|  |
| --- |
| Miquel Simonet1, Joseph V. Casillas, & Alex Aldrich3 |
| 1 University of Arizona |
| 2 Rutgers University |
| 3 Concordia College |
|  |

# Author note

Correspondence concerning this article should be addressed to Miquel Simonet, Postal address. E-mail: [my@email.com](mailto:my@email.com)

# Overview

## Bayesian data analysis

This study employs Bayesian Data Analysis for quantitative inferential statistics. Specifically, this implies that we use Bayesian *credible intervals* to draw statistical inferences. A Bayesian model calculates a posterior distribution, i.e., a distribution of plausible parameter values, given the data, a data-generating model, and any prior information we have about those parameter values. Posterior distributions are computationally costly. For this reason, we use the Hamiltonian Markov Chain Monte Carlo algorithm to obtain a sample that incldues thousands of values from the posterior distribution. In practical terms, what this means is that we do not calculate a single point estimate for an effect , but rather we draw a sample of 4,000 plausible values for . This allows us to quantify our uncertainty regarding by summaryzing the distribution of those values. We will use 3 statistics to describe the posterior distribution: (1) the mean, (2) a credible interval, and (3) the posterior probability that the effect is greater than 0. The mean provides a point estimate for the distribution, the 95% credible interval provides bounds for the effect, and the posterior probability that the effect is greater than zero further quantifies our uncertainty. For example, if a hypothesis states that , we judge there to be *compelling evidence* for this hypothesis if the mean point estimate is a positive number, is (by a reasonably clear margin) not included in the 95% credible interval of , and the posterior is close to one.

These three statistics provide an intuitive interpretation of any given effect. For instance, consider a case in which the posterior mean of is 100 and the 95% credible interval is [40, 160]. The interval tells us that we can be 95% certain the *true* value of is between 40 and 160, given the data, our model, and our prior information. Furthermore, the interval allows us to specify areas of uncertainty. In this example, we can conclude that the effect is almost certain to be positive. The lower interval value of 40 tells us that 95% of the plausible values are greater than 40. We also note that the interval covers a wide range of values, thus we also conclude that we are not very certain about the size of the effect. This type of interepretation is not possible under a frequentist paradigm.

## About this document

This document was written in RMarkdown using papaja (Aust & Barth, 2018) and serves as a project report for our research group. From this point foward, the document is written as if it were the results section of a future manuscript. This implies that it is written in a way that allows it to be copy and pasted into the actual manuscript once it is available.

# Method

## Participants

The data include 42 participants from 3 populations: monolingual English speakers, monolingual Spanish speakers, Bilingual English-Spanish speakers.

### Monolingual English speakers

The study ingludes 8 monolingual English speakers from the U.S. Southwest. Age and other info here.

### Monolingual Spanish speakers

The monolingual Spanish speakers come from Majorca, Spain. There are 8 of them. Age and other info here.

### Bilingual speakers

The English-Spanish bilinguals (n = 26) are from Southern Arizona, Northern Mexico. Age, BLP and other info. Note there are two samples from this population. The coronal dataset includes 17 speakers and the bilabial dataset includes 9 speakers.

## Metrics

F1/F2, voice onset time, relative intensity, center of gravity, standard deviation, skewness, kurtosis

## Procedure

Decide if we will present 3 separate experiments with 3 different methods sections.

## Statistical analyses

All analyses were conducted in R (R Core Team, 2019, version 3.6.0). We use Bayesian multilevel models fitted in Stan using brms (Bürkner, 2017, 2018, version 2.10.0). Bayesian Data Analysis (BDA) has become a welcome alternative to frequentist statistical analysis. See XXX and XXX for tutorials and in depth explanations related BDA in the psychological and speech sciences. For all models, the criterion was standardized, or converted to z-scores, in order to facilitate comparibility between metrics. Continuous predictors were also standardized and categorical predictors were sum-to-zero coded. Thus for all models the intercept represents the outcome variable at the average state. We use regularizing, weakly informative priors in all models (specifics below) with 4,000 iterations (2,000 warmup) running on 16 processing cores. We quantify our uncertainty regarding a given effect by reporting point estimates derived from the posterior predictive distribution, including the 95% Highest Density Credible Intervals (HDI). Additionally, we assume a negligible effect size of ± 0.1 (CITE COHEN) in order to establish a Region of Practical Equivalence (ROPE), for which we assess the proportion of the HDI that falls within this interval. Finally, we report the Maximum Probability of Effect (MPE), or the Probability of Direction, as the proportion of the posterior distribution that is of the median’s sign. We assume there to be compelling evidence for a given effect when the HDI of the posterior distribution does not contain 0 nor fall within the ROPE by a reasonable margin and the MPE is greater than 97%.

# Results

The results are divided into 4 subsections dealing with (1) phonetic development over time, (2) learning trajectories, (3) comparisons with native bilinguals, and (4) individual differences.

## Experiment 0: Vowels

![Figure 1: F1 and F2 of /a/ from monolingual speakers as a function of language (English, Spanish). Transparent points represent raw data. Solid points indicate posterior means ± 95% and 80% credible intervals.](data:application/pdf;base64,)

Figure 1: F1 and F2 of /a/ from monolingual speakers as a function of language (English, Spanish). Transparent points represent raw data. Solid points indicate posterior means ± 95% and 80% credible intervals.

![Figure 2: Model plot.](data:application/pdf;base64,)

Figure 2: Model plot.

## Experiment 1: Monolinguals

![Figure 3: VOT and burst metrics of coronal stops (/d/, /t/) from monolingual speakers as a function of language (English, Spanish). Transparent points represent raw data. Solid points indicate posterior means ± 99% and 80% credible intervals.](data:application/pdf;base64,)

Figure 3: VOT and burst metrics of coronal stops (/d/, /t/) from monolingual speakers as a function of language (English, Spanish). Transparent points represent raw data. Solid points indicate posterior means ± 99% and 80% credible intervals.

![Figure 4: Model plot.](data:application/pdf;base64,)

Figure 4: Model plot.

## Experiment 2: Bilinguals

![Figure 5: VOT and burst metrics of coronal stops (/d/, /t/) from bilingual speakers as a function of language (English, Spanish). Transparent points represent raw data. Solid points indicate posterior means ± 99% and 80% credible intervals.](data:application/pdf;base64,)

Figure 5: VOT and burst metrics of coronal stops (/d/, /t/) from bilingual speakers as a function of language (English, Spanish). Transparent points represent raw data. Solid points indicate posterior means ± 99% and 80% credible intervals.

![Figure 6: Model plot.](data:application/pdf;base64,)

Figure 6: Model plot.

## Experiment 3: Bilingual POA data

![Figure 7: VOT and burst metrics of voiceless stops from bilingual speakers as a function of language (English, Spanish), place of articulation (Coronal, Bilabial). Transparent points represent raw data. Solid points indicate posterior means ± 99% and 80% credible intervals.](data:application/pdf;base64,)

Figure 7: VOT and burst metrics of voiceless stops from bilingual speakers as a function of language (English, Spanish), place of articulation (Coronal, Bilabial). Transparent points represent raw data. Solid points indicate posterior means ± 99% and 80% credible intervals.

![Figure 8: Model plot.](data:application/pdf;base64,)

Figure 8: Model plot.

# Results

# Discussion

# References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from <https://github.com/crsh/papaja>

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. <https://doi.org/10.18637/jss.v080.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>

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Appendix A

test

| Metric | Parameter | Estimate | HDI | ROPE | MPE |
| --- | --- | --- | --- | --- | --- |
| F1 | Intercept | -0.014 | [-0.245, 0.219] | 0.650 | 0.550 |
|  | Language | 0.184 | [-0.001, 0.353] | 0.154 | 0.979 |
|  | Phoneme | -0.021 | [-0.096, 0.053] | 1.000 | 0.717 |
|  | Item rep. | 0.008 | [-0.074, 0.091] | 1.000 | 0.573 |
| F2 | Intercept | -0.094 | [-0.368, 0.176] | 0.467 | 0.759 |
|  | Language | 0.196 | [-0.037, 0.409] | 0.181 | 0.958 |
|  | Phoneme | -0.006 | [-0.09, 0.082] | 1.000 | 0.555 |
|  | Item rep. | 0.045 | [-0.033, 0.123] | 0.938 | 0.874 |

Appendix B

test

| Metric | Parameter | Estimate | HDI | ROPE | MPE |
| --- | --- | --- | --- | --- | --- |
| VOT | Intercept | 0.011 | [-0.124, 0.129] | 0.926 | 0.575 |
|  | Language | 0.666 | [0.567, 0.767] | 0.000 | 1.000 |
|  | Phoneme | -0.616 | [-0.691, -0.548] | 0.000 | 1.000 |
|  | F1 | 0.010 | [-0.03, 0.052] | 1.000 | 0.704 |
|  | F2 | -0.006 | [-0.048, 0.035] | 1.000 | 0.626 |
|  | Item rep. | -0.013 | [-0.044, 0.019] | 1.000 | 0.790 |
|  | Language x Phoneme | 0.111 | [0.042, 0.178] | 0.364 | 0.998 |
| RI | Intercept | -0.120 | [-0.392, 0.146] | 0.409 | 0.821 |
|  | Language | 0.077 | [-0.142, 0.29] | 0.564 | 0.768 |
|  | Phoneme | -0.097 | [-0.227, 0.038] | 0.518 | 0.925 |
|  | F1 | -0.113 | [-0.222, -0.001] | 0.393 | 0.977 |
|  | F2 | -0.221 | [-0.334, -0.105] | 0.000 | 1.000 |
|  | Item rep. | 0.058 | [-0.025, 0.147] | 0.855 | 0.914 |
|  | Language x Phoneme | 0.022 | [-0.108, 0.148] | 0.904 | 0.635 |
| COG | Intercept | 0.048 | [-0.195, 0.296] | 0.593 | 0.658 |
|  | Language | 0.638 | [0.41, 0.88] | 0.000 | 1.000 |
|  | Phoneme | -0.209 | [-0.276, -0.141] | 0.000 | 1.000 |
|  | F1 | -0.015 | [-0.065, 0.035] | 1.000 | 0.728 |
|  | F2 | 0.035 | [-0.025, 0.095] | 1.000 | 0.881 |
|  | Item rep. | -0.028 | [-0.074, 0.019] | 1.000 | 0.878 |
|  | Language x Phoneme | 0.058 | [-0.007, 0.127] | 0.916 | 0.957 |
| Kurtosis | Intercept | -0.117 | [-0.297, 0.072] | 0.430 | 0.901 |
|  | Language | -0.615 | [-0.781, -0.469] | 0.000 | 1.000 |
|  | Phoneme | 0.216 | [0.113, 0.327] | 0.000 | 1.000 |
|  | F1 | -0.028 | [-0.089, 0.034] | 1.000 | 0.823 |
|  | F2 | -0.008 | [-0.063, 0.051] | 1.000 | 0.608 |
|  | Item rep. | 0.013 | [-0.05, 0.073] | 1.000 | 0.668 |
|  | Language x Phoneme | -0.240 | [-0.348, -0.136] | 0.000 | 1.000 |
| SD | Intercept | 0.036 | [-0.213, 0.278] | 0.603 | 0.627 |
|  | Language | 0.497 | [0.281, 0.71] | 0.000 | 1.000 |
|  | Phoneme | -0.282 | [-0.397, -0.162] | 0.000 | 1.000 |
|  | F1 | -0.005 | [-0.063, 0.054] | 1.000 | 0.560 |
|  | F2 | 0.018 | [-0.041, 0.086] | 1.000 | 0.719 |
|  | Item rep. | -0.027 | [-0.093, 0.04] | 1.000 | 0.798 |
|  | Language x Phoneme | 0.205 | [0.087, 0.315] | 0.016 | 0.999 |
| Skewness | Intercept | -0.025 | [-0.197, 0.155] | 0.768 | 0.608 |
|  | Language | -0.502 | [-0.642, -0.37] | 0.000 | 1.000 |
|  | Phoneme | 0.296 | [0.167, 0.431] | 0.000 | 1.000 |
|  | F1 | -0.044 | [-0.108, 0.024] | 0.979 | 0.911 |
|  | F2 | 0.005 | [-0.05, 0.061] | 1.000 | 0.569 |
|  | Item rep. | 0.017 | [-0.036, 0.069] | 1.000 | 0.731 |
|  | Language x Phoneme | -0.251 | [-0.377, -0.123] | 0.000 | 1.000 |

Appendix C

test

| Metric | Parameter | Estimate | HDI | ROPE | MPE |
| --- | --- | --- | --- | --- | --- |
| VOT | Intercept | -0.075 | [-0.216, 0.072] | 0.638 | 0.850 |
|  | Language | 0.476 | [0.391, 0.565] | 0.000 | 1.000 |
|  | Phoneme | -0.622 | [-0.737, -0.497] | 0.000 | 1.000 |
|  | F1 | 0.001 | [-0.028, 0.031] | 1.000 | 0.536 |
|  | F2 | -0.009 | [-0.041, 0.021] | 1.000 | 0.705 |
|  | Item rep. | -0.003 | [-0.034, 0.026] | 1.000 | 0.575 |
|  | Language x Phoneme | -0.007 | [-0.085, 0.07] | 1.000 | 0.565 |
| RI | Intercept | -0.025 | [-0.226, 0.184] | 0.706 | 0.598 |
|  | Language | 0.135 | [0.026, 0.237] | 0.240 | 0.994 |
|  | Phoneme | -0.084 | [-0.158, -0.014] | 0.677 | 0.988 |
|  | F1 | -0.271 | [-0.377, -0.166] | 0.000 | 1.000 |
|  | F2 | -0.193 | [-0.248, -0.138] | 0.000 | 1.000 |
|  | Item rep. | 0.004 | [-0.043, 0.052] | 1.000 | 0.566 |
|  | Language x Phoneme | 0.005 | [-0.062, 0.071] | 1.000 | 0.555 |
| COG | Intercept | -0.107 | [-0.286, 0.073] | 0.464 | 0.885 |
|  | Language | 0.581 | [0.441, 0.721] | 0.000 | 1.000 |
|  | Phoneme | -0.218 | [-0.272, -0.164] | 0.000 | 1.000 |
|  | F1 | 0.020 | [-0.023, 0.063] | 1.000 | 0.821 |
|  | F2 | -0.021 | [-0.064, 0.018] | 1.000 | 0.842 |
|  | Item rep. | 0.020 | [-0.019, 0.055] | 1.000 | 0.852 |
|  | Language x Phoneme | -0.047 | [-0.11, 0.016] | 0.977 | 0.929 |
| Kurtosis | Intercept | 0.066 | [-0.095, 0.226] | 0.675 | 0.797 |
|  | Language | -0.596 | [-0.693, -0.498] | 0.000 | 1.000 |
|  | Phoneme | 0.254 | [0.182, 0.322] | 0.000 | 1.000 |
|  | F1 | -0.000 | [-0.051, 0.052] | 1.000 | 0.503 |
|  | F2 | 0.005 | [-0.044, 0.055] | 1.000 | 0.584 |
|  | Item rep. | -0.044 | [-0.089, 0.003] | 1.000 | 0.969 |
|  | Language x Phoneme | -0.110 | [-0.209, -0.012] | 0.408 | 0.984 |
| SD | Intercept | -0.127 | [-0.285, 0.025] | 0.363 | 0.948 |
|  | Language | 0.556 | [0.436, 0.68] | 0.000 | 1.000 |
|  | Phoneme | -0.220 | [-0.296, -0.147] | 0.000 | 1.000 |
|  | F1 | 0.003 | [-0.045, 0.051] | 1.000 | 0.548 |
|  | F2 | -0.013 | [-0.061, 0.031] | 1.000 | 0.706 |
|  | Item rep. | 0.028 | [-0.015, 0.07] | 1.000 | 0.909 |
|  | Language x Phoneme | 0.074 | [-0.011, 0.165] | 0.740 | 0.953 |
| Skewness | Intercept | 0.122 | [-0.041, 0.276] | 0.377 | 0.935 |
|  | Language | -0.514 | [-0.597, -0.427] | 0.000 | 1.000 |
|  | Phoneme | 0.306 | [0.225, 0.394] | 0.000 | 1.000 |
|  | F1 | -0.010 | [-0.056, 0.035] | 1.000 | 0.663 |
|  | F2 | -0.003 | [-0.047, 0.044] | 1.000 | 0.543 |
|  | Item rep. | -0.009 | [-0.057, 0.039] | 1.000 | 0.651 |
|  | Language x Phoneme | -0.202 | [-0.297, -0.113] | 0.000 | 1.000 |

Appendix D

test

| Metric | Parameter | Estimate | HDI | ROPE | MPE |
| --- | --- | --- | --- | --- | --- |
| VOT | Intercept | -0.094 | [-0.212, 0.023] | 0.550 | 0.944 |
|  | Language | 0.773 | [0.663, 0.878] | 0.000 | 1.000 |
|  | Place | 0.140 | [0.032, 0.242] | 0.211 | 0.996 |
|  | F1 | 0.017 | [-0.014, 0.048] | 1.000 | 0.853 |
|  | F2 | -0.006 | [-0.04, 0.029] | 1.000 | 0.628 |
|  | Item rep. | 0.002 | [-0.037, 0.039] | 1.000 | 0.538 |
|  | Language x Place | 0.025 | [-0.079, 0.133] | 0.934 | 0.686 |
| RI | Intercept | 0.211 | [-0.009, 0.428] | 0.139 | 0.970 |
|  | Language | 0.115 | [0.033, 0.192] | 0.341 | 0.997 |
|  | Place | -0.410 | [-0.622, -0.179] | 0.000 | 1.000 |
|  | F1 | -0.220 | [-0.309, -0.126] | 0.000 | 1.000 |
|  | F2 | -0.176 | [-0.252, -0.103] | 0.000 | 1.000 |
|  | Item rep. | 0.010 | [-0.062, 0.081] | 1.000 | 0.622 |
|  | Language x Place | 0.028 | [-0.055, 0.108] | 0.984 | 0.752 |
| COG | Intercept | -0.386 | [-0.52, -0.25] | 0.000 | 1.000 |
|  | Language | 0.303 | [0.196, 0.411] | 0.000 | 1.000 |
|  | Place | 0.647 | [0.522, 0.771] | 0.000 | 1.000 |
|  | F1 | 0.042 | [-0.005, 0.092] | 1.000 | 0.953 |
|  | F2 | -0.037 | [-0.083, 0.011] | 1.000 | 0.942 |
|  | Item rep. | 0.001 | [-0.049, 0.049] | 1.000 | 0.520 |
|  | Language x Place | 0.291 | [0.178, 0.396] | 0.000 | 1.000 |
| Kurtosis | Intercept | 0.433 | [0.275, 0.589] | 0.000 | 1.000 |
|  | Language | -0.233 | [-0.301, -0.163] | 0.000 | 1.000 |
|  | Place | -0.829 | [-0.969, -0.677] | 0.000 | 1.000 |
|  | F1 | 0.003 | [-0.04, 0.048] | 1.000 | 0.552 |
|  | F2 | 0.009 | [-0.042, 0.056] | 1.000 | 0.656 |
|  | Item rep. | -0.044 | [-0.092, 0.004] | 1.000 | 0.963 |
|  | Language x Place | -0.191 | [-0.259, -0.123] | 0.000 | 1.000 |
| SD | Intercept | -0.501 | [-0.633, -0.37] | 0.000 | 1.000 |
|  | Language | 0.234 | [0.145, 0.321] | 0.000 | 1.000 |
|  | Place | 0.782 | [0.661, 0.9] | 0.000 | 1.000 |
|  | F1 | 0.011 | [-0.034, 0.058] | 1.000 | 0.687 |
|  | F2 | -0.017 | [-0.066, 0.029] | 1.000 | 0.765 |
|  | Item rep. | 0.024 | [-0.026, 0.071] | 1.000 | 0.843 |
|  | Language x Place | 0.204 | [0.118, 0.291] | 0.000 | 1.000 |
| Skewness | Intercept | 0.546 | [0.35, 0.745] | 0.000 | 1.000 |
|  | Language | -0.126 | [-0.228, -0.027] | 0.286 | 0.992 |
|  | Place | -0.865 | [-1.042, -0.68] | 0.000 | 1.000 |
|  | F1 | 0.006 | [-0.059, 0.076] | 1.000 | 0.555 |
|  | F2 | 0.014 | [-0.03, 0.059] | 1.000 | 0.732 |
|  | Item rep. | -0.013 | [-0.066, 0.046] | 1.000 | 0.684 |
|  | Language x Place | -0.092 | [-0.192, 0.007] | 0.561 | 0.967 |