Emotional Arousal Predicts Voting on the U.S. Supreme Court

Bryce J. Dietrich* Ryan D. Enos† Maya Sen‡

First Draft: August 24, 2015
This Draft: December 18, 2017

Abstract

Do judges telegraph their preferences during oral arguments? Using the U.S. Supreme Court as our example, we demonstrate that Justices implicitly reveal their leanings during oral arguments, even before arguments and deliberations have concluded. Specifically, we extract the emotional content of over 3,000 hours of audio recordings spanning 30 years of oral arguments before the Court. We then use the level of emotional arousal in each of the Justices' voices during these arguments to accurately predict many of their eventual votes on these cases. Our approach yields predictions that are statistically and practically significant and robust to including a range of controls; in turn, this suggests that subconscious vocal inflections carry information that legal, political, and textual information do not.

*Department of Political Science, University of Iowa, 341 Schaeffer Hall, Iowa City, IA 52242 (bryce-dietrich@uiowa.edu, http://www.brycejdietrich.com).

†Department of Government, Harvard University, 1637 Cambridge St., Cambridge, MA 02138 (renos@gov.harvard.edu, http://ryandenos.com).

‡Corresponding Author. John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138 (maya_sen@hks.harvard.edu, http://scholar.harvard.edu/msen).

Introduction

In February of 2017, the Trump Administration issued an executive order that banned the entry of people from seven majority-Muslim countries into the U.S. and unleashed chaos in airports all over the world. Legal challenges were swift and, within two weeks, the 9th Circuit Court of Appeals had scheduled oral arguments concerning the order. Public interest in how the court would rule was significant. More than 130,000 people listened to the arguments, and hundreds of experts weighed in on how the judges would vote. Many of these predictions relied on the three judges' emotional reactions and vocal expressions during oral argument. For example, the *New York Times* provided a live analysis of the judges' reactions, assessing whether they were "pretty skeptical" or "friendly."[1] Ultimately, the three-judge panel ruled unanimously against the White House, but it was not without days of painful uncertainty for those affected.

In this paper, we do what observers of the 9th Circuit oral argument were attempting by asking whether we can systematically detect how judges will vote based on emotional responses at oral argument. We address this by using as our context the U.S. Supreme Court, which has decades of audio data. Specifically, we examine the audio from nearly 3,000 hours of oral arguments from the last 30 years and find that *vocal pitch alone* is strongly predictive of Supreme Court Justices' votes. The results are robust to the inclusion of other factors and predict outcomes at least as well as more complex models accounting for substantive features of the actors and cases, suggesting that vocal pitch predicts decisions in ways that characteristics like ideology or legal issue areas do not. In results presented in the Supplemental Information, moreover, we extend our findings to the 9th Circuit's Executive Order argument, suggesting that our approach has good external validity.

As we discuss below, our findings are consistent with several causal narratives. For example, it could be the case that Justices actively rely on their emotions in reaching important

[1] `https://www.nytimes.com/interactive/2017/02/07/us/ninth-circuit-oral-arguments-trump-immigration.html`.

decisions. It might also be the case that judges experience aroused responses during oral arguments because they receive information conflicting with previously made decisions. Although we cannot disaggregate these explanations, our results clearly show that non-substantive and implicit signals, even among elite actors such as federal judges and Supreme Court justices, can provide additional meaningful information on their attitudes beyond what can be found in their explicit pronouncements.

Description of Supreme Court Oral Arguments Audio Data, Emotional Arousal, and Vocal Pitch

We are not the first to suggest that emotion plays an important role in judicial oral arguments (e.g., Johnson et al., 2009; Shullman, 2004). For example, in Black et al. (2011), the authors argue that emotional displays commit Supreme Court Justices to certain courses of action. Consequently, the "tenor" of oral arguments can be used as a barometer of how Justices will rule in a given case. Even though others have utilized the number of questions directed towards each side (Roberts, Jr., 2005; Epstein, Landes and Posner, 2010), Black et al. (2011) captured the emotions expressed during oral arguments using the number of "pleasant" and "unpleasant" words. Ultimately they found the more unpleasant words Justices use towards an attorney, the less likely that attorney will prevail in the case. Even though we acknowledge the importance of what Justices say, we suggest *how* they say those words is of equal, if not greater, import.

First, changes in vocal inflections, like pitch, often occur unbeknownst to the speaker (Ekman et al., 1991). For the Justices, emotional arousal becomes more likely, especially when interacting with someone with whom they disagree. When this occurs, the heart begins to race, palms begin to sweat, and all muscles, including the vocal cords, tighten (Posner, Russell and Peterson, 2005). This is the primary reason why mean F_0 has been consistently shown to be associated with emotional activation (Calvo and D'Mello, 2010). Higher levels of emotional arousal are associated with higher levels of vocal pitch and amplitude; in con-

trast, emotions associated with low levels of physiological arousal are associated with lower mean F_0, F_0 variability, and vocal intensity (lower decibel level), as well as decreases in F_0 over time (Mauss and Robinson, 2009). Due to the automatic nature of this response, a speaker's vocal pitch will often provide insights into a speaker's level of activation beyond their conscious communication. In turn, this suggests that such unconscious reactions are indicative of Justices' ultimate preferences towards a case. This is the case regardless of whether the Justice formulates her response contemporaneously or whether she is reacting on the basis of predispositions about the case.

Second, some Justices, like Antonin Scalia, are more willing to express emotion as compared to others. Such differences are problematic for text-based measures since they only capture emotion that is verbalized. According to Russell (2003), this ignores a range of emotions which occur prior to conscious awareness. As analogy, consider felt body temperature. Even though our body's temperature changes all the time, we do not always identify those changes as being *hot* or *cold*. For some, a small decrease in temperature may be enough to say, "I am cold!" For others, that same decrease may not even be recognized. Emotional expression on the Supreme Court functions in a similar way. For some Justices, an attorney's error may be egregious enough to warrant calling it "idiotic," while for others Justices that same error may not even raise an eyebrow, verbal or otherwise. Text-based measures are extraordinarily useful when one is interested in understanding the former, but struggle with latter. Our study hopes to gain a better understanding of these more subtle forms of emotional expression, such as changes in vocal pitch.

To achieve this end, we collected audio recordings from oral arguments in 1,773 cases, beginning in 1982 and ending in 2014. Using the timestamps provided by the Oyez Project,[2] we further parsed these cases into discrete segments of audio uttered by (1) the Justices themselves, (2) the lawyer/s representing the petitioner, and (3) the lawyer/s representing

[2]https://www.oyez.org. More information on the methodology behind the audio data acquisition is provided in the Supplemental Information.

the respondent.[3] Lawyers spoke for 2,137 hours. Justices spoke for 502 hours. For the Justices, this represented 146,335 discrete utterances. Additional descriptive statistics are provided in the Supplemental Information, Tables S1 – S2.

Results: How Emotion Arousal Predicts Supreme Court Justices' Voting

We expect that a Justice who is more emotionally activated when speaking towards an attorney will be more likely to vote against that attorney. If this is correct, a higher vocal pitch will predict a stronger negative response. To analyze this, we code whether a Justice votes in favor of the petitioner, a 1 or 0 variable.[4] We construct a measure of "Pitch Difference" by subtracting vocal pitch in questions directed towards petitioners from vocal pitch in questions directed toward respondents. (Vocal pitch was measured using *Praat*, a speech synthesis program that estimates the fundamental frequency by dividing the autocorrelation of a windowed signal by the autocorrelation of the window itself. More details can be found in the Supplemental Information.) For each Justice, we converted vocal pitch to standard deviations above and below his or her average vocal pitch, which accounts for systematic differences between Justices (for example between male and female Justices).[5]

The main results are presented in Table 1, Model 1. These results show *that the higher emotional arousal or excitement directed at an attorney compared to his or her opponent, the less likely that attorney is to win the Justice's vote* ($p < 0.001$).[6] From Model 1, when the vocal pitch of questions directed to both sides is the same, the predicted probability

[3]The "petitioner" is the party bringing the case to the Court; the "respondent" is the party responding to the petitioner's claim. Both parties are represented by separate counsel, often highly experienced attorneys drawn from the elite group of appellate attorneys that comprise the Supreme Court bar.

[4]Since the outcome is 1 (Justice votes for the petitioner) or 0 (Justice votes against the petitioner), we use a multilevel logistic regression, which we implemented here via the `lme4` package in `R` statistical software language.

[5]In general, longer vocal cords produce a lower vocal pitch. Since male vocal cords are typically longer (17.5–25mm) than female vocal cords (12.5–17.5mm), men tend to naturally talk at a lower vocal pitch.

[6]Since we cannot assume that Justices' votes within cases are independent, we include a randomly varying intercept for each Justice. This additional parameter also helps account for other within-Justice differences.

Table 1: Does Vocal Pitch Predict Votes in Favor of the Petitioner?

	No Controls	DAL	Harvard IV	LIWC
	(1)	(2)	(3)	(4)
Fixed Effects				
Constant	0.178***	−0.025	−0.027	−0.026
	(0.055)	(0.160)	(0.160)	(0.160)
Pitch Difference	−0.266***	−0.214***	−0.215***	−0.214***
	(0.036)	(0.038)	(0.038)	(0.038)
Percent More Unpleasant Words Directed at Petitioner		−1.971	0.071	−2.129
		(1.471)	(0.846)	(1.308)
Percent More Pleasant Words Directed at Petitioner		−1.647	0.272	−1.673
		(1.086)	(0.685)	(1.046)
Number More Questions Directed at Petitioner		−0.057***	−0.057***	−0.057***
		(0.008)	(0.008)	(0.008)
Political Ideology$_{t-1}$		0.158***	0.158***	0.158***
		(0.033)	(0.033)	(0.033)
Lower Court Decision Was Conservative		0.011	0.012	0.012
		(0.073)	(0.073)	(0.073)
Political Ideology$_{t-1}$ × Lower Court Decision Was Conservative		−0.263***	−0.264***	−0.263***
		(0.034)	(0.034)	(0.034)
Solicitor General as Amicus Supporting Petitioner		0.540***	0.543***	0.540***
		(0.079)	(0.079)	(0.079)
Solicitor General as Amicus Supporting Respondent		−0.672***	−0.666***	−0.673***
		(0.104)	(0.104)	(0.104)
Number of Amicus Briefs Supporting Petitioner		0.039***	0.039***	0.039***
		(0.008)	(0.008)	(0.008)
Number of Amicus Briefs Supporting Respondent		−0.058***	−0.058***	−0.058***
		(0.007)	(0.007)	(0.007)
Petitioner's Level of Resources		0.045***	0.046***	0.045***
		(0.014)	(0.014)	(0.014)
Respondent's Level of Resources		−0.003	−0.003	−0.003
		(0.014)	(0.014)	(0.014)
Random Effects				
Intercept	0.03***	0.02**	0.02**	0.02**
	(0.16)	(0.13)	(0.12)	(0.12)
N_1	5,209	4,977	4,977	4,977
N_2	18	18	18	18
$\log L$	−3,551.721	−3,201.611	−3,203.521	−3,201.128
AIC	7,109.441	6,433.222	6,437.041	6,432.256
Percent Correctly Predicted	56.01	62.78	62.63	62.84

Note: Each model is a multilevel logistic regression with a random intercept for each Justice. Outcome is whether the Justice voted in favor of petitioner. Unit of analysis is each Justice's vote. Models include statements with question marks. The average vocal pitch in questions directed towards the petitioner ("Petitioner Pitch") minus the average vocal pitch in questions directed towards the respondent ("Respondent Pitch") is captured in "Pitch Difference" (Petitioner Pitch - Respondent Pitch). Model 2 uses the Dictionary of Affect in Language (DAL). Model 3 uses the Harvard-IV dictionary. Model 4 uses the Linguistic Inquiry and Word Count (LIWC) dictionary. The rest of the controls are the same as Black et al. (2011). 10-fold cross-validation is used to estimate the percent of votes correctly predicted. Please refer to the Supplemental Information for more details about each dictionary, the controls, and our cross-validation approach. Levels of significance are reported as follows: *p < .1; **p < .05; ***p < .01.

of a Justice voting for the petitioner is 0.55. However, the probability of a Justice voting for the petitioner drops by 7 percentage points if the difference between the vocal pitch directed to the petitioner is one standard deviation higher than the vocal pitch directed at the respondent. The overall prediction rate is also reported.[7] Here, we are able to predict 56.01 percent of Justices' votes accurately (see Table 1) and 66.55 percent of overall case outcomes accurately (see Table S4) using only pitch difference, suggesting that vocal pitch predicts not only how individual Justices vote but also the eventual disposition of the case.

Models 2, 3, and 4 include the controls used by Black et al. (2011), as well as the differences in the use of "pleasant" and "unpleasant" words as defined by the Dictionary of Affect in Language (DAL), the Harvard IV dictionary (also known as the General Inquirer), and the Linguistic Inquiry and Word Count (LIWC) dictionary, respectively. Since the Harvard IV dictionary is publicly available, we provide the words used for Model 3 in the Supplemental Information. For Models 2 and 4, we provide some examples of "positive" and "negative" words. Unlike Black et al. (2011), we use the Martin-Quinn scores estimated in the previous term, as Martin-Quinn scores are dynamically estimated within each term using Justices' votes, which would introduce endogeneity if not lagged. These are continuous measures from liberal (-1.0) to conservative (+1.0) and vary from Justice to Justice and from term to term. In addition, note that all models are multilevel logistic regressions with random intercepts for each Justice.[8]

After accounting for vocal pitch, including these other variables only increases the predictive power of the model by around seven percentage points, suggesting that vocal pitch has unique predictive value. To further assess the substantive importance of vocal pitch, we compared the performance of vocal pitch (and of only vocal pitch) to a widely known algorithm developed by Katz, Bommarito and Blackman (2014), known as {Marshall}+.

[7] We used a .50 threshold for these calculations, meaning when the model returned a predicted probability greater than .50, we predicted the Justice would vote for the petitioner.

[8] Due to space limitations, we do not describe all the control variables in the main text. Full descriptions and additional model specifications can be found in the Supplemental Information.

This algorithm uses 95 variables to predict Supreme Court Justices' voting and is known as one of the most predictive algorithms currently available.[9] As explained in the Supplemental Information, we restricted our analysis to the period from 1998 to 2012. In total, {Marshall}+ successfully predicts 64.76 percent of cases correctly, which is 1.79 percentage points *lower* than our prediction rate of 66.55 percent of cases. A simple χ^2 test reveals the models are similarly predictive ($\chi^2 = 0.52$, $df = 1$, $p > 0.05$), suggesting we are able to equal the predictive power of a model that uses 95 predictors using only one—vocal pitch.

Model 1 also outperforms traditional petitioner-based models in which a "plaintiff always wins" rule is applied. Although seemingly simple, such a rule is actually fairly sophisticated and takes into account a lot of what scholars know about strategic planning, the rule of four, and principles of precedent. Using only vocal pitch, Model 1 significantly ($p \leq 0.001$) outperforms this baseline by 2.25 percentage points. Even though Models 2, 3, and 4 all perform better, when one only uses the number of "positive" and "negative" words, the prediction rates are substantially worse. For example, when the only predictors are the percent more positive and negative words directed at the petitioner, the model successfully predicts 0.81 and 0.72 percentage points better than the "plaintiff always wins" model, depending on whether one uses the DAL or LIWC dictionaries, respectively. The Harvard-IV dictionary actually predicts 8.25 percentage points *less* than this baseline. As we show in the Supplemental Information, regardless of the text-based measure one uses, vocal pitch does substantially better at predicting both cases and votes. These results are not meant to suggest that vocal pitch is the *only* variable that should be incorporated into models of oral arguments. Nor do our results suggest text-based measures have no place in the study of emotional expression on the Supreme Court. Rather, our results demonstrate vocal pitch should be one of many variables that should be taken into consideration when assessing oral arguments.

[9]These 95 variables include case information, ideological information, Supreme Court trends, Justice background characteristics, etc.). See `http://lexpredict.com/portfolio/predicting-the-supreme-court`.

Discussion

For scholars interested in predicting Justice votes weeks, if not months, before the Court's ruling is released, oral arguments "provide a barometer of how justices will rule in a given case" (Black et al., 2011, 574). While we are not the first to suggest emotional expressions are an important part of such prognostications, the vast majority of these studies have only considered text-based measures (for an exception, see Schubert et al., 1992). We show vocal pitch on its own is about as predictive of Justices' votes and overall case outcomes as models that use all publicly available quantitative legal and non-legal information, including additional textual information related to emotion. These comparisons are not meant to suggest that vocal pitch is the *only* variable scholars should use when assessing emotional expression on the Supreme Court. We argue the {Marshall}+ algorithm, text-based measures, and the "petitioner always wins" rule can (and should) be used to predict Justice votes. However, non-verbal signals, including changes in vocal pitch, also carry considerable weight. Justices choose their words carefully, but have far less control over how those words are spoken—and these subconscious vocal cues, our findings show, carry important information about eventual rulings.

References

Black, Ryan C., Sarah A. Treul, Timothy Johnson and Jerry Goldman. 2011. "Emotions, Oral Arguments, and Supreme Court Decision Making." *The Journal of Politics* 73(2):572–581. 2, 6, 8

Calvo, Rafael A. and Sidney D'Mello. 2010. "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications." *IEEE Transactions on Affective Computing* 1(1):18–37. 2

Ekman, Paul, Maureen O'Sullivan, Wallace V. Friesen and Klaus R. Scherer. 1991. "In-

vited Article: Face, Voice, and Body in Detecting Deceit." *Journal of Nonverbal Behavior* 15(2):125–135. 2

Epstein, Lee, William M. Landes and Richard A. Posner. 2010. "Inferring the Winning Party in the Supreme Court from the Pattern of Questioning at Oral Argument." *The Journal of Legal Studies* 39(2):433–467. 2

Johnson, Timothy R., Ryan C. Black, Jerry Goldman and Sarah A. Treul. 2009. "Inquiring Minds Want to Know: Do Justices Tip Their Hands with Questions at Oral Argument in the U.S. Supreme Court?" *Washington University Journal of Law & Policy* 29:241–261. 2

Katz, Daniel Martin, Michael James Bommarito and Josh Blackman. 2014. "Predicting the Behavior of the Supreme Court of the United States: A General Approach." *Available at SSRN 2463244* . 6

Mauss, Iris B. and Michael D. Robinson. 2009. "Measures of Emotion: A Review." *Cognition and Emotion* 23(2):209–237. 3

Posner, Jonathan, James A. Russell and Bradley S. Peterson. 2005. "The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive development, and Psychopathology." *Development and Psychopathology* 1y:715–734. 2

Roberts, Jr., John G. 2005. "Inquiring Minds Want to Know: Do Justices Tip Their Hands with Questions at Oral Argument in the U.S. Supreme Court?" *Journal of Supreme Court History* 30:68–81. 2

Russell, James A. 2003. "Core Affect and the Psychological Construction of Emotion." *Psychological Review* 110:145–172. 3

Schubert, James N., Steven A. Peterson, Glendon Schubert and Stephen Wasby. 1992. "Observing Supreme Court Oral Argument: A Biosocial Approach." *Politics and the Life Sciences* 11(1):35–51. 8

Shullman, Sarah Levien. 2004. "The Illusion of Devil's Advocacy: How the Justices of the Supreme Court Foreshadow Their Decisions during Oral Argument." *Journal of Appellate Practice and Process* 6:271–293. 2