

## SI3. Linear Mixed-Effects Model

Chiara Santolin\*      Gonzalo García-Castro      Martin Zettersten  
Nuria Sebastian-Galles      Jenny R. Saffran

Updated: January 17th, 2020

We fit a Linear Mixed-Effects Model (LMEM) with looking time (*LT*) as response variable, that included test item (*Item*, Familiar vs. Novel), number of HeadTurn Preference Procedure experiments completed by infants (*HPP*), and their interaction ( $Item \times HPP$ ), as fixed effects. The *Item* predictor was treatment-coded, with *Familiar* trials as the baseline. Participant (*Participant*) and study (*Study*) were included as random effects. Following Barr, Levy, Scheepers, & Tily (2013), we fit a model with a maximal random effects structure. This maximal model included by-participant, by-study random intercepts, by-study *HPP* random slopes, and all variance components of the variance-covariance matrix for the by-study random effects. By-participant *HPP* random slopes were not included, as the sample was cross-sectional.

This model accounts for cross-participants variability in overall looking time (i.e. some infants are long lookers, some are short lookers), and for cross-studies differences in overall looking time (i.e. infants from some studies may look longer overall than infants from other studies), and allows the effect of *HPP* to vary across studies. We specified by-study random effects for two strong reasons. First, participants from different linguistic/cultural environments were included in both studies. This may have led to participants in one of the locations to looking longer in average than those from the other location. Second, in spite of their similarity both studies were not identical, which could also have led to differences in overall looking time. We used the `lmer` function of the `lme4` R package (???) to fit the model. We used the following formula to specify the model:

```
LookingTime ~ Item * HPP + (1 | Participant) + (1 + HPP | Study)
```

This model converged successfully. The correlation parameter (by-study intercepts and HPP slopes) was at boundary (-1). A closer look at the distribution of the residuals showed strong support against the normal distribution of the residuals (Shapiro-Wilk normality test:  $W = 0.98$ ,  $p = 0.001$ ). For this reason, we refit the model after log-transforming the looking times (???). We used the following formula to fit the log-transformed model:

```
LogLookingTime ~ Item * HPP + (1 | Participant) + (1 + HPP | Study)
```

This model converged successfully as well. The Cholesky factor on this model was singular. Although near-boundary correlation parameters or singular Cholesky factors are sub-optimal, the interpretation of such parameters is not involved in our hypothesis (???; ???). For these reasons, we continued interpreting the model. We found no evidence of non-normality of residuals in this model, (Shapiro-Wilk normality test:  $W = 0.99$ ,  $p = 0.557$ ). Fig. 1 shows the distribution of the residuals in the model before and after log-transforming looking times.

Significance testing was conducted by performing an *F*-test with Kenward-Roger approximation to degrees of freedom (???) on the coefficients of the fixed effects. We used the `Anova` function of the `car` R package (???) to do so. We found a statistically significant interaction term,  $F(1, 100) = 14.21$ ,  $p = 2.765483e-04$ , 95% CI = [0.04, 0.16], suggesting that the effect of trial type on looking time was influenced by the number of HPP experiments each participant participated in. The interaction shows that experience with a higher number of HPP experiments is associated with a stronger novelty preference. We also fit a Bayesian Linear-Mixed Model with the same formula using the `brm` function of the `brms` R package (???). We specified a normal prior for the three fixed effect coefficients with mean 0 and standard deviation 3.

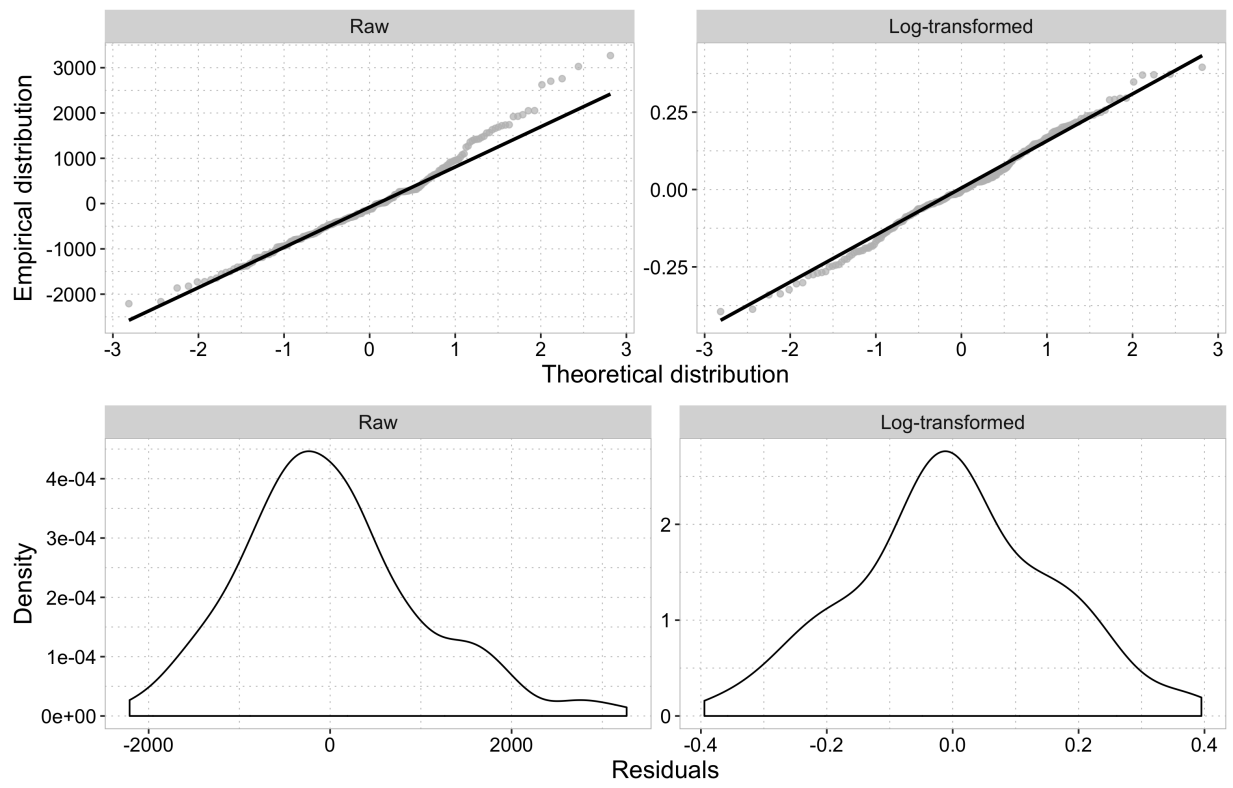
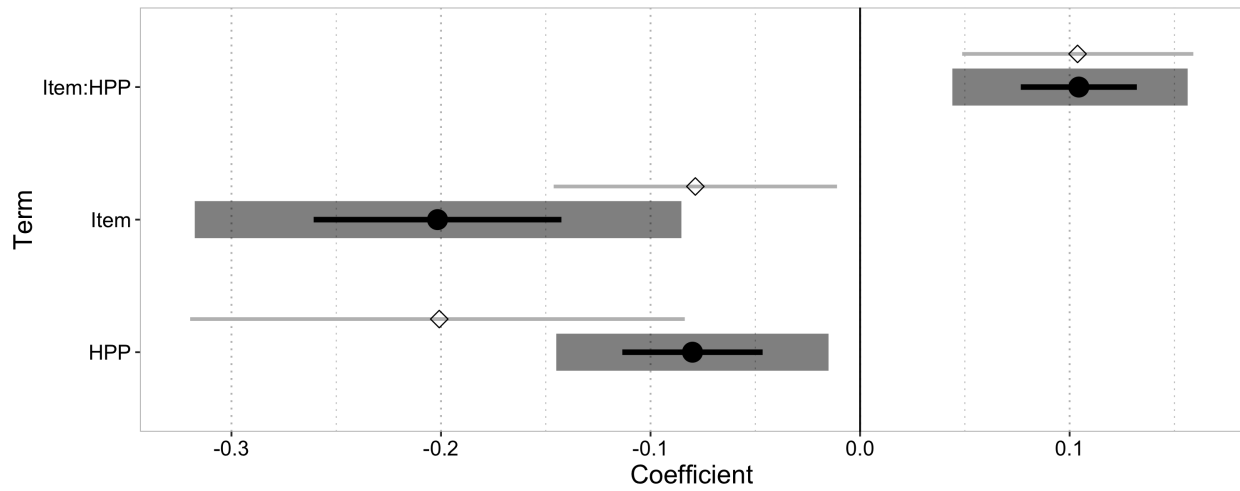


Figure 1: Distribution of residuals in relation to the normal distribution.

Although the 95% credible intervals of the *Item* and *HPP* effects contain 0, indicating that such main effects may be overestimated in the frequentist model, the 95% credible interval for coefficient for the interaction (*Item*  $\times$  *HPP*) does not contain 0, and overlaps almost entirely with the 95% frequentist confidence interval.

Table 1: Estimates of the linear mixed-effects model and outcomes of the Kenward-Roger F-tests performed on fixed effects. 95% confidence intervals were bootstrapped.

Term	Coefficient	Coefficient (Bayes)	SEM	95%CI	95% Cred. Int	F	Den. df	p
Intercept	8.8827	8.8637	0.1027	8.68, 9.14	8.36, 9.22	7270.8430	7.5463	0.0000
Item	-0.2017	-0.0786	0.0591	-0.32, -0.09	-0.15, -0.01	11.6345	100.0000	0.0009
HPP	-0.0800	-0.2008	0.0335	-0.14, -0.02	-0.32, -0.08	5.4379	138.3449	0.0211
Item * HPP	0.1044	0.1038	0.0277	0.04, 0.16	0.05, 0.16	14.2118	100.0000	0.0003



\begin{figure}  
\caption{Coefficients of the model. Black dots and error bars indicate the coefficient and SEM estimated by the frequentist model. Grey boxes indicate the 95% confidence interval. Black diamonds and grey error bars indicate the coefficient and 95% credible intervals estimated by the bayesian model.} \end{figure}

We observed little evidence of multicollinearity in the predictors we included in the model:

Term	VIF	Tolerance
Item	4.33	0.23
HPP	1.21	0.83
ItemCenter * HPP	4.54	0.22

## Comparison between models with different coding for *Item*

Model with *Item* dummy-coded with baseline on familiar trials (reported above)

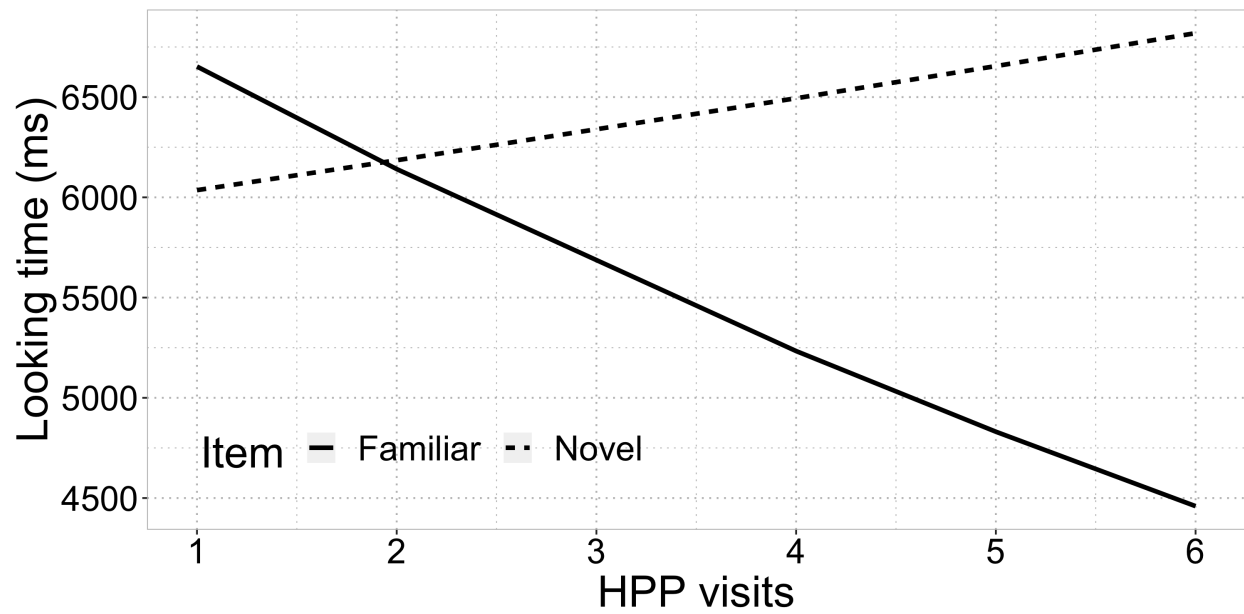


Figure 2: Predicted looking times plotted against HPP

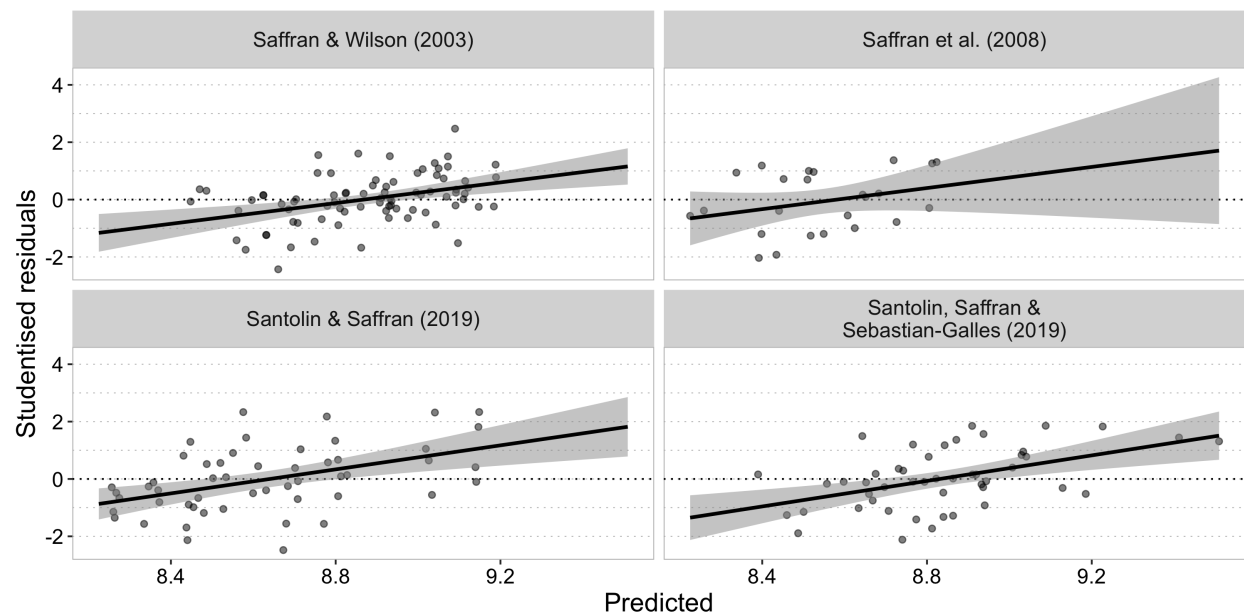


Figure 3: Distribution of residuals across predicted looking times.

Table 3: Estimates of the Item-dummy-coded linear mixed-effects model (with familiar trials as baseline) and outcomes of the Kenward-Roger F-tests performed on fixed effects. 95% confidence intervals were bootstrapped. This model is identical to the one reported above.

Term	Coefficient	SEM	95%CI	F	Den. df	p
Intercept	8.8827	0.1027	8.68, 9.14	7270.8430	7.5463	0.0000
Item	-0.2017	0.0591	-0.32, -0.09	11.6345	100.0000	0.0009
HPP	-0.0800	0.0335	-0.14, -0.02	5.4379	138.3449	0.0211
Item*HPP	0.1044	0.0277	0.04, 0.16	14.2118	100.0000	0.0003

#### Model with Item dummy-coded with baseline on novel trials

Table 4: Estimates of the Item-dummy-coded linear mixed-effects model (with novel trials as baseline) and outcomes of the Kenward-Roger F-tests performed on fixed effects. 95% confidence intervals were bootstrapped.

Term	Coefficient	SEM	95%CI	F	Den. df	p
Intercept	8.6811	0.1027	8.46, 8.89	6944.4631	7.5463	0.0000
ItemNovel	-0.2017	0.0591	-0.32, -0.11	11.6345	100.0000	0.0009
HPP	0.0244	0.0335	-0.04, 0.1	0.5059	138.3449	0.4781
ItemNovel*HPP	0.1044	0.0277	0.06, 0.16	14.2118	100.0000	0.0003

#### Model with Item effect-coded

Table 5: Estimates of the Item-effect-coded linear mixed-effects model and outcomes of the Kenward-Roger F-tests performed on fixed effects. 95% confidence intervals were bootstrapped.

Term	Coefficient	SEM	95%CI	F	Den. df	p
Intercept	8.7819	0.0983	8.58, 8.97	7729.0694	6.3508	0.0000
ItemCenter	-0.2017	0.0591	-0.31, -0.09	11.6345	100.0000	0.0009
HPP	-0.0278	0.0305	-0.09, 0.04	0.7844	99.0893	0.3779
ItemCenter*HPP	0.1044	0.0277	0.05, 0.16	14.2118	100.0000	0.0003

## References

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>