Experience with research paradigms relates to infants' direction of preference.

Chiara Santolin[1], Gonzalo Garcia-Castro[1], Martin Zettersten[2], Nuria Sebastian-Galles[1], & Jenny Saffran[2]

[1] Centre for Brain and Cognition, Universitat Pompeu Fabra

[2] Waisman Center & Department of Psychology, University of Wisconsin-Madison

Preprint submitted to peer-review on February, 17th, 2020

Author Note

Correspondence concerning this article should be addressed to Chiara Santolin, Edifici Merce Rodereda, Calle Ramón Trias Fargas, 25, 08018 Barcelona. E-mail: chiara.santolin@upf.edu

Abstract

Interpreting and predicting direction of preference in infant behavioral research has been a thorny issue for decades. Several factors have been proposed to account for familiarity and novelty preferences in habituation and familiarization studies, including infant age, length of exposure and task complexity. The current study explores an additional dimension that may affect direction of preference: amount of experience with the experimental task. To test this hypothesis, we re-analyzed the data from 4 experiments on artificial grammar learning in 12-month-old infants run using the Head- turn Preference Procedure (HPP). The participants in these studies varied substantially in their number of laboratory visits. Linear mixed-effects results show that the number of HPP studies in which infants had previously participated is related to infants' direction of preference: infants who had no (or limited) experience with the HPP setting were more likely to show familiarity preferences than infants who had amassed more experience with this task in prior study visits. Interestingly, the effect is driven by a significant decrease in looking time for familiar trials. These results have important implications for the interpretation of experimental results: infants' experience with a given paradigm or, more broadly, with the lab environment, may affect their patterns of preferences.

Keywords: preferential looking, familiarity preference, novelty preference, head-turn preference procedure, linear mixed-effects model

Word count: 2939

Experience with research paradigms relates to infants' direction of preference.

Introduction

In infancy research, the importance of changes in preferential looking has been recognized since at least the 1960s, when Fantz (1964) showed that young infants preferentially attend to novel visual stimuli. Subsequent studies extended this evidence to other domains, including acoustic perception and cognition, revealing differences in direction of preference. Theories intended to account for such differences suggest that novelty preferences arise when infants have completed the processing of a (familiar) stimulus (see Houston-Price and Nakai (2004) and Aslin (2007) for reviews).

Rather than representing a binary distinction, direction of preference can be better described as a continuum from more familiar to more novel (e.g., Thiessen, Hill, & Saffran, 2005). Hunter and Ames (1988) provide the most comprehensive model of preferential looking with three central factors hypothesized to affect the strength and direction of preference: age, familiarization duration and task complexity. In a given task, younger infants tend to prefer the familiar stimulus whereas older infants are more likely to prefer the novel one (e.g., Colombo & Bundy, 1983; though see Bergmann & Cristia, 2016, for a meta-analysis suggesting that age does not predict shifts in preference). A shorter exposure to familiar stimulus prior to testing also leads infants to subsequently prefer the familiar object (for reviews, see also Rose, Feldman, & Jankowski, 2004). Task complexity refers to the stage of stimulus processing. For example, in a visual recognition task, 4-month-old infants revealed a systematic preference for the familiar object prior to showing a strong preference for the novel object (Roder, Bushneil, & Sasseville, 2000). Task complexity can also refer to the complexity of stimuli. For example, sequential stimuli put greater strain on memory resources than stimuli in which all components are simultaneously available (e.g., Ferguson, Franconeri, & Waxman, 2018). A related dimension is the similarity across stimuli used at familiarization and test: when there is a close perceptual match such as

same colors or sounds during training and test, infants are more likely to show a novelty preference (e.g., Hunter & Ames, 1988; Thiessen & Saffran, 2003).

The combination of these factors informs predictions concerning direction of preference in systematic ways. For example, Thiessen et al. (2005) manipulated length of exposure and observed a flip from familiarity to novelty preference after doubling the amount of familiarization received by the infants. Similarly, Ferguson et al. (2018) manipulated sequential vs. spatial presentation of visual patterns, and observed stronger novelty effects with (a) increasing age and (b) spatial presentation. That said, it is also frequently the case that the observed direction of preference does not conform with expectations based on the dimensions noted above; the infancy literature is rife with examples of counterintuitive patterns of preference (e.g., Fiser & Aslin, 2001; Bosch & Sebastián-Gallés, 2001; Dawson & Gerken, 2009 for novelty preference in 4-month-olds; Johnson et al., 2009 for both novelty and familiarity preference in 11-months-olds; Jusczyk & Aslin, 1995 for familiarity preference in 7-months-olds; Sebastián-Gallés & Bosch, 2009; Thiessen, 2012 for post-habituation familiarity preference).

One factor that is frequently overlooked in considerations of infant data is that infants do not arrive at the lab as naïve participants. Like adults, they bring significant prior experience that may influence their performance in lab tasks. In many instances, researchers attempt to override or sidestep those experiences by using novel stimuli (e.g., unfamiliar languages, shapes or sounds), or by integrating those experiences into their experimental designs (e.g., monolingual vs. bilingual infants). But there may also be forms of experience that go unidentified by researchers. One such factor is that many infants participate in more than one experiment over the course of weeks or months. Testing the same participants in multiple (putatively unrelated) experiments is a common practice in infant research, reflecting the challenges of advancing a field of investigation that is based on a limited and hard-to-recruit population. Researchers are typically very careful to avoid stimulus contagion across unrelated studies, but it is possible that prior lab experience

impacts infants' performance.

The purpose of this article is to explore the effect of experience with experimental paradigms on direction of preference in learning tasks. This idea emerged from a puzzling pattern of results in a replication of a published study focused on non-linguistic artificial grammar learning in 12-month-olds (Santolin & Saffran, 2019). We observed a flip in preference from novelty to familiarity between the original study and its replication (Santolin et al., 2019), despite the use of identical stimuli and procedures. While there were some differences between the studies (most notably, in the location in which the studies were run), one main factor jumped out at us: many of the infants in the study that elicited a novelty preference had participated in prior studies with Head-turn Preference Procedure (HPP), whereas most of the infants in the study that elicited a familiarity preference were first-time HPP participants. We reasoned that the more familiarity infants had with the lab apparatus and task demands, the more likely they would be to learn rapidly, leading to a novelty preference. To explore this question, we combined the data from these two experiments with the data from two other published artificial grammar learning tasks with similar designs that included 12-month-olds with a range in number of lab visits (Saffran et al., 2008, Exp. 1 Language P; and Saffran & Wilson, 2003, Exp. 2). Our hypothesis was that the amount of infants' prior experience with HPP would affect direction of preference.

Methods

A brief description of the four experiments included in this analysis, and our rationale for selecting them, is provided in the Appendix, Section 1 (see Figure A1 for a summary of the results). A fully reproducible repository hosting data and analyses is available at https://osf.io/g95ub/.

We modeled results of all infants (N=102) who participated in the four studies. Number of HPP visits varied from one to six total visits (including the current visit). We

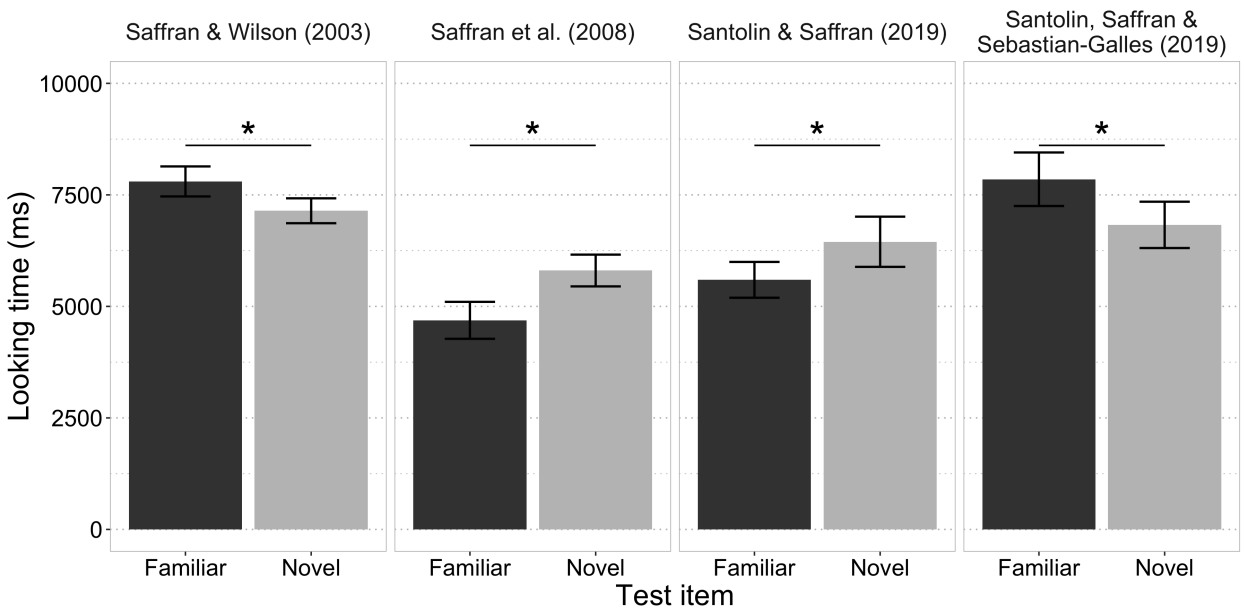

Figure 1. Looking time for familiar and novel test stimuli of the original studies. Stimuli vary based on the experiment. Error bars indicate the standard error of the mean.

fit a linear mixed-effect model including *Looking time* as response variable, and *Test item* (Familiar vs. Novel), *HPP* (number of experiments completed by infants) and their interaction as fixed effects. We also included by-participant and by-study random intercepts (4 levels: Santolin & Saffran, 2019; Saffran et al., 2008, and @saffran2003; Santolin et al., 2019). The *HPP* predictor was coded as a continuous variable indicating the overall number of experiments with the Head-turn Preference Procedure the infants participated in. *Test Item* was centered on familiar test items (Familiar = 0; Novel = 1). Since the experiments differ at distinct levels (e.g., different stimuli, lab location), the model accounted for cross-participant and cross-study differences in looking time. Degrees of freedom were approximated using the Kenward-Roger approach (Judd, Westfall, & Kenny, 2012), which can result in non-integer values. Further modeling details are provided in Appendix, Section 3.

We predicted a *Test item* (familiar vs. novel) by number of *HPP* interaction, indicating that the duration of infants' looking towards familiar versus novel items would

depend on infants' HPP experience. An interaction could result from at least three different patterns of results: an increase in looking time for novel items, a decrease in looking time for familiar items, or both, as a result of additional experience in HPP studies.

Results

We found a statistically significant interaction ($F(1,100.00) = 11.99$, $p = .001$) suggesting that the effect of Test Items on looking time differences was affected by the number of HPP experiments infants had participated in (Table 1, Fig. 2). In line with our predictions, the size of the difference between looking times on familiar and novel test items changed as a function of number of HPP visits.

We also found a significant main effect of the HPP predictor ($F(1,133.10) = 4.80$, $p = .030$) indicating that the Test Item by HPP interaction is mainly driven by a significant decrease in looking time to familiar items as the number of HPP visits increases. There was no evidence that a greater number of HPP visits was accompanied by longer looking to the novel item ($F(1,133.10) = 0.27$, $p = .606$).

Results also hold when reducing the data to infants with less than six HPP indicating that the interaction effect was not driven exclusively by participants with an unusually high number of visits [HPP 1-5: $F(1,99.00) = 10.29$, $p = .002$; HPP 1-4: $F(1,98.00) = 10.43$, $p = .002$; HPP 1-3: $F(1,92.00) = 4.56$, $p = .035$]. Notably, the interaction is significant even when sub-setting the data to the infants who participated in two HPP experiments only (HPP 1-2: $F(1,78.00) = 4.05$, $p = .048$).

Table 1

Summary of the results of the linear mixed-effects model. Degrees of freedom were approximated using the Kenward-Rogers approach, thus sometimes result in non-integers.

	Coefficient	***SEM***	**95% CI**	***F***	**Den. *df***	***p***
Intercept	7,679.11	673.14	6390.13, 9294.12	124.69	9.06	$< .001$
Test Item	-1,398.77	411.31	-2204.84, -589.11	11.57	100.00	.001
HPP	-539.70	238.69	-999.88, -74.83	4.80	133.10	.030
Test Item × *HPP*	667.11	192.64	247.23, 1028.51	11.99	100.00	.001

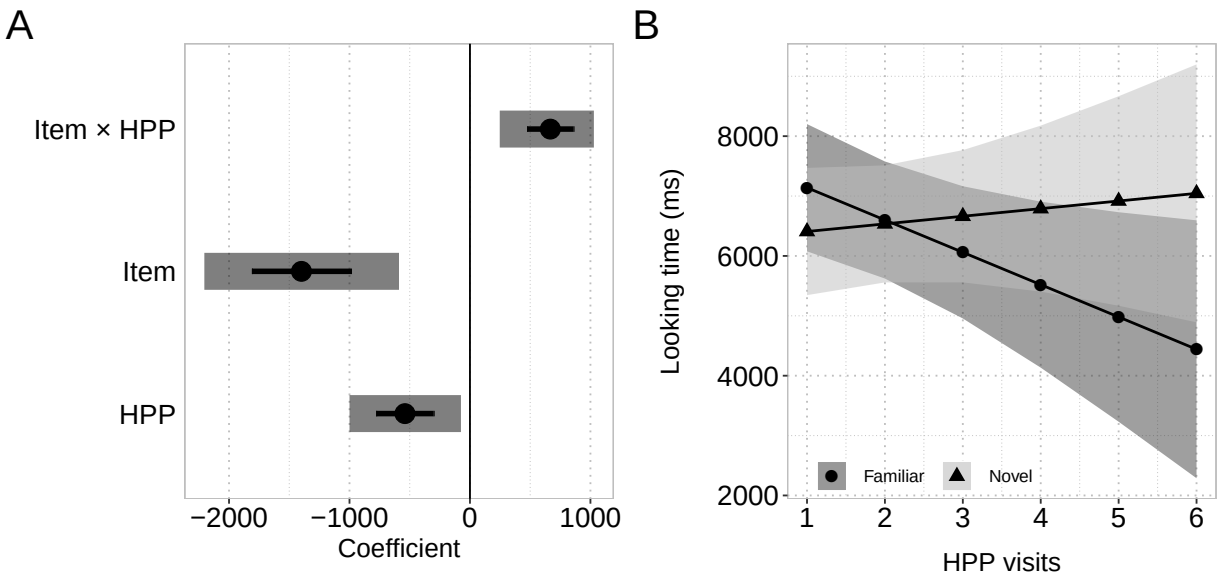

Figure 2. Panel A: graphic representation of the coefficients of the fixed effects of the model. Black dots represent the point estimate of the coefficient, black whiskers represent the standard error of the mean, and grey boxes represent the bootstrapped 95% confidence interval around the point estimate. Panel B: predicted looking time (in ms) for familiar and novel test items plotted against number of HPP visits. Black dots and lines represent the mean predicted looking time, and grey shaded areas represent the standard error of the mean.

Discussion

Results reported in this article are consistent with our hypothesis that experience with the Head-turn Preference Procedure affects direction of preference. The model combined four experiments with 12-month-old infants performing simple artificial grammar learning tasks, and showed that infants who had *not* previously experienced the HPP setting were more likely to show familiarity preferences than infants who had prior experience. One possible explanation for this finding relates to the structure of the HPP task. There are at least two types of information that must be simultaneously encoded by the infant at her first HPP experiment: 1) visual-auditory contingency (i.e., sounds appear contingently on the infant looking at the screen), and 2) the experiment stimuli (e.g., word sequences, sound streams). Therefore, infants have to engage in two concurrent learning when experiencing HPP for the first time: learning the structure of the HPP method, and solving the learning problem itself (e.g., discriminating sound strings following/breaking the grammar patterns). Such double-processing of information likely increases the task complexity, biasing results towards familiarity preferences. Infants who return to the lab for subsequent HPP experiments may be more able to focus on the learning problem, resulting in better learning as evidenced in novelty preferences.

It is important to notice that this effect may not just be limited to experiencing the HPP setting *per se*, but can be caused by the entire laboratory visit. When infants visit the lab for the very first time, they face a challenging situation: a new environment with new people interacting with them, testing rooms with a peculiar design (e.g., all-black or all-white walls with big screens), and novel sounds and images. This is a great amount of novel information for a young infant to process at once. In contrast, as infants come back to the lab for subsequent studies, the location, testing room and research staff may become more familiar, reducing the information load. The present study cannot discern which type of previous experience (HPP setting or lab) is responsible for the observed results.

Our findings have important implications for the interpretation of directions of preference in future studies. Infants' prior experience with a lab or research paradigm could account for distinct, and sometimes counterintuitive, patterns of preference. A related hypothesis suggests that less-common directions of preference for studies adressing a given topic (e.g., rule learning)

Our results are consistent with existing theories of cognitive development suggesting that, in spite of their limited capacities, infants 1) constantly gather input from the natural environment, 2) selectively sample the information to learn, and 3) direct their resources to examine the most relevant and informative input (e.g., Bates et al., 1996; Kidd, Piantadosi, & Aslin, 2014; Saffran & Kirkham, 2018; Santolin & Saffran, 2018). Our results suggest that infants actively process information about the lab environment and, consequently, their test performance are affected by how much lab experience they have accumulated. The learning outcome, in fact, seems to be constrained by the amount of novel information infants have to process in parallel when visiting the lab.

Evidence provided in this article has important implications for future interpretation of directions of preference. Infants' prior experience with the lab or a given research paradigm can account for different, and sometimes counterintuitive, patterns of preference. A related hypothesis suggests that less-frequent directions of preference with respect to the pattern of preferences shown in the literature of a given topic (e.g., rule learning) likely represent sign errors as opposed to true infant preferences (Bergmann, Rabagliati, & Tsuji, 2019, RR; Rabagliati, Ferguson, & Lew-Williams, 2019). While this may be the case, it is also possible that discrepancies in preferential looking are related to factors like those investigated in the current study: prrior experience with the testing environment. For this reason, discrepancies such as unexpected directions of preference may actually be meaningful and informative about the state of infant learners in specific studies.

thes results would allow to update the model of the dactors inducing different

patterns of preferences in infant studies (e.g., Hunter & Ames, 1988). Here we propose that the dimension *task complexity* could be expanded beyond the specific task content (e.g., how complex are the stimuli presented) to include infants' familiarity with the task. Our findings, in fact, suggest that the learning outcome of a given task is constrained by how much task experience infants have accumulated through prior lab visits. Therefore, the amount of novel information infants have to process in parallel during a study increases the task demands, and the likelihood of showing a familiarity preference. This may well include the novelty of the experimental paradigm. It would be of great interest to expand our findings to other preferential paradigms (e.g., infant-controlled preferential looking procedures, visual-world paradigms) to advance our understanding of how such experiences modulate infants' performance at test.

206 **References**

207 Aslin, R. N. (2007). What's in a look? *Developmental Science*, *10*(1), 48–53.
208 https://doi.org/10.1111/j.1467-7687.2007.00563.x

209 Bates, E., Elman, J., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K.
210 (1996). *Rethinking Innateness*. MIT press.

211 Bergmann, C., & Cristia, A. (2016). Development of infants' segmentation of words from
212 native speech: A meta-analytic approach. *Developmental Science*, *19*(6), 901–917.
213 https://doi.org/10.1111/desc.12341

214 Bergmann, C., Rabagliati, H., & Tsuji, S. (2019). What's in a looking time preference?
215 https://doi.org/10.31234/osf.io/6u453

216 Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of Early Language Discrimination
217 Abilities in Infants From Bilingual Environments. *Infancy*, *2*(1), 29–49.
218 https://doi.org/10.1207/S15327078IN0201_3

219 Colombo, J., & Bundy, R. S. (1983). Infant response to auditory familiarity and novelty.
220 *Infant Behavior & Development*, *6*(3), 305–311.
221 https://doi.org/10.1016/S0163-6383(83)80039-3

222 Dawson, C., & Gerken, L. (2009). From Domain-Generality to Domain-Sensitivity:
223 4-Month-Olds Learn an Abstract Repetition Rule in Music That 7-Month-Olds Do
224 Not. *Cognition*, *111*(3), 378–382. https://doi.org/10.1016/j.cognition.2009.02.010

225 Fantz, R. L. (1964). Visual Experience in Infants: Decreased Attention to Familiar
226 Patterns Relative to Novel Ones. *Science*, *146*(3644), 668–670.
227 https://doi.org/10.1126/science.146.3644.668

228 Ferguson, B., Franconeri, S. L., & Waxman, S. R. (2018). Very young infants learn
229 abstract rules in the visual modality. *PloS One*, *13*(1), e0190185.
230 https://doi.org/10.1371/journal.pone.0190185

231 Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial
232 structures from visual scenes. *Psychological Science*, *12*(6), 499–504.
233 https://doi.org/10.1111/1467-9280.00392

234 Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in
235 infant preference procedures. *Infant and Child Development*, *13*(4), 341–348.
236 https://doi.org/10.1002/icd.364

237 Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel
238 and familiar stimuli. In *Advances in infancy research, Vol. 5.* (pp. 69–95).
239 Westport, CT, US: Ablex Publishing.

240 Johnson, S. P., Fernandas, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., &
241 Slemmer, J. A. (2009). Abstract Rule Learning for Visual Sequences in 8- and
242 11-Month-Olds. *Infancy : The Official Journal of the International Society on*
243 *Infant Studies*, *14*(1), 2–18. https://doi.org/10.1080/15250000802569611

244 Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in
245 social psychology: A new and comprehensive solution to a pervasive but largely
246 ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69.
247 https://doi.org/10.1037/a0028347

248 Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in
249 fluent speech. *Cognitive Psychology*, *29*(1), 1–23.
250 https://doi.org/10.1006/cogp.1995.1010

251 Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks Effect in Infant
252 Auditory Attention. *Child Development*, *85*(5), 1795–1804.
253 https://doi.org/10.1111/cdev.12263

254 Rabagliati, H., Ferguson, B., & Lew-Williams, C. (2019). The profile of abstract rule
255 learning in infancy: Meta-analytic and experimental evidence. *Developmental*
256 *Science*, *22*(1), e12704. https://doi.org/10.1111/desc.12704

Roder, B. J., Bushneil, E. W., & Sasseville, A. M. (2000). Infants' Preferences for Familiarity and Novelty During the Course of Visual Processing. *Infancy*, *1*(4), 491–507. https://doi.org/10.1207/S15327078IN0104_9

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2004). Infant visual recognition memory. *Developmental Review*, *24*(1), 74–100. https://doi.org/10.1016/j.dr.2003.09.004

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*(2), 479–500. https://doi.org/10.1016/j.cognition.2007.10.010

Saffran, J. R., & Kirkham, N. Z. (2018). Infant Statistical Learning. *Annual Review of Psychology*, *69*(1), 181–203. https://doi.org/10.1146/annurev-psych-122216-011805

Saffran, J. R., & Wilson, D. P. (2003). From Syllables to Syntax: Multilevel Statistical Learning by 12-Month-Old Infants. *Infancy*, *4*(2), 273–284. https://doi.org/10.1207/S15327078IN0402_07

Santolin, C., & Saffran, J. R. (2018). Constraints on Statistical Learning Across Species. *Trends in Cognitive Sciences*, *22*(1), 52–63. https://doi.org/10.1016/j.tics.2017.10.003

Santolin, C., & Saffran, J. R. (2019). Non-Linguistic Grammar Learning by 12-Month-Old Infants: Evidence for Constraints on Learning. *Journal of Cognition and Development*, *20*(3), 433–441. https://doi.org/10.1080/15248372.2019.1604525

Santolin, C., Saffran, J. R., & Sebastian-Galles, N. (2019). Non-linguistic artificial grammar learning in 12-month-old infants: A cross-lab replication study. In. Potsdam, Germany.

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it? *Developmental Science*, *12*(6), 874–887.

https://doi.org/10.1111/j.1467-7687.2009.00829.x

Thiessen, E. D. (2012). Effects of inter- and intra-modal redundancy on infants' rule learning. *Language Learning and Development*, *8*(3), 197–214. https://doi.org/10.1080/15475441.2011.583610

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716. https://doi.org/10.1037/0012-1649.39.4.706

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
## Warning: package 'kableExtra' was built under R version 3.5.2
```

```
## Warning: package 'purrr' was built under R version 3.5.2
```

Appendix A

Experiments included in the linear mixed-effects model.

The selected experiments consist of an artificial grammar learning task with 12-month-old infants. These experiments are characterized by variability in the number of infants' prior HPP visits[1]. They include all studies run in the two senior authors' labs that included (a) 12- to 13-month-old participants; (b) HPP; (c) artificial grammar learning (linguistic or non-linguistic); (d) 2 to 5 minutes of exposure; (e) an *a priori* hypothesis that infants would show learning; (f) visit numbers recorded at the time of testing. The studies are thus as well matched as is possible given the retrospective nature of this analysis.

Saffran & Wilson (2003) demonstrated that 12-month-old infants can compute multiple regularities from a finite-state grammar. Infants were able to first segment words from running speech based on transitional probabilities, then detect permissible orderings of the segmented words. Test items consisted of grammatical and ungrammatical sentences that could only be discriminated based on word-level information (transitional probabilities between syllables were not informative about the "grammaticality" of test items). Infants showed a significant familiarity preference: $F(1, 38)= 5.37$, $p<.05$.

Saffran et al. (2008) demonstrated that infants could detect simple phrases (i.e., clusters of nonsense words grouped together based on statistical regularities) from artificial grammars. In Exp. 1, infants in the Predictive Language condition were familiarized with a grammar including predictive (statistical) dependencies between words. The test items consisted of familiar sentences vs. novel sentences violating the grammar. Infants showed a

[1] At the time of publication of Saffran & Wilson (2003), the first author noted that there appeared to be an association between the number of prior studies completed by the infants and the direction of preference. The analysis was included in the original manuscript submission but was removed from later revisions based on the manuscript reviews.

significant novelty preference: $t(11)= 2.52$, $p<.05$.

Santolin & Saffran (2019) is a conceptual replication of Saffran et al. (2008) using non-linguistic sounds (e.g., computer alert sounds) to implement the grammars. Infants exposed to the Predictive language showed a significant novelty preference: $t(26)=2.45$, $p=.021$, $d=0.47$.

We replicated the Predictive Language condition of Santolin & Saffran (2019) at the University Pompeu Fabra, Barcelona (Santolin et al., 2019), using identical stimuli and procedures. We found significant discrimination of the test stimuli but observed the opposite direction of preference: Infants listened longer to familiar than novel strings: $t(23)=2.30$, $p=.030$, $d=0.47$. All results are shown in Fig. 1 of the main manuscript.

Participants information

We retrieved data from 102 12-month-old infants who had participated in a range of 1-6 studies. Three of the studies were run in Madison, WI (University of Wisconsin-Madison): Saffran & Wilson, 2003 (Exp. 2): $N=40$, mean age: 11.5 months; Saffran et al., 2008 (Exp. 1, Condition P-Language): $N=12$), mean age: 12.8 months; Santolin & Saffran, 2019 (Condition 1); $N=26$, mean age: 12.9 months. One study was run in Barcelona, Spain (Universitat Pompeu Fabra): Santolin, Saffran & Sebastian-Galles, 2019: $N=24$, mean age: 13 months. Two data points (average looking time for familiar and novel test items) were available for each participant. Participants included in the current analysis are those included in the final version of the studies.

Linear mixed-effects model - additional information

We fit a model predicting looking time (LT) including $TestItem$ (Familiar vs. Novel), number of Head-turn Preference Procedure experiments completed by infants (HPP), and their interaction ($Item \times HPP$) as fixed effects. Participant and study (4 levels: Santolin

& Saffran (2019), Santolin et al. (2019), Saffran et al. (2008), and Saffran & Wilson (2003)) were included as random effects. Following Barr, Levy, Scheepers, & Tily (2013), we fit a model with a maximal random effects structure including random intercepts by-participant and by-study, and random slopes of HPP by-participant and by-study. However, due to lack of convergence, we pruned the random effects structure until convergence was achieved (e.g., Brauer & Curtin, 2018). The final model included by-participant and by-study random intercepts only. The particular random effects structure chosen does not qualitatively impact the estimates and conclusions from the model. This model accounts for cross-participant variability in overall looking time (as some infants look longer than others), and for cross-studies differences in overall looking time. The model was fit using the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015) from the R environment (R Core Team, 2018). We used the `Anova` function from the `car` R package (Fox & Weisberg, 2019) to perform F-tests on fixed effects using Kenward-Roger's approximation to degrees of freedom (e.g., Judd, Westfall, & Kenny, 2012).

Results sub-setting data to participants with less than 6, 5, 4, and 3 HPP studies

Consistent with the results of the entire dataset, we found a statistically significant interaction of *Test Item* with the number of *HPP* visits when reducing the sample to the infants who participanted in less than 6, 5, 4, and 3 HPP experiments. Below, we report the output of the linear mixed-effects model fitted on the original and reduced samples (Table A1, Fig. A1).

Table A1

Output of the Linear Mixed Effects-Model fitted on the subsetted data. Results are reported for the original analysis, and the HPP5, HPP4, HPP3, and HPP2 subsets. Degrees of freedom were approximated using the Kenward-Roger approach, this sometimes resulting in non integers.

Subset	**Term**	**Coefficient**	***SEM***	**95% CI**	**F**	**Den. *df***	***p***
Original	*Intercept*	7,679.11	673.14	[6390.13, 9294.12]	124.69	9.06	$< .001$
	Test Item	-1,398.77	411.31	[-2204.84, -589.11]	11.57	100.00	.001
	HPP	-539.70	238.69	[-999.88, -74.83]	4.80	133.10	.030
	Test Item × *HPP*	667.11	192.64	[247.23, 1028.51]	11.99	100.00	.001
HPP 1-5	*Intercept*	7,674.97	691.60	[6452.66, 9029.35]	118.33	10.07	$< .001$
	Test Item	-1,416.12	435.42	[-2237.65, -543.72]	10.58	99.00	.002
	HPP	-535.65	261.21	[-1081.11, -37.53]	3.97	133.57	.048
	Test Item × *HPP*	677.84	211.27	[241, 1081.95]	10.29	99.00	.002
HPP 1-4	*Intercept*	7,611.06	719.37	[6188.42, 9275.57]	107.35	10.46	$< .001$
	Test Item	-1,491.11	452.08	[-2348.39, -578.5]	10.88	98.00	.001
	HPP	-500.79	278.70	[-1070.22, 98.18]	3.04	131.92	.083
	Test Item × *HPP*	726.22	224.92	[294.24, 1144.99]	10.43	98.00	.002
HPP1-3	*Intercept*	7,470.05	794.61	[6007.13, 9172.54]	83.78	14.30	$< .001$
	Test Item	-1,349.93	532.33	[-2426.86, -267.64]	6.43	92.00	.013
	HPP	-395.84	366.62	[-1182.55, 316.11]	1.09	122.31	.299
	Test Item × *HPP*	627.92	294.09	[2.97, 1252.57]	4.56	92.00	.035
HPP 1-2	*Intercept*	7,301.91	1,009.89	[5252.93, 9362.1]	48.86	23.94	$< .001$
	Test Item	-1,783.67	726.69	[-3199.64, -360.67]	6.02	78.00	.016
	HPP	-261.34	586.82	[-1449.77, 991.74]	0.19	106.96	.667
	Test Item × *HPP*	969.50	481.79	[38.29, 2010.43]	4.05	78.00	.048

In all models, the $Item \times HPP$ interaction term was statistically significant [HPP 1-5: $F(1, 99) = 10.29$, $p = .002$; HPP 1-4: $F(1, 98) = 10.43$, $p = .002$; HPP 1-3: $F(1, 92) = 4.56$, $p = .035$; HPP 1-2: $F(1, 78) = 4.05$, $p = .048$] . These results provide evidence that the effect of HPP on the preference pattern we observed in our main analysis is not entirely dependent on any subsample of the HPP variable.

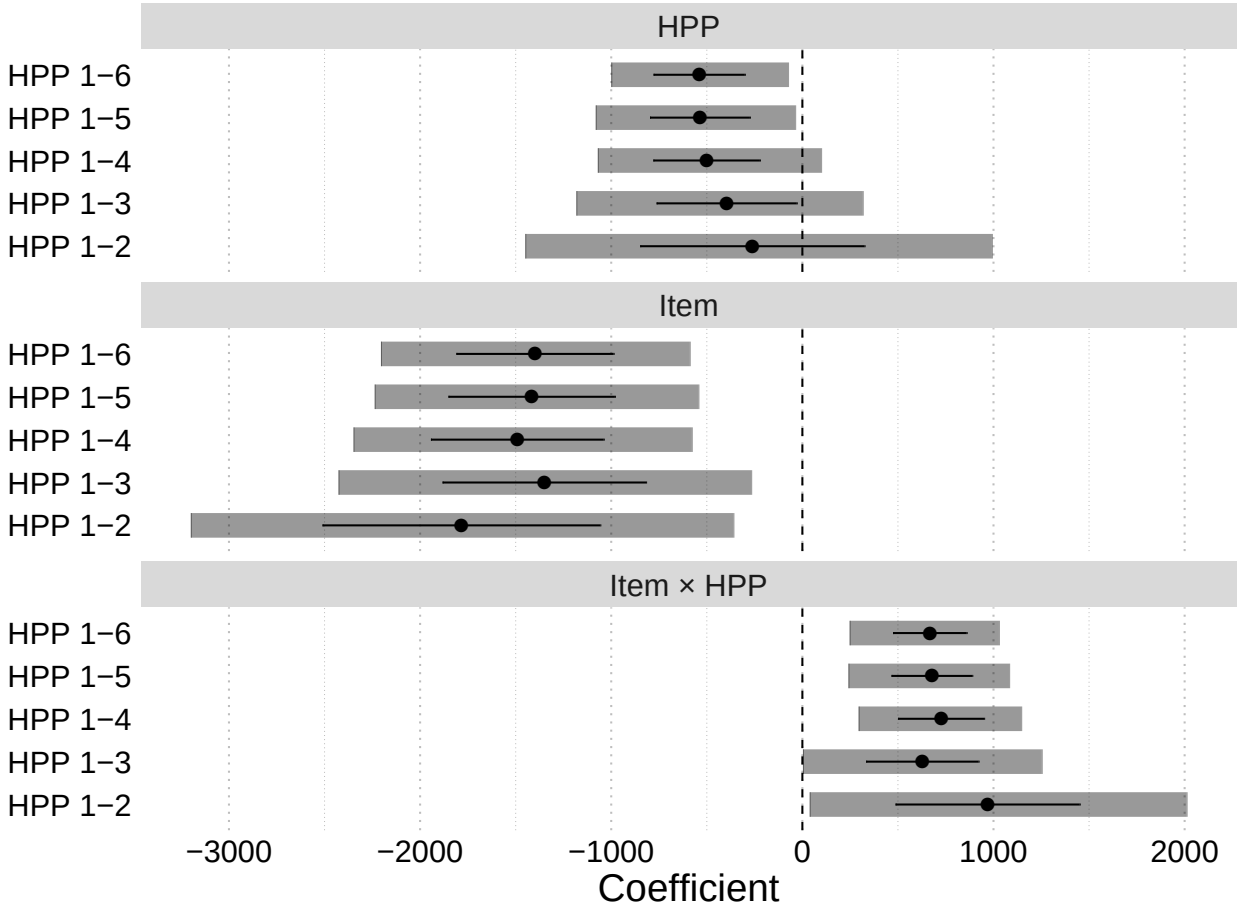

Figure A1. Graphical representation of the coefficients of the Test Item by HPP visits interaction term of the model fitted on the complete data-set (reported in the main manuscript), and of the models fitted on the reduced data-sets. Black dots represent the point estimate of the coefficient, black whiskers represent the standard error of the mean, and grey boxes represent the bootstrapped 95% confidence interval around the point estimate.

Appendix B

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

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

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/

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

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

Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*(2), 479–500. https://doi.org/10.1016/j.cognition.2007.10.010

Saffran, J. R., & Wilson, D. P. (2003). From Syllables to Syntax: Multilevel Statistical

388 Learning by 12-Month-Old Infants. *Infancy*, *4*(2), 273–284.
389 https://doi.org/10.1207/S15327078IN0402_07

390 Santolin, C., & Saffran, J. R. (2019). Non-Linguistic Grammar Learning by 12-Month-Old
391 Infants: Evidence for Constraints on Learning. *Journal of Cognition and*
392 *Development*, *20*(3), 433–441. https://doi.org/10.1080/15248372.2019.1604525

393 Santolin, C., Saffran, J. R., & Sebastian-Galles, N. (2019). Non-linguistic artificial
394 grammar learning in 12-month-old infants: A cross-lab replication study. In.
395 Potsdam, Germany.