

Milestone 5

Diego Arias

4/3/2020

Contents

1 Overview of Replication Paper	1
2 Beautiful? Graphic	2
3 Table	2
4 Appendix	3
5 What worked, what did not work	3
6 Proposed Extension	3
7 References	4

1 Overview of Replication Paper¹

In their paper, “Emotional Arousal Predicts Voting on the U.S. Supreme Court”, Dietrich and colleagues study over 3,000 hours of Supreme Court audio recordings to try to assess if judges implicitly reveal their leanings during the oral arguments. The predictor in their analysis is “Pitch Difference”, which was calculated by subtracting the judges’ vocal pitches in questions directed to the respondents by their pitches in questions directed to the respondents. It is important to note that the judges’ vocal pitch at different instances was quantified in the number of standard deviations above or below his/her average vocal pitch in the trial. The outcome variable is the judge’s final vote (more specifically if he votes in favor of the petitioner (1) or against (0)).

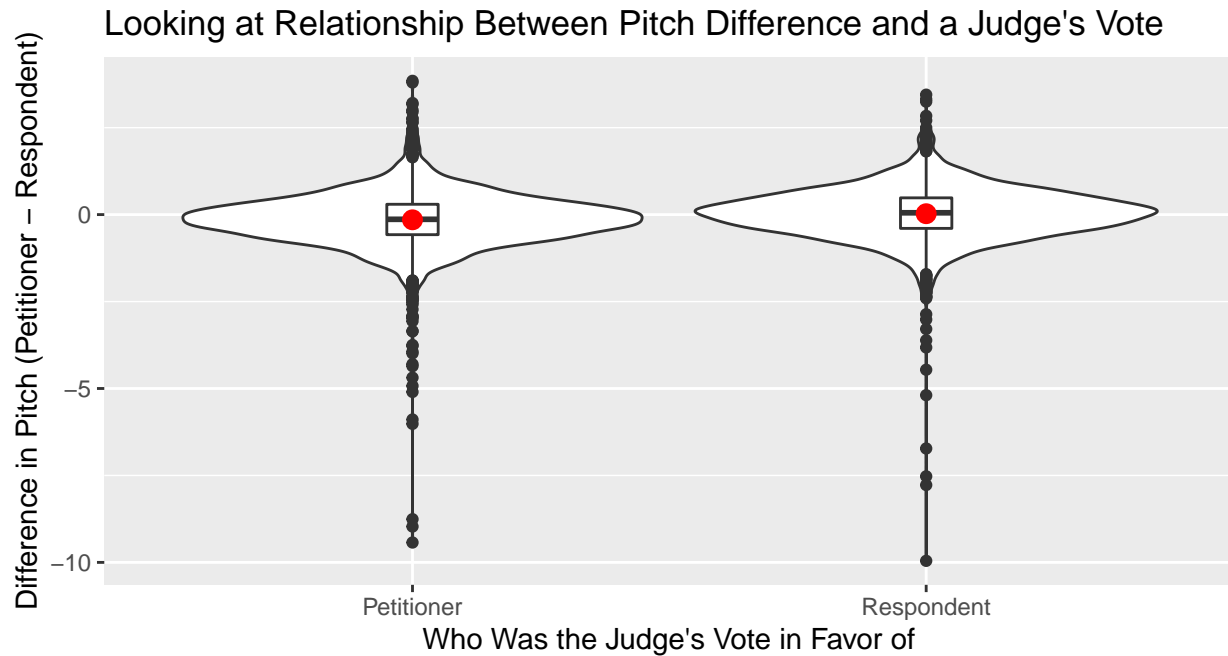
In their effort to find study the effect of vocal pitch on voting, the authors use a multilevel logistic regression model that uses pitch difference as a predictor for voting behavior. Ultimately, the authors find that the more emotionally aroused a justice’s questions are at an attorney compared to his/her opponent, the less likely that attorney is to win the Justice’s vote. Interestingly, the model goes on to show that if the vocal pitch of the questions directed to both the petitioner and respondents is the same, the probability of a justice voting for a petitioner is .55. The model goes on to show the effect of vocal pitch because if the difference between the pitches of questions directed to the petitioner is one standard deviation higher than that related to the respondent (meaning the judge is more emotionally aroused) the probability of a justice voting for a petitioner drops to .48.

This multilevel logistic regression model ends up predicting 57.5% of the judge’s votes accurately and 66.55% of overall case outcomes accurately, which is incredible judging how the model just uses one implicit measure to predict something as complicated as voting behavior. To put this in perspective, the authors of the study compared their model to a widely known one called {Marshall}+, which uses 95 predictors and successfully predicts 64.76% of Supreme Court Cases.

¹Please refer to the Github repository of my final project for further information., Github repository I make use of Xie (2019) and Wickham (2019)

This project will aim to replicate the aforementioned logistic regression model (and others included to answer secondary questions in the study). It will also take the analysis a step further by providing a new avenue to analyze the already collected data.

2 Beautiful? Graphic



While hard to see in the boxplot, judges who vote in favor of the petitioner will average have a more negative difference in pitch (which is measured as the pitch toward the petitioner minus the pitch toward the respondent) compared to judges who vote in favor of the respondent

Data from Dietrich et al. (2011)

3 Table

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
## Approximation) [glmerMod]
## Family: binomial ( logit )
## Formula: petitioner_vote ~ pitch_diff + (1 | justiceName)
## Data: sc
##
##      AIC      BIC    logLik deviance df.resid
##  7109.4   7129.1  -3551.7   7103.4     5206
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -4.0233 -1.0556  0.7648  0.9101  1.6206
##
## Random effects:
## Groups      Name      Variance Std.Dev.
```

```
## justiceName (Intercept) 0.02598 0.1612
## Number of obs: 5209, groups: justiceName, 18
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.17811    0.05469   3.256 0.00113 **
## pitch_diff  -0.26596    0.03617  -7.354 1.93e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## pitch_diff 0.043
```

4 Appendix

5 What worked, what did not work

I was able to replicate all of the main analysis, which was found in the finalizing results file that the authors provided. This did not take a while, as there was little main analysis done in the paper and there was only one table provided (and the code was able to smoothly run on my computer) Thus, my goal is to also look at some of the supplemental figures and tables (which are now in Python) and replicate them because it feels like the main analysis is not enough. I will do this after meeting with Alice!

6 Proposed Extension

I have a few ways where I think I could further take this paper and would love to talk about the pros and cons of them.

For one, I think there must be a lot of individual differences with how much the justice's pitch difference can be used to predict voting behavior. Namely, some judges's pitch may be more revealing for whatever reason, while others may be able to better control their pitch and thus make it harder to predict their voting outcome from it. A way to test this out would be to use a `stan_glm` model that models voting behavior by using the pitch difference and justice variables (with an interaction). The interaction coefficient would be able to tell us how much the effect that pitch difference depends has on voting varies depending on the justice. I could then make a graph that is similar to the one on the supplemental figure 4, which looks at this effect for each judge. Still, the fact that this was done in the supplemental figure makes me feel like this possible extension isn't original enough.

Thus, an idea that came to mind is to add a column in the data which codes the gender of the judge, and then see if there are any differences in how vocal pitch predicts voting across male and female justices, and then create a plot that demonstrates the difference (or similarity). This stems from learning that women on average have a broader vocal pitch, which might mean that there is actually a bigger opportunity for the vocal pitch of the female justices to vary, and thus might better predict voting behavior. Another possible idea is to see how this effect changes depending on the age of the justice. Maybe the older justice's are, the more experience they have and the better than can control their voice to seem neutral even when they have already made up their mind. This could easily be done by including the date of birth as a separate column and subtracting this number from the year of the case to get how old the justice was during the case.

Truthfully, I need to better understand Black et al's 2011 paper and also the current papers coding of pleasant and unpleasant words to better understand model's 2-4. After doing this, I might be able to propose some way to create a better model for voting behavior. This might just be combining one of these models with

additional predictors which I find to be important, whether they are the specific justice, the gender of the justice, or the age of the justice (as previously explained). This is because models 2-4 have many more variables than the original model, but at most have a predictive power of only 7 percentage points higher. Maybe I can find a way to better code for the meaning and pleasantness of the words to make a more predictive model, or maybe its the case that the vocal pitch already reflects a lot of the positiveness the words (which I would be able to test with a simple correlational model between the two predictors)

7 References

- Dietrich, Bryce J., Ryan D. Enos, and Maya Sen. “Emotional arousal predicts voting on the US supreme court.” *Political Analysis* 27.2 (2019): 237-243
- Miller, Gerald R., et al. “The effects of videotape testimony in jury trials: Studies on juror decision making, information retention, and emotional arousal.” *BYU L. Rev.* (1975): 331.
- Oliver, Edward, and William Griffitt. “Emotional arousal and ‘objective’ judgment.” *Bulletin of the Psychonomic Society* 8.5 (1976): 399-400
- Wickham, Hadley. 2019. *Stringr: Simple, Consistent Wrappers for Common String Operations*. <https://CRAN.R-project.org/package=stringr>.
- Xie, Yihui. 2019. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://CRAN.R-project.org/package=knitr>.