SI3. Linear Mixed-Effects Model

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

Updated: January 20th, 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 dummt-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 that included all random effects our design could allow. Initially, this maximal model included by-participant, by-study random intercepts, by-study HPP and $Item$ 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. 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` package (Bates, Mächler, Bolker, & Walker, 2015) of the R environment (R Core Team, 2018) to fit the model. We simplified the random effects structure of the model until we obtained plausible values of the variance components of the variance-covariance matrix (i.e., correlation parameter was not near-boundary and variance components were not zero). The resulting model only included by-participant and by-study random intercepts. Its correspondent formula for being fitted in `lme4` was the following:

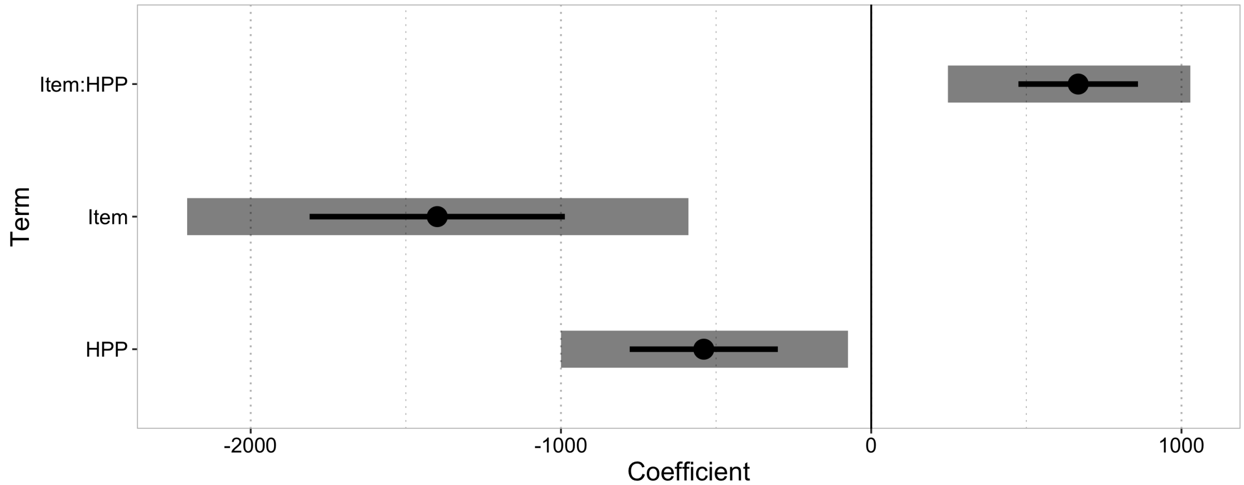
```
LookingTime ~ Item * HPP + (1 | Participant) + (1 | Study)
```

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). This model converged successfully.

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, $F(1, 100) = 11.99$, $p = 0.0008$, 95% CI = [247.23, 1028.51], 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.

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	SEM	95%CI	F	Den. df	p
Intercept	7679.1090	673.1428	6390.13, 9294.12	124.6894	9.0645	0.0000
Item	-1398.7746	411.3075	-2204.84, -589.11	11.5654	100.0000	0.0010
HPP	-539.7046	238.6913	-999.88, -74.83	4.8005	133.0997	0.0302
Item * HPP	667.1130	192.6372	247.23, 1028.51	11.9927	100.0000	0.0008



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

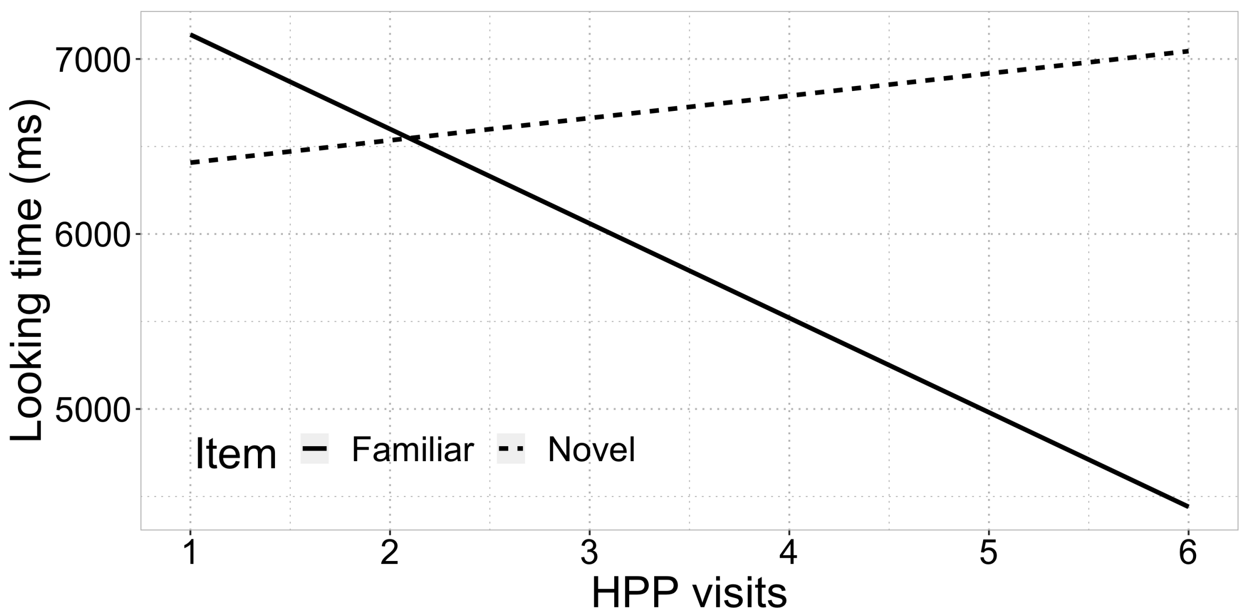


Figure 1: Predicted looking times plotted against HPP.

Comparison between models with different coding for *Item*

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

Table 2: 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	7679.1090	673.1428	6390.13, 9294.12	124.6894	9.0645	0.0000
Item	-1398.7746	411.3075	-2204.84, -589.11	11.5654	100.0000	0.0010
HPP	-539.7046	238.6913	-999.88, -74.83	4.8005	133.0997	0.0302
Item*HPP	667.1130	192.6372	247.23, 1028.51	11.9927	100.0000	0.0008

Model with Item dummy-coded with baseline on novel trials

Table 3: 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	6280.3344	673.1428	4790.33, 7691.79	83.4014	9.0645	0.0000
ItemNovel	-1398.7746	411.3075	-2236.23, -726.47	11.5654	100.0000	0.0010
HPP	127.4084	238.6913	-369.48, 640.51	0.2675	133.0997	0.6059
ItemNovel*HPP	667.1130	192.6372	354.06, 1057.97	11.9927	100.0000	0.0008

Model with Item effect-coded

Table 4: 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	6979.7217	640.9585	5605.52, 8175.28	113.1281	7.4572	0.0000
ItemCenter	-1398.7746	411.3075	-2169.1, -638.57	11.5654	100.0000	0.0010
HPP	-206.1481	218.3948	-663.89, 250.42	0.8268	96.6948	0.3655
ItemCenter*HPP	667.1130	192.6372	272.38, 1020.71	11.9927	100.0000	0.0008

HPP3

Given the reduced number of infants that had participated in the studies more than 3 times, we tested the robustness of the estimated coefficients by fitting a model that only included participants with experience in 1 up to 3 HPP studies. This model showed similar outcomes. We found a statistically significant interaction, $F(1, 92) = 3.67$, $p = 0.0584$, 95% CI = [-0.01, 0.17], 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. The main effect of HPP, however, was no longer statistically significant.

Table 5: 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. Participants with more than 3 HPP studies have been excluded. 95% confidence intervals were bootstrapped.

Term	Coefficient	SEM	95%CI	F	Den. df	p
Intercept	8.8374	0.1176	8.62, 9.08	5439.2240	11.6225	0.0000
Item	-0.1670	0.0758	-0.32, -0.01	4.8570	92.0000	0.0010
HPP	-0.0487	0.0510	-0.16, 0.05	0.8639	126.8885	0.6059
Item * HPP	0.0802	0.0419	-0.01, 0.17	3.6727	92.0000	0.0008

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

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.

R Core Team. (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/