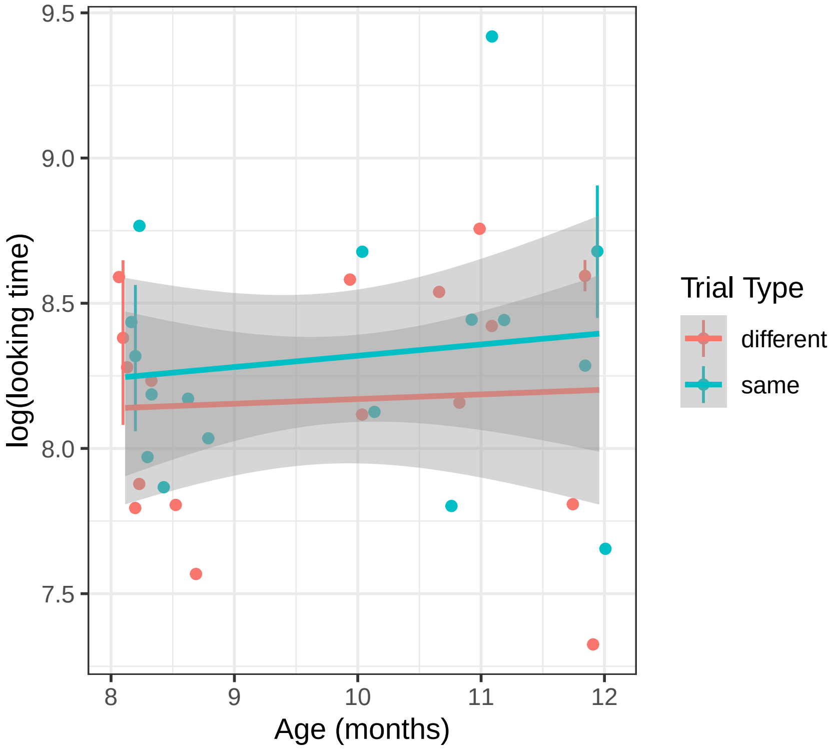

Figure 2: Log(looking time) by trial type and age, and bootstrapped 95% confidence intervals.