

# Methods

## Participants

We retrieved data from 90 participants: 24 from Santolin & Saffran (2019) (26 tested in Wisconsin, the replication study [Santolin et al. (2019); 24] tested in Barcelona, and 40 from Saffran & Wilson (2003), tested in Wisconsin. Two datapoints were available for each participant: one for the mean looking time in “familiar” trials, and one for the mean looking time in “novel” trials. A total of 180 datapoints were included in the analysis. Participants included in our analysis were those included in the final version of both studies.

## Studies

The two studies conducted by Santolin and colleagues compared looking times in novel and familiar trials, and revealed that infants displayed a preference toward one of the trial types. The direction of the preference, however, was different for the two studies.

## Data analysis

We fit a model predicting looking time ( $LT$ ) from the fixed effects trial type ( $TrialType$ , centered; Familiar = -0.5 and Novel = 0.5), the number of HeadTurn Preference Procedure experiments completed by infants ( $HPP$ ), and their interaction ( $TrialType \times HPP$ ) as fixed effects. *Familiar* trials were set as the baseline. Participant ( $Participant$ ) and study ( $Study$ ) were included as random effects. Following Barr et al. (2013), we fit a model with the maximal random effects structure that allowed the model to converge. The maximal random effects structure included a by-participant and by-study random intercepts, and a random slope of  $HPP$  by  $Study$ . We subsequently pruned the random effects structure until convergence was achieved. The final model included by-participant and by-study random intercepts. The particular random effects structure chosen does not qualitatively impact the estimates and conclusions from the model

This model accounts for across-participants variability in overall looking time (i.e. some infants are long lookers, some are short lookers), and for across-studies differences in overall looking time, 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.

Term	Coefficient	95% CI	t	df	p
Intercept	7064.258	6037.34, 8091.18	-	-	-
Trial type	-318.405	-738.95, 102.14	2.202	88.000	0.141
HPP	-310.027	-824.58, 204.52	1.395	85.139	0.241
Trial type * HPP	665.558	208.22, 1122.9	8.136	88.000	0.005

We found a statistically significant interaction term,  $t(88) = 8.14$ ,  $p = 0.005$ , 95% CI = [208.22, 1122.9], 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.

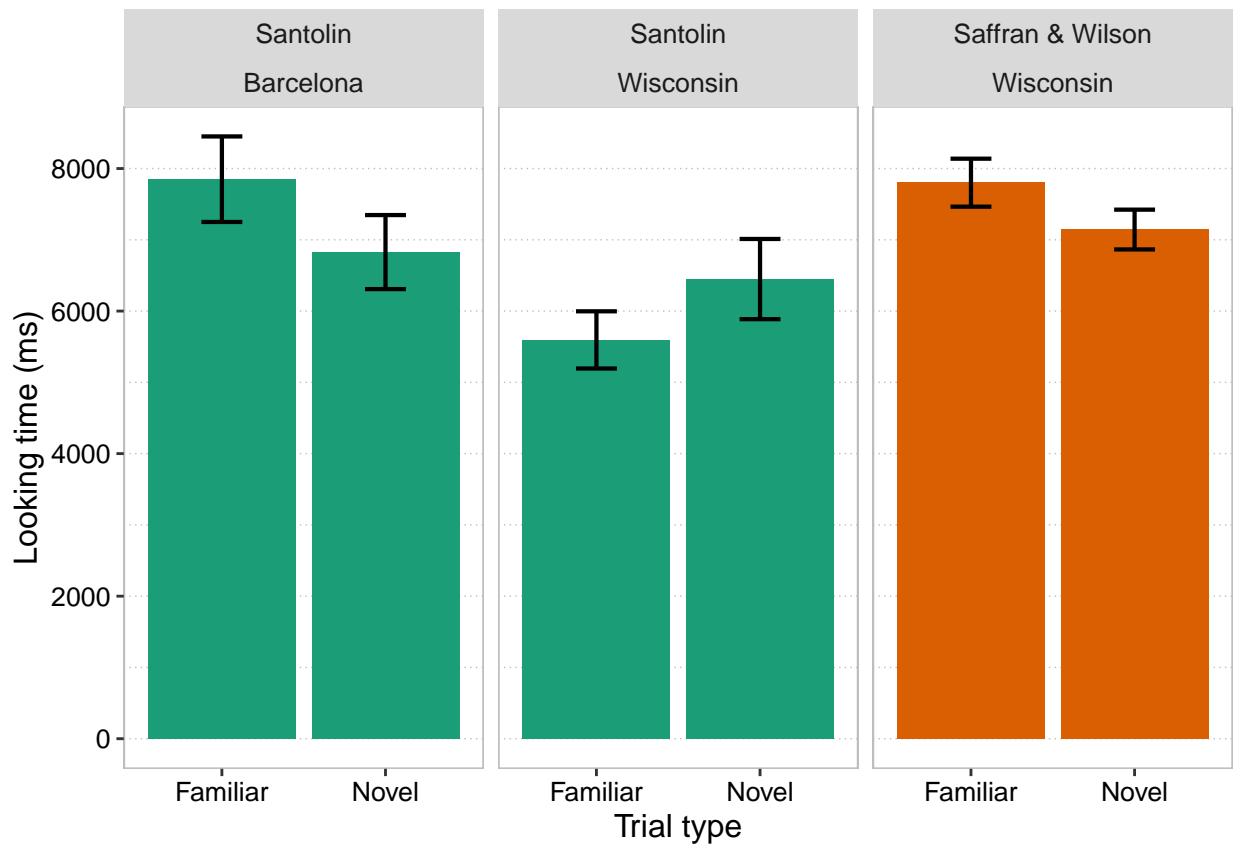


Figure 1: Looking time (ms) in familiar and novel trials split by Study and Location. Error bars indicate the standard error of the mean, respectively. Grey lines indicate participant level mean looking times.

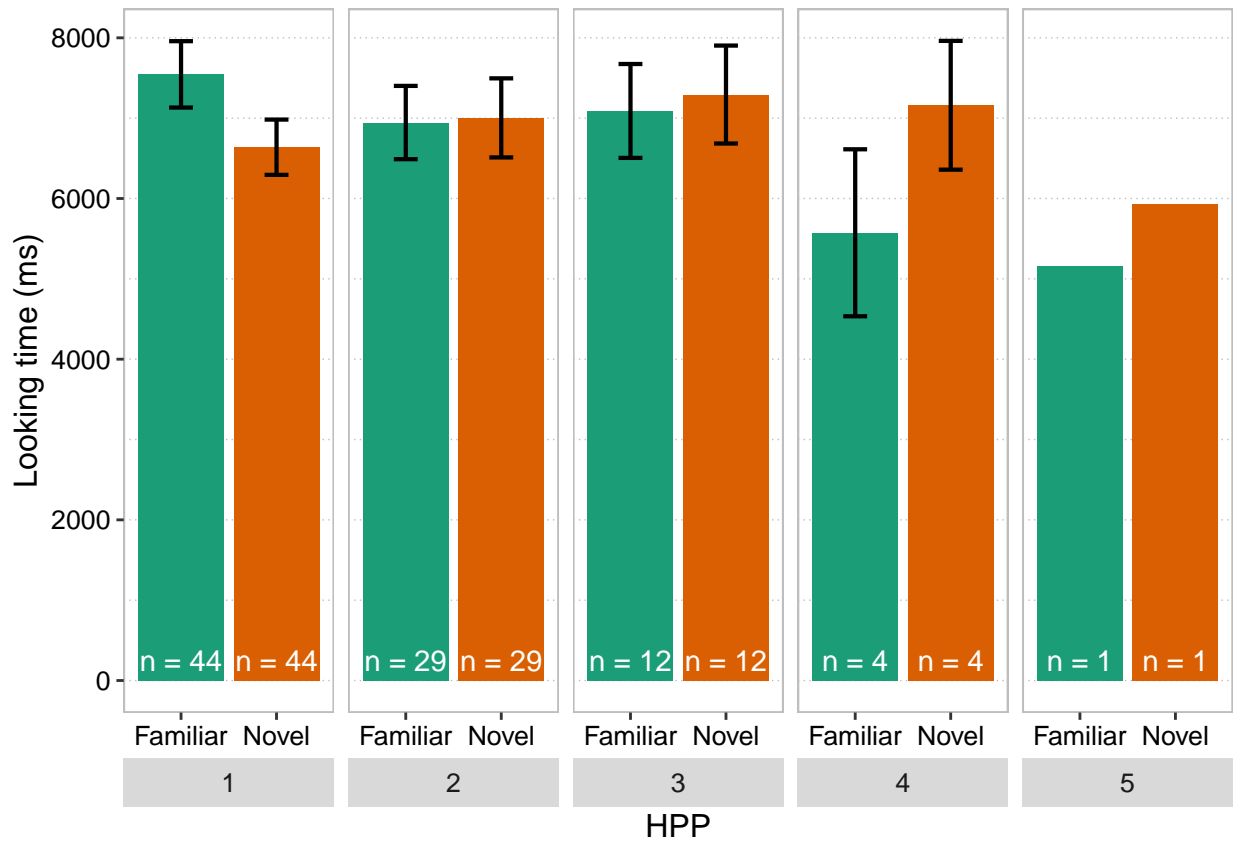


Figure 2: Looking times of familiar and novel trials, split by number of HeadTurn Preference Procedure experiments. Black points and error bars represent the group-level mean looking time and SEM, respectively. Grey lines represent participant-level mean looking time.

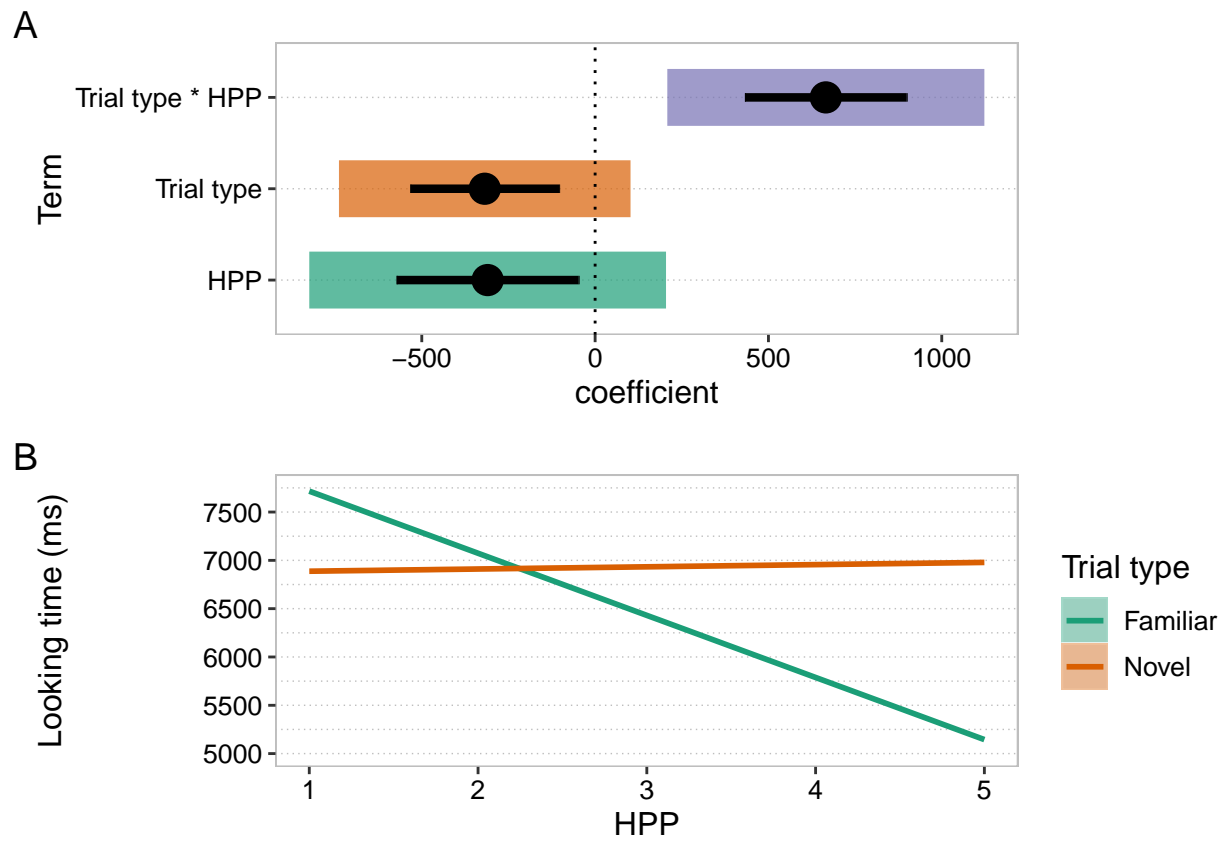


Figure 3: (A) Coefficients of the fixed effects. Dots, whiskers, and shaded boxes represent the estimated coefficients, standard error, and 95 percent Wald confidence intervals, respectively. (B) Predicted looking times plotted against HPP, split by study.

# Appendices

## Appendix 1: Linear Mixed Model

Following Barr, Levy, Scheepers, & Tily (2013) guidelines, we initially specified a maximal random structure, including random intercepts for *Participant* and *Study*, and random slopes for *TrialType*, *HPP*, and *TrialType*  $\times$  *HPP*, by *Participant* and *Study*. Due to lack of convergence, the random structure was simplified until the model fit was no longer singular. The code and formula of the mode are presented below:

```
lookingTime ~ TrialType * HPP + (1 | Participant) + (1 | Study) + (1 + HPP | Study)
```

$$\begin{aligned} LT_{ips} = & \\ & \beta_0 + Participant_{0p} + Study_{0s} + \\ & (\beta_1 + Study_s)HPP_i + \\ & \beta_2 TrialType_i + \\ & (\beta_3 + Study_s)HPP \times TrialType_i + \\ & e_{ips}, e_{ips} \sim N(0, \sigma^2), Participant_{0p} \sim N(0, \tau_{00}^2), Study_{0p} \sim N(0, \omega_{00}^2) \end{aligned}$$

Where:

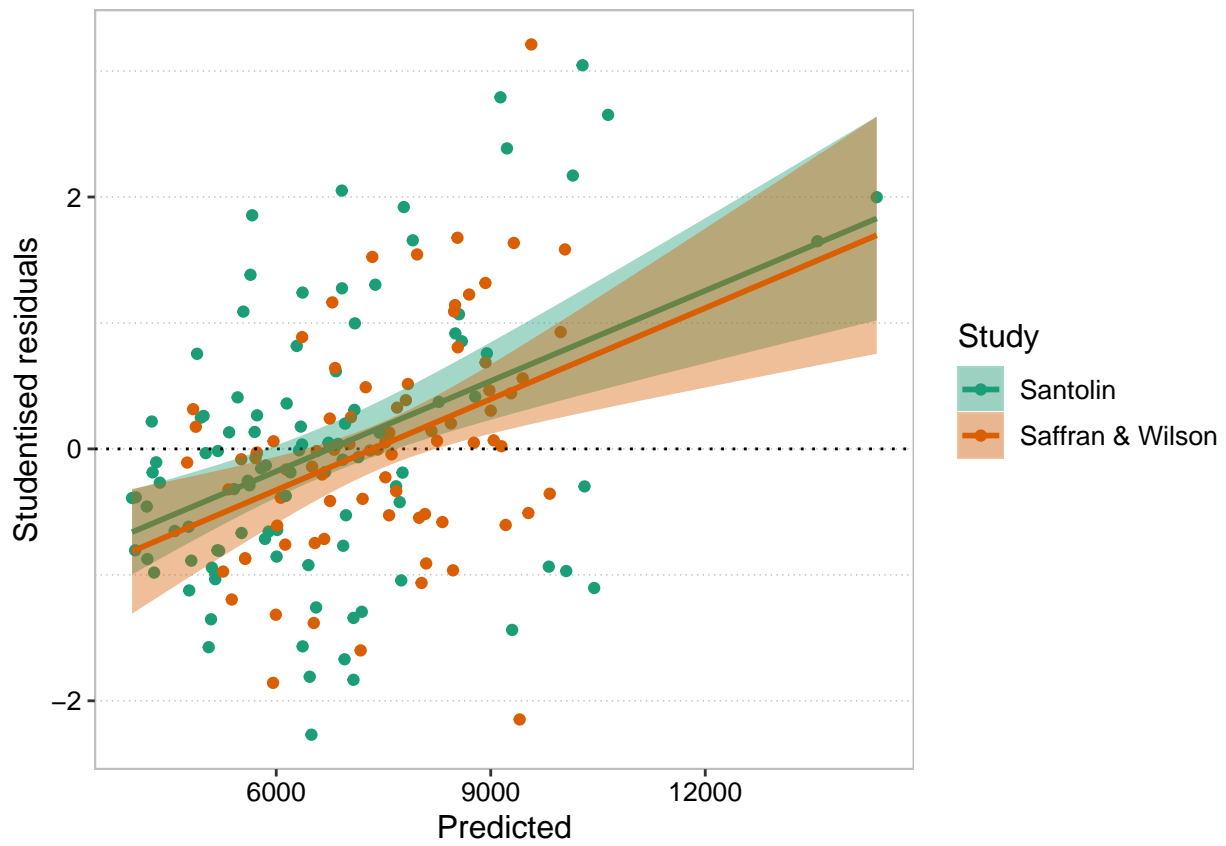
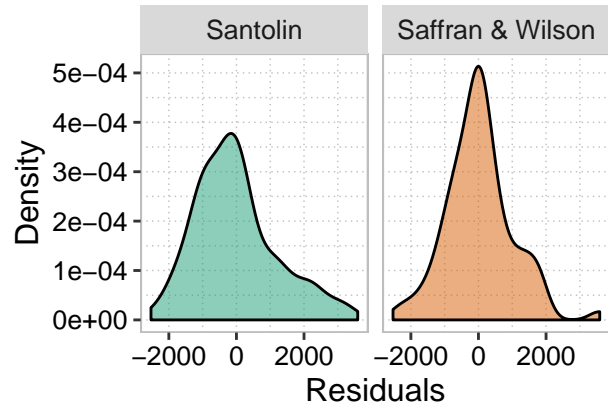
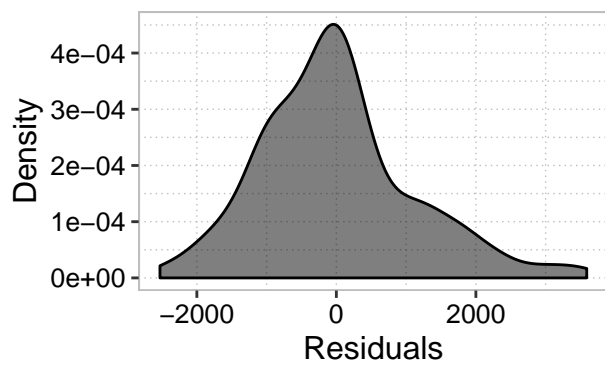
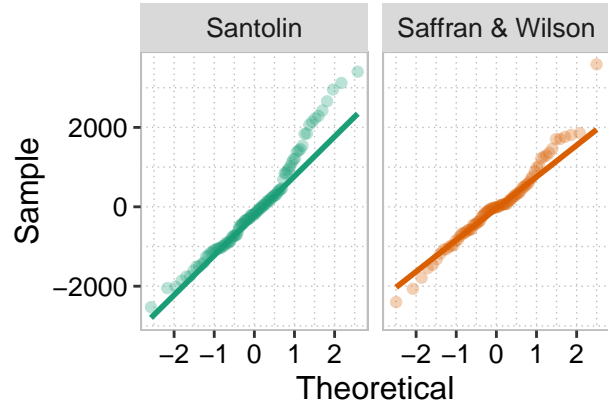
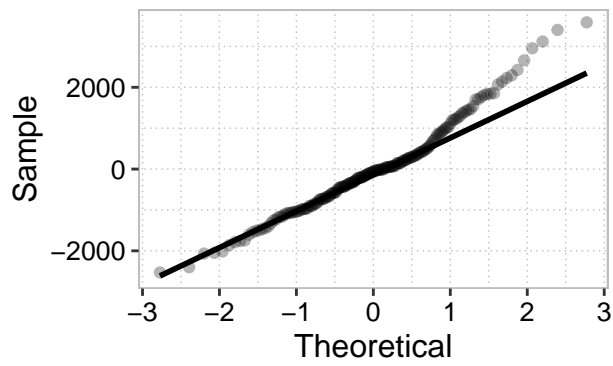
- $LT_{ips}$  is the looking time in trial  $i$ , in participant  $p$  from study  $s$ .
- $\beta_0$  is the fixed intercept (the grand mean of all looking times from all trials).
- $Participant_{0p}$  is the by-participant random intercept (the overall looking time of participant  $p$ , assumed to have been sampled from a normal distribution with mean 0 and variance  $\tau_{00}^2$ ).
- $Study_{0s}$  is the by-study random intercept (the overall looking time in study  $s$ , assumed to have been sampled from a normal distribution with mean 0 and variance  $\omega_{00}^2$ ).
- $\beta_1$  is the coefficient of the fixed effect of the *HPP* predictor.
- $\beta_2$  is the coefficient of the fixed effect of the *TrialType* predictor.
- $\beta_3$  is the coefficient of the fixed effect of the *TrialType*  $\times$  *HPP* interaction.
- $e_{ips}$  is the error of the model in trial  $i$ , of participant  $p$  from study  $s$ , assumed to be normally distributed with mean 0 and variance  $\sigma^2$ .

Before fitting the model, we group mean-centered the *HPP* predictor. The model was fitted using the **lme4** R package (Bates, Mächler, Bolker, & Walker, 2015). We used Maximum Likelihood criterion to estimate the coefficients, assuming an unstructured variance-covariance structure. We used the **lmerTest** package (Kuznetsova, Brockhoff, & Christensen, 2017) to compute  $p$ -values using the Satterthwaite's approximation to degrees of freedom (Satterthwaite, 1946).

Posterior predicted simulations (Gelman & Hill, 2006) indicated that the coefficients obtained by our model are plausible: The observed inter-quartile range (IQR) of looking times lied within the IQR obtained from 1000 simulated datasets generated from the model in more than 95% of the simulations ( $p = 0.443$ ).

## Appendix 2: Checking assumptions of the Linear Mixed Model

Residuals seem to approximate a normal distribution, though the some scores in the Santolin & Saffran (2019) and Santolin et al. (2019) studies seem to be somewhat shifted.



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

	VIF	Tolerance
Trial type	1	1
HPP	1	1
Trial type * HPP	1	1

## 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>
- Gelman, A., & Hill, J. (2006). *Data analysis using regression and Multilevel/Hierarchical models*. Cambridge University Press. Retrieved from <https://books.google.es/books?id=c9xLKzZWoz4C>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- 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](https://doi.org/10.1207/S15327078IN0402_07)
- 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.
- Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, 2(6), 110–114. <https://doi.org/10.2307/3002019>