SI3. Linear Mixed-Effects Model

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

Updated: January 13th, 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 specifiyied 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 (Bates et al., 2015b) 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 (Csibra, Hernik, Mascaro, Tatone, & Lengyel, 2016). 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 (Bates et al., 2015a, 2015b). For these reasons, we continued interpreting the model. We found no evidence of non-normality of residuals in this model, (Shapiro-Wilk normality test: $W = 0.99$, $p = 0.524$). 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 (Kenward & Roger, 2009) on the coefficients of the fixed effects. We used the `Anova` function of the `car` R package (Fox & Weisberg, 2019) to do so. We found a statistically significant interaction term, $F(1, 100) = 14.21$, $p = 2.764819\text{e-}04$, 95% CI = [0.06, 0.15], 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

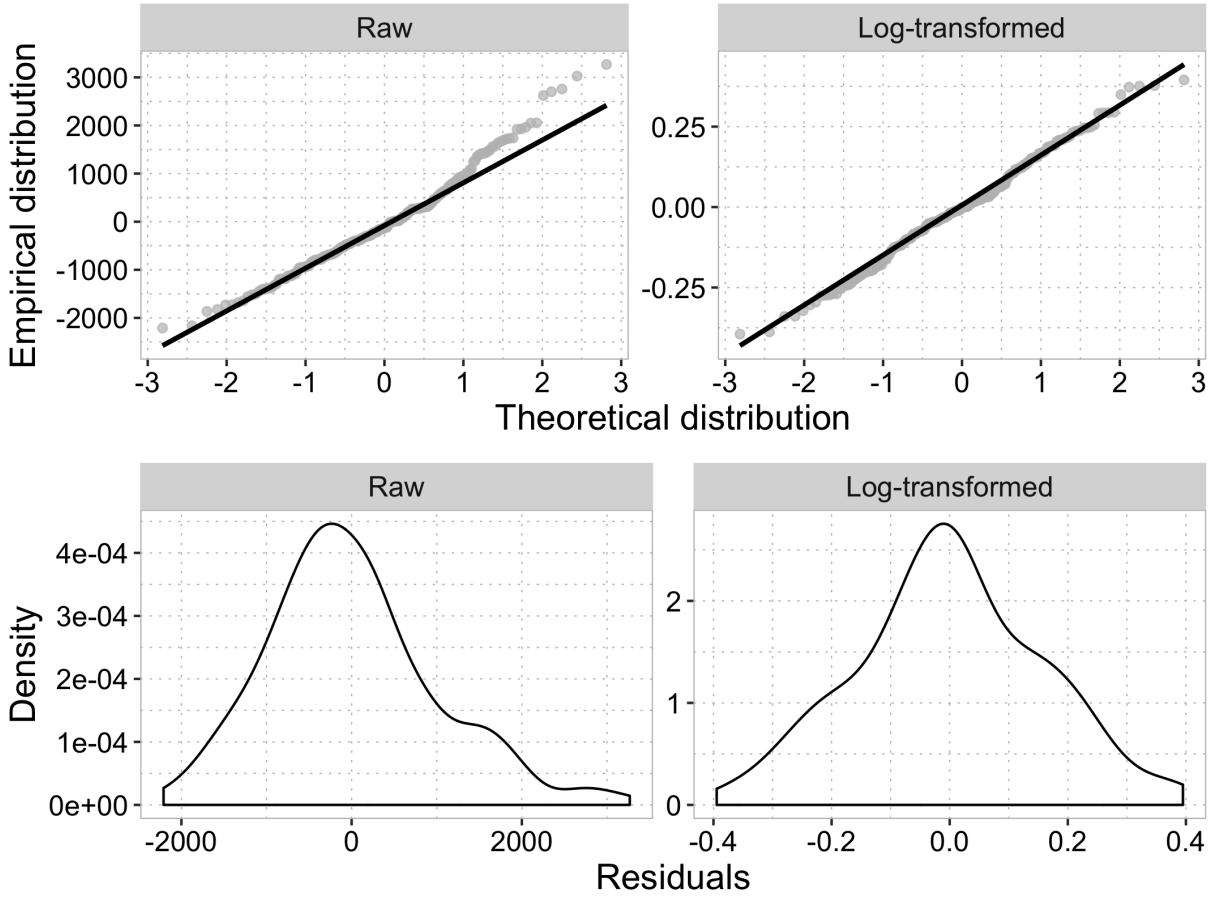

Figure 1: Distribution of residuals in relation to the normal distribution.

package (Bürkner, 2018). We specified a normal prior for the three fixed effect coefficients with mean 0 and standard deviation 3.

Although the 95% credible intervals of the *Item* and *HPP* effects contain 0, indicating that such main effects may be overstimated 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	8.8716	8.8791	0.1242	8.62, 9.12	8.43, 9.3	4618.6143	3.1579	0.0000
Item	-0.2017	-0.0797	0.0591	-0.31, -0.1	-0.21, 0.06	11.6349	100.0000	0.0009
HPP	-0.0763	-0.2006	0.0350	-0.15, 0	-0.32, -0.08	3.2236	2.6254	0.1834
Item * HPP	0.1044	0.1039	0.0277	0.06, 0.15	0.05, 0.16	14.2123	100.0000	0.0003

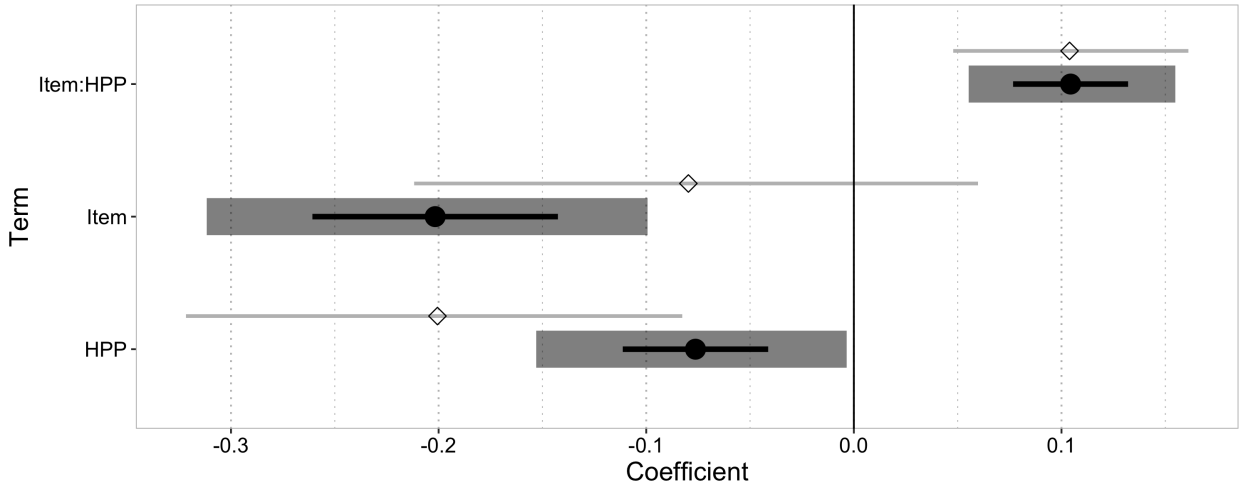

Figure 2: Predicted looking times plotted against HPP

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

Term	VIF	Tolerance
Item	4.33	0.23
HPP	1.19	0.84
ItemCenter * HPP	4.52	0.22

Comparison between models with different coding for *Item*

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

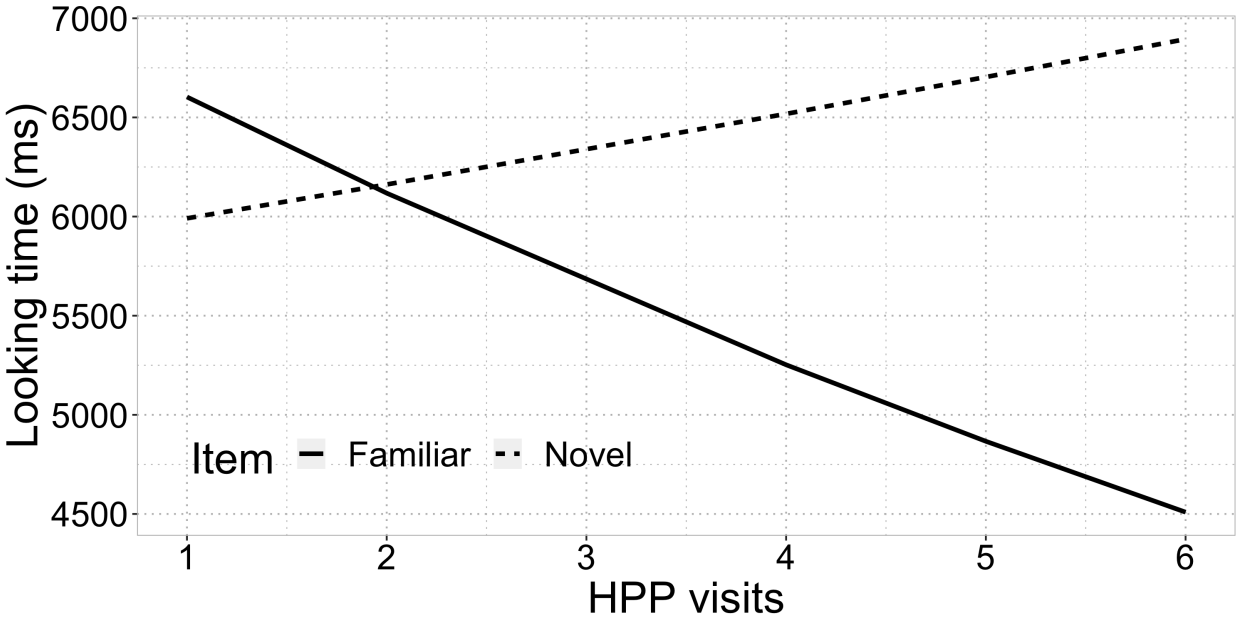

Figure 3: Predicted looking times plotted against HPP

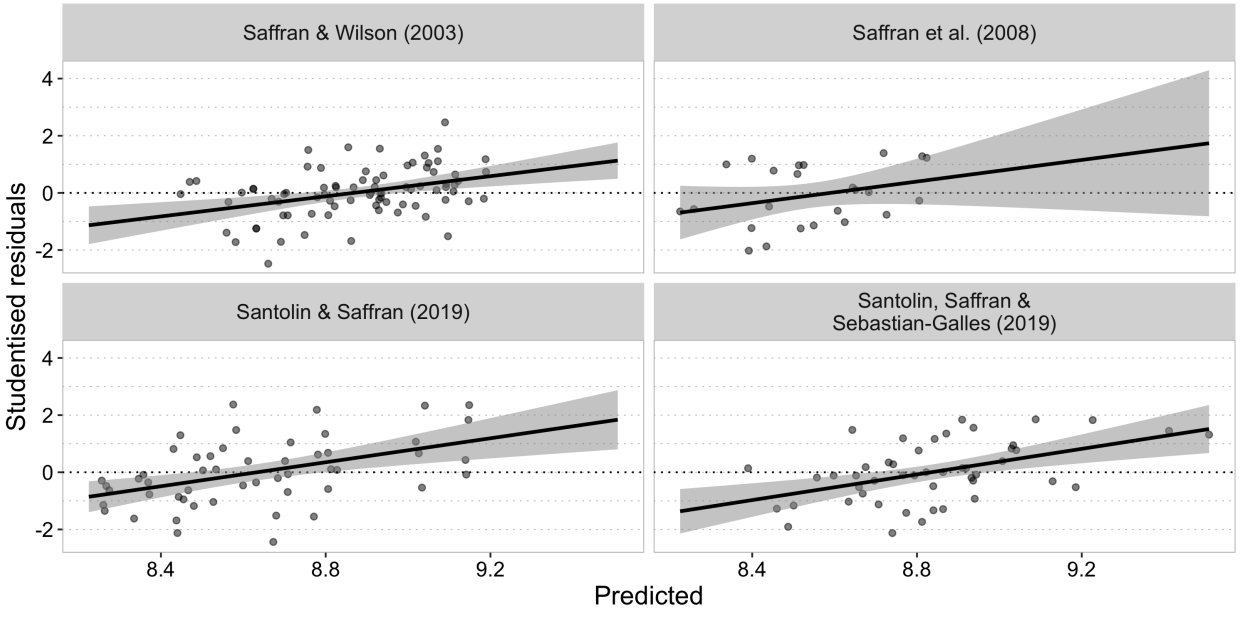

Figure 4: Distribution of residuals across predicted looking times.

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	8.8716	0.1242	8.62, 9.12	4618.6143	3.1579	0.0000
Item	-0.2017	0.0591	-0.31, -0.1	11.6349	100.0000	0.0009
HPP	-0.0763	0.0350	-0.15, 0	3.2236	2.6254	0.1834
Item*HPP	0.1044	0.0277	0.06, 0.15	14.2123	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	8.6699	0.1242	8.43, 8.92	4411.0322	3.1579	0.0000
ItemNovel	-0.2017	0.0591	-0.32, -0.07	11.6349	100.0000	0.0009
HPP	0.0281	0.0350	-0.05, 0.1	0.4373	2.6254	0.1834
ItemNovel*HPP	0.1044	0.0277	0.05, 0.16	14.2123	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	8.7708	0.1206	8.53, 9.01	4758.2075	2.8103	0.0000
ItemCenter	-0.2017	0.0591	-0.32, -0.07	11.6349	100.0000	0.0009
HPP	-0.0241	0.0322	-0.1, 0.05	0.3598	1.8700	0.1834
ItemCenter*HPP	0.1044	0.0277	0.05, 0.16	14.2123	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

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015b). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*,

10(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Csibra, G., Hernik, M., Mascaro, O., Tatone, D., & Lengyel, M. (2016). Statistical treatment of looking-time data. *Developmental Psychology*, *52*(4), 521.

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third). Thousand Oaks CA: Sage. Retrieved from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Kenward, M. G., & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis*, *53*(7), 2583–2595. https://doi.org/10.1016/j.csda.2008.12.013