

Covering: Mutable Characteristics and Perceptions of Voice in the U.S. Supreme Court

Journal of Law and Empirical Analysis
2025, Vol. 0(0) 1–31
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2755323X251340759
journals.sagepub.com/home/lex



Daniel L. Chen¹ , Yosh Halberstam^{2,*}, and Alan Yu³

Abstract

The growing emphasis on “fit” as a hiring criterion introduces the potential for a new, subtle form of discrimination (Bertrand & Duflo, 2017). Analysis of 1,901 U.S. Supreme Court oral arguments from 1998 to 2012 documents that voice-based snap judgments predict court outcomes. Male petitioners who rank below median in perceived masculinity are 7 percentage points more likely to win. This negative correlation between perceived masculinity and winning cases in the Supreme Court is more pronounced in masculine industries. Perceived femininity of women lawyers also predicts court outcomes. Democrats favor men with less masculine-sounding voices. Perceived masculinity explains additional variance in Supreme Court decisions beyond what is predicted by the best random forest prediction model. A de-biasing experiment using information and incentives in factorial design is consistent with misperceptions and taste for masculine-sounding lawyers explaining the negative correlation between perceived masculinity and Supreme Court wins.

Keywords

gender, identity, speech, judicial decision making

JEL Codes

J15, J78, K41

1. Introduction

Courts routinely distinguish between immutable and mutable characteristics, and between being a member of a legally protected group as opposed to behavior associated with that group, which is not protected (Yoshino, 2006). Legal theorists suggest that discrimination, once aimed at entire groups, now aims at subsets that refuse to *cover*, that is, to assimilate to dominant norms (Goffman, 1963). Mutable characteristics have recently entered economic models of identity formation (Austen-Smith & Fryer, 2005; Bertrand & Duflo, 2017; Neumark, 2018). For example, African-Americans cannot be fired for their skin color, but they can be for wearing cornrows. This distinction between *being* (immutable) and *doing* (mutable) incentivizes assimilation. Should we understand such legally sanctioned differential treatment to be harmful? Are individuals “punished” for not conforming when it comes to mutable characteristics (Yoshino, 2000)? Building on

Kenji Yoshino’s notion of “covering,” we use the phrase “punish non-conformity” to capture how society might indirectly burden individuals whose voices deviate from a perceived norm.

The primary focus of our empirical analysis is judicial outcomes—namely, which attorney prevails at the Supreme Court—this emphasis motivates a selection puzzle rooted in

¹Toulouse School of Economics, Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France

²Economics Department, University of Toronto, Toronto, ON, Canada

³Linguistics Department, University of Chicago, Chicago, IL, USA

*Senior Financial Economist, Federal Deposit Insurance Corporation

Corresponding Author:

Daniel L. Chen, Toulouse School of Economics, Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France.
Email: daniel.chen@iast.fr



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

discrimination theory. In principle, if law firms or parties aim to maximize the probability of winning, they should promote or retain lawyers best suited to succeed before the Court. Classic models of taste-based or statistical discrimination (Becker, 1957; Phelps, 1972) suggest that if firms discriminate in favor of “masculine-sounding” voices merely due to preference or misbeliefs, then we would observe a mismatch if those favored traits do not align with actual performance. By testing whether perceived voice masculinity (or femininity) correlates negatively with success, we create an “outcomes test,” akin to the approach used in labor-market discrimination studies (Charles & Guryan, 2008). Put simply, if market participants rationally rewarded masculine voices, we would not expect to find that those same voices underperform at the Supreme Court. Yet that is precisely the pattern our data reveal. This incongruity points to a selection dynamic—firms appear to “choose the wrong attorneys” from a pure win-maximization standpoint—and underscores the broader significance of how extralegal traits (like voice) can affect both the makeup of the advocacy pool and the ultimate decisions handed down by the Justices. The selection puzzle we describe is purely suggestive; we do not have a controlled experiment manipulating the voice of advocates at the moment they are being selected by firms.

This study defines voice as a “mutable characteristic” because, although not trivially changed, vocal registers and pitch are subject to deliberate training and self-conscious modification. Professional speech coaching (e.g., Margaret Thatcher’s documented pitch-lowering exercises (Atkinson, 1984)) illustrates that people can—and do—adjust elements of their vocal style over time, placing voice in a middle ground between strictly immutable traits like race or age, and freely chosen characteristics like clothing. By “mutable,” we thus highlight how speakers can face social pressure to “cover” or conform their voice, even if such transformations demand substantial effort.

Likewise, our analysis of “perceptions” relies on external assessments: we ask a broad panel of online participants (via Amazon Mechanical Turk) to rate how “masculine” or “feminine” a single spoken phrase sounds—“Mr Chief Justice, (and) may it please the Court?”—thus standardizing word choice and capturing how everyday listeners might rapidly form impressions of an advocate’s voice. Because it is infeasible to poll the Justices themselves, these crowd-sourced ratings serve as a proxy for snap judgments any audience member (including the Court) could form upon hearing a brief utterance.

Our central inquiry is whether these externally measured, potentially mutable vocal cues correlate with attorney success in the Supreme Court. If extralegal traits—here, “perceived masculinity” or “perceived femininity”—predict outcomes net of conventional factors (e.g., case facts, attorney credentials), it raises questions about the subtle influences that mutable identity markers might have in high-stakes judicial settings.

Political scientists and legal scholars have long debated the extent to which Supreme Court decision-making is driven by legal doctrine versus extra-legal or attitudinal factors (Segal & Spaeth, 2002; Epstein & Knight, 1997; Epstein et al., 2013). Most of that work, however, emphasizes either rationally strategic behavior or broad ideological preferences. By contrast, we align more closely with a growing subfield that examines unconscious biases and heuristics—for example, how Justices might rely on System 1 thinking (Kahneman, 2011; Rachlinski et al., 2009a) even in a high-level adjudicative context. In this sense, our approach differs from purely rationalist extralegal accounts: we investigate a truly extraneous influence—vocal traits—that can shape outcomes in a way that does not appear to fit conventional strategic or attitudinal models. Scholarly work on facial appearance (Todorov et al., 2005; Benjamin & Shapiro, 2009; Berggren et al., 2010), emotional cues (Schubert et al., 2002), and cognitive biases (Rachlinski et al., 2009a) underscores that courts—like other deliberative bodies—can be susceptible to impressions that extend beyond the strict letter of the law.

Hence, while prior scholarship offers robust evidence of extra-legal judicial decision-making, it generally locates that behavior within an explicitly rational framework (e.g., Justices optimizing personal policy goals or responding strategically to political constraints). Here, by contrast, the role of vocal cues suggests a less canonical mechanism: snap judgments or involuntary biases that operate beneath conscious strategies. Positioning our findings within this newer subfield of judicial behavior helps illustrate that subtle, potentially irrational factors can still affect which side prevails—even in a context where law and policy are presumed dominant.

Against this backdrop, we examine voice-based perceptions as one more factor that might correlate with Supreme Court outcomes, echoing research showing that first-impression judgments—however fleeting—can influence real-world decisions (Ambady & Rosenthal, 1992). Our measure of “masculinity” in voice is used to explore whether an attribute that is tangential to legal merits can nonetheless predict who wins and who loses. Although we do not claim that judicial reasoning explicitly turns on an advocate’s vocal register, the data reveal consistent associations between perceived vocal traits and voting patterns, adding to the empirical record that external cues may have a subtle but significant impact in high-level adjudication.

Moreover, these findings hold relevance beyond the Supreme Court. Situations ranging from lower-court litigation to corporate hiring frequently involve consequential interactions where snap judgments about individuals arise—be it in job interviews, negotiations, or high-stakes presentations (Schroeder & Epley, 2015; Klofstad et al., 2012). In each case, a person’s vocal qualities can spark biases (positive or negative) that shape decisions. Hence, although the Supreme Court is an elite and idiosyncratic institution, our results

resonate with a broader phenomenon: voice-based heuristics can help or hinder individuals in settings where outcomes hinge on persuasion, credibility, or perceived competence.

We use “first impressions” to refer to the immediate, real-time judgments formed when an advocate begins speaking before the Court, rather than any assertion that the Justices come in with no prior knowledge of the lawyer’s identity or legal arguments. Indeed, many Supreme Court advocates appear repeatedly, and the Justices typically read briefs in advance (Hazelton & Hinkle, 2022). However, social cognition research (Ambady & Rosenthal, 1992) indicates that even well-informed decision-makers can form or update rapid heuristics upon encountering someone’s live voice or demeanor.

Scholars have extensively studied the “elite Supreme Court bar,” in which a small cadre of repeat advocates shape both advocacy strategy and judicial perception. While Lazarus (2008) focuses on the modern rise of specialized SCOTUS practitioners, a long-standing political-science literature (McGuire, 1993; 1995; Spriggs et al., 1995) emphasizes the repeat-player phenomenon, documenting how multiple appearances can foster ongoing relationships with the Justices. We recognize that many attorneys return for multiple cases and, accordingly, some of our specifications incorporate lawyer fixed effects, allowing us to examine whether within-lawyer variation in perceived vocal traits still predicts outcomes. Meanwhile, our dataset also includes a substantial fraction of single-appearance advocates, preserving variation along the “novelty” dimension.

Moreover, even for repeat players, the immediate auditory cues—pitch, resonance, intonation—can still prompt a new or reinforced perception during oral argument. We do not claim that these vocal impressions override ideology or the substantive content of briefs, but rather propose that voice-based attributes may be one additional extralegal influence on Justices’ reactions and, ultimately, on case outcomes. Thus, “first impressions” here reflects the psychologically grounded process triggered upon hearing an advocate’s live speech, highlighting how vocal qualities can still matter in a highly prepared, information-rich environment.

While the classical Becker (1957) theory suggests that complete markets would eradicate prejudice, as discriminatory employers would ultimately fail, evidence suggests otherwise. Numerous studies, including correspondence tests (Bertrand & Mullainathan, 2004) and analyses linking labor market inequality to prejudicial attitudes (Charles & Guryan, 2008), indicate persistent discrimination. Yet, distinguishing between prejudice and statistical discrimination remains complex, often hindered by limited data on productivity (Heckman & Siegelman, 1993; Neumark, 2018). Innovative experiments across various contexts (List, 2004; Mobius & Rosenblat, 2006; Rao, 2014) have attempted to dissect these discrimination forms, but real-world, high-stake settings remain challenging to study. We contribute to this discourse by analyzing 1,901 oral arguments in the U.S. Supreme Court

between 1998 and 2012, focusing on how snap judgments of lawyers’ voices correlate with court outcomes. Our analysis leverages voice samples (e.g., Sample 1 (<https://goo.gl/ZPdCkU>) and Sample 2 (<https://goo.gl/mbhuLF>)) holding fixed the words spoken – “Mr Chief Justice, (and) may it please the Court?” – to measure perceived masculinity and its association with case success.

For the analysis of observational ratings, participants simply rate already-existing voice clips (with or without reversing the clips) on perceived traits. We treat these ratings as a measurement of how typical listeners hear an attorney’s voice. In a randomly assigned 2×2 design (Information/No-Information $\times$ Incentive/No-Incentive), participants guess who won the case under varying levels of feedback and rewards. By observing changes in their correlation between perceived win rates and perceived masculinity, we can isolate how much of the correlation might stem from taste-based preference versus misbelief.

Our research design deliberately uses only the standard opening phrase, “Mr Chief Justice, (and) may it please the Court?”, in order to hold word choice constant and ensure that any perceived “masculinity” or “femininity” reflects vocal qualities rather than differences in lexical content. Moreover, we reverse some clips so that even English comprehension is removed, which further confirms that the acoustic signal alone drives our measured impressions.

To bolster confidence that this short standardized sample does not “miss out” on relevant vocal information, Dietrich et al. (2019) analyze entire oral arguments—without restricting to a single sentence—and rely on direct pitch measurements instead of listener ratings. Crucially, they reach the same conclusion: an advocate’s higher-pitched voice correlates with a higher likelihood of winning at the Supreme Court. Together, these approaches underscore that voice-based cues meaningfully predict Court outcomes, even beyond any specific word choices or rhetorical content.

Our results align with a growing body of empirical work suggesting that judicial outcomes can be influenced by factors not captured in traditional legal or attitudinal models (see, e.g., Chen & Jess, 2017; Rachlinski et al., 2009b). If vocal traits—which do not bear on the substantive merits of a case—correlate with winning, then no purely doctrinal or ideological account can fully explain how justices decide. This underscores the difficulty of building comprehensive models of judicial behavior and complicates the longstanding debate over whether decisions reflect law, policy, or personal biases (Segal & Spaeth, 2002). See also Epstein et al. (2018), Waterbury (2024), and the experiments by Rachlinski, Wistrich, and Guthrie for a few examples of how in-group bias, attractiveness, and extralegal factors can enter judicial decision-making.

More crucially, these findings raise questions of fairness and equal treatment under the law. If an advocate’s voice—something arguably tangential to the legal issues at stake—affects a party’s likelihood of prevailing, then systemic

reliance on such traits calls into question the principle that cases should be judged solely on their legal merits. From a normative perspective, allowing vocal attributes to shape outcomes challenges the fundamental expectation that litigants stand on equal footing regardless of personal characteristics.

This issue resonates with broader worries about implicit bias and the subtle ways in which perceived identity markers (gender presentation, race, accent) can permeate judicial or jury decisions. Whether or not these biases are deliberate, their presence in Supreme Court advocacy is particularly consequential, given the Court's role as the final arbiter of crucial legal and policy questions.

Katz et al. (2017) develop a time-evolving random forest classifier for Supreme Court outcomes, harnessing the Supreme Court Database (SCDB) across two centuries (1816–2015). They encode legal issues, court/Justice characteristics, and historical patterns (e.g., reversal rates) into a rich feature set. By retraining or ‘growing’ the random forest each term, Katz et al. show 70% accuracy in predicting affirm-or-reverse decisions across vastly different Court compositions.

In our work, we adapt the Katz framework not to maximize predictive accuracy per se, but rather to serve as a comprehensive control for all known correlates of Supreme Court outcomes—ideology, lower-court data, etc.—and then test whether vocal traits add explanatory power. Our main interest is causal inference: isolating whether lawyers' perceived masculinity (or femininity) has an independent effect above and beyond the Katz-style best-guess model.

Our main interest lies in identifying whether perceived vocal traits influence Supreme Court outcomes once we've accounted for all other known correlates (ideology, issue area, lower court data, etc.). The RF model itself is used primarily as a comprehensive control or “best guess” baseline—akin to how one might include a highly flexible function of observables in a regression—to ensure that any remaining correlation with masculinity/femininity is not due to omitted variables.

This paper proceeds as follows: Section 2 provides theoretical background; Section 3 describes our experimental methods and data; Section 4 presents baseline associations between voice impressions and Supreme Court outcomes, exploring the dynamics of lawyer variation and extending the analysis to female lawyers; and Section 5 offers concluding remarks.

2. Background and Theory

Previous research has extensively analyzed the nuances of speech, including aspects like vowels, pitch, diction, and intonation. However, the impact of speech variation, particularly beyond lexical choices, on real-world behavior remains underexplored. Our study addresses this gap by investigating the significance of vocal cues in high-stakes policy-making environments, specifically the U.S. Supreme

Court. We delve into whether these vocal cues hold relevance even when advocates (i.e., lawyers) deliver identical sentences, thus bringing a new dimension to understanding the influence of speech in critical decision-making settings.

First impressions based on voice have been shown to significantly influence decisions in diverse contexts, ranging from mate and leader selection to housing choices, consumer behavior, and even financial markets, as evidenced by vocal analysis in earnings conference calls (Nass & Lee, 2001; Klostad et al., 2012; Purnell et al., 1999; Scherer, 1979; Tighe et al., 2012; Mayew & Venkatachalam, 2012). These observations align with the broader behavioral literature on intuitive (System I) versus analytical (System II) thinking. Notable examples include studies showing that facial cues can predict electoral outcomes (Todorov et al., 2005; Benjamin & Shapiro, 2009; Berggren et al., 2010) and research demonstrating the influence of nonverbal behavior on teacher evaluations based on silent video clips (Ambady & Rosenthal, 1992). Intriguingly, Schroeder and Epley (2015) found that even in the presence of visual information, voice-based impressions carry more weight for employers in assessing job applicants. This body of work underscores the powerful role of vocal cues in shaping perceptions and decisions, setting the stage for our exploration of their impact in the high-stakes setting of the U.S. Supreme Court.

There are compelling arguments from various disciplines as to why voice impressions should ostensibly hold no sway in Supreme Court decisions. A rational perspective, as posited by Posner (1973), suggests that factual information should supplant initial impressions. Ideologically, some argue that court outcomes are largely political, with predetermined results (e.g., Cameron, 1993). Legally, it is advocated that decisions should hinge solely on the legal merits of arguments (Kornhauser, 1999). Economically, consistent correlations between voice and court outcomes are unlikely as law firms and advocates are expected to adapt their strategies to negate such biases (Becker, 1957; Knowles et al., 2001).

However, from a behavioral standpoint, the nuances of speech can reveal significant insights into one's personality and identity (Babel et al., 2013; Hodges-Simeon et al., 2010; McAleer et al., 2014; Bordalo et al., 2014). Past research underscores the potential impact of vocal cues in diverse environments, including courtrooms (Schubert et al., 2002) and presidential debates (Gregory & Gallagher, 2002). Furthermore, the alignment of a lawyer's spoken words with a SCOTUS judge's linguistic style has been shown to predict the judge's vote (Danescu-Niculescu-Mizil et al., 2012). This body of evidence suggests that despite theoretical arguments to the contrary, voice impressions may indeed play a role in high-stakes legal settings, a hypothesis our study aims to explore.

In the last few decades, the Supreme Court bar has been characterized by the emergence of an elite group of private sector attorneys not witnessed since the early 1800s (Lazarus, 2008). Elite Supreme Court legal practices are increasingly

led by “rainmakers” and landing a Supreme Court case is incredibly important to them. There are a few specialists, most working on behalf of businesses, who have enjoyed heightened success (Fisher, 2013). A new Supreme Court Pro Bono Bar has also emerged (Morawetz, 2011).

The selection of oral advocates by firms is when perceived masculinity may *positively* affect selection. However, firms that excessively prioritize masculinity in their selection criteria may inadvertently create a negative correlation between perceived masculinity and actual performance in court.

Positive selection may occur because deep-voiced individuals are perceived as having many positive attributes (Klofstad et al., 2012; Tigue et al., 2012; Apple et al., 1979; Buller et al., 1996). Margaret Thatcher and George H. W. Bush were coached to be less shrill (Kramer, 1987). Via humming exercises, Thatcher made her voice more masculine by the amount equivalent to half (<https://goo.gl/8bMkut>) the male-female difference (Atkinson, 1984), though her natural voice occasionally slipped out (<https://goo.gl/WNDgr0>) (Nallon, 2014). As women have entered the workforce and positions of authority, their voices have moved closer to a masculine standard (Pemberton et al., 1998) and they have been rewarded for this (Case, 1995). This phenomenon is not limited to women or leaders: in an employment discrimination case involving Sears, job applicants were asked: “Do you have a low-pitched voice?” (EEOC v. Sears, Roebuck & Co., 628 F. Supp. 1264, 1300). The employer preferred employees with masculine voices, even if they performed worse on the job (Case, 1995).

More formally, suppose there are two advocates M and F , who either win or lose. Consider the following utility:

$$U_v = \alpha\pi_v + V \quad (1)$$

where $\pi_v = \text{Prob}(v \text{ wins})$ represents the individual’s beliefs about whether the advocate chosen by him will actually win, and V represents the individual’s preference (taste) for advocate V independent of the economic consequences. In sum, an individual chooses an advocate based on his or her beliefs about whether that lawyer will win, represented by the probability π_v , and a taste for an advocate with voice V . The parameter α represents stakes—e.g., “Billions of dollars at stake in Supreme Court power market fight” (Crawford & Malik, 2016).¹

Individuals will choose advocate F over M if and only if the difference in the probability of F winning rather than M winning exceeds the relative taste individuals have for advocates with voice M :

$$\alpha(\pi_F - \pi_M) \geq d \quad (2)$$

where $d \equiv M - F$. Similarly, an individual chooses advocate M over F , if and only if, the difference in beliefs in the probability of M winning over F winning exceeds the difference in the taste for F over M , weighted by the stakes.

Suppose individuals are more likely to choose M over F . There are two reasons individuals may choose M over F : information and taste, i.e., due to statistical discrimination (information, $\pi_F < \pi_M$) or prejudice (taste, $d > 0$). Information can be used to update one’s beliefs about $\pi_F - \pi_M$, and any changes in behavior would be due to information. Likewise, the incentives to choose correctly erode the effect of taste on choices ($\pi_F - \pi_M > d/\alpha$). Any changes in behavior would be due to existence of preference (taste) for a type of advocate independent of the economic consequences. Incentives increase α , so any response would imply that $d > 0$.

In the experiment, if subjects perceive masculine voices to be more likely to win, their prior beliefs are that $\pi_M > \pi_F$. To summarize,

- $\frac{\partial(\pi_F - \pi_M)}{\partial \text{Info}} > 0$ would be evidence of statistical discrimination
- $\frac{\partial^2(\pi_F - \pi_M)}{\partial \text{Info} \partial \text{Incentives}} > 0$ would be evidence of prejudice

As $\alpha \rightarrow \infty$, only a large d (on the order of US\$10 million) explains the negative correlation between perceived masculinity and court outcomes. Further, in industries with higher $d > 0$, masculine voices would do worse as firms indulge in taste d at cost of $\alpha\Delta\pi_v$.

In sum, this theoretical analysis motivates a debiasing experiment where we randomize both the information and the incentives. We will investigate whether information reduces the correlation between perceptions of masculinity and perceptions of winning. We will then investigate if incentives further reduce this correlation. Providing incentives in the model is analogous to increasing the stakes, which would reduce the influence of taste on choices, but only if there is a positive taste for masculine lawyers. If the only reason that decision makers prefer masculine lawyers is due to misbeliefs, then only information would affect decisions. In the model, this would mean that d/α would be 0 regardless of the size of α .

3. Experiment

This study’s design builds on a rich interdisciplinary tradition examining how minimal voice cues influence judgments in high-stakes settings. Research in sociophonetics shows that brief audio clips can convey potent social signals (Purnell et al., 1999; Babel et al., 2014). In parallel, psychology experiments on “thin slices” of expressive behavior demonstrate that even a few seconds of speech or nonverbal interaction can reliably affect evaluators’ perceptions and subsequent decisions (Ambady & Rosenthal, 1992). Meanwhile, scholarship specific to the U.S. Supreme Court (e.g., Epstein et al., 2010) highlights the potential impact of oral argument dynamics on case outcomes, including how advocates’ manner or style may subtly shape judicial impressions. Our methodology—extracting standardized, short

voice samples (“Mr Chief Justice, may it please the Court”) and collecting crowd-sourced ratings—follows the broader logic of using minimal cues to uncover bias or discrimination (Bertrand & Mullainathan, 2001). By situating the data-collection approach within these literatures, we aim to illuminate whether similarly rapid, voice-based judgments might correlate with real legal outcomes in the highest court.

3.1. Design

Oral arguments at the SCOTUS have been recorded since the installation of a recording system in October 1955. The recordings and the associated transcripts were made available to the public in electronically downloadable format by the Oyez Project (<https://www.oyez.org/>), which is a multimedia archive at the Chicago-Kent College of Law devoted to the SCOTUS and its work. The audio archive contains more than 110 million words in more than 9000 hours of audio synchronized, based on the court transcripts, to the sentence level. Oral arguments are, with rare exceptions, the first occasion in the processing of a case in which the Court meets face-to-face to consider the issues. Usually, counsel representing the competing parties of a case each have 30 minutes in which to present their side to the Justices. The Justices may interrupt these presentations with comments and questions, leading to interactions between the Justices, the lawyers and, in some cases, the *amici curiae*, who are not a party to the case but nonetheless offer information that bears on the case to assist the Court.

This paper presents analysis of clips taken from the period 1998–2012. Over 80% of the advocates featured in these clips argued only once before the SCOTUS, and 169 advocates in the sample—about 15%—were female. We hired 748 MTurk raters from the US to evaluate 1,901 audio clips comprising 1085 lawyers. Each rater evaluated 60 random clips, producing roughly 20 ratings for each clip. Six clips were randomly chosen to be played twice for each rater, resulting in each rater providing a total of 66 evaluations.

A little over half (382) of the 748 MTurk raters were female. Two-thirds were aged between 20 and 35 years old, and one-third were older than 35. Likewise, one-third indicated they had some college education, whereas one-third claimed to have a bachelor’s degree. The median annual income of those who completed the survey was about \$40,000. Their racial and geographical distribution broadly reflects that of the US population.

The raters were asked to use headphones and to rate on a Likert scale from 1 (low) to 7 (high) the characteristics of masculinity, attractiveness, confidence, intelligence, trustworthiness, and aggressiveness. These six traits were selected based on previous research on listeners’ perceptual evaluations of linguistic variables (Eckert, 2008; Campbell-Kibler, 2010; McAleer et al., 2014). They are also similar to the ones used in Todorov et al. (2005), which presented subjects with pictures of electoral candidates’ faces and asked them to rate

their perceived attributes. That study found that only perceptions of competence predicted election outcomes. To assess whether raters were making global judgments about candidates that were not specific to competence, the authors also elicited judgments of candidates’ intelligence, leadership, honesty, trustworthiness, charisma, and likability. We take a similar approach by analyzing how judgments of masculinity, while correlated with judgments of other voice attributes, are the only ones that predict court outcomes in a consistent and robust manner. Subjects were also asked to predict whether the lawyer would win.

Male and female lawyers were rated in separate blocks, such that participants either rated male advocates or female advocates but not both, so raters would not be comparing females and males on the degree of masculinity. Female lawyers were rated in terms of femininity instead of masculinity.

We randomized the order and polarity of the questions (e.g., “*very masculine*” and “*not at all masculine*” would appear on the left and right of a 7-point scale, respectively, or on the right and left in the opposite polarity); the order and polarity of questions were held fixed for any particular rater to minimize cognitive fatigue. For additional nudges across experimental designs, to ensure attention by the rater, we included listening attention checks. Raters who failed these checks were dropped from the sample. There were six alertness trials, three with beeps and three without. The beep comes at the beginning of the lawyer’s voice. Subjects were asked if they heard a beep, but not to rate the lawyer’s voice.

Several clips were repeated to check for intra-rater reliability. Raters were also asked to rate the quality of the recording. While there is no time limit on how long a subject can spend on each trial, they were given a minimum of 5 seconds to respond; they were not allowed to proceed to the next trial until the 5 seconds was up (and all the questions completed) in order to ensure that subjects were given enough time to complete the ratings and to discourage them from speeding through the trials. An example of what each questionnaire looked like is provided in Figure 1. No information regarding the identity of the lawyer or the nature of the case was given to the participants.²

We also reversed the voice clips so the sentence no longer sounded anything like English.³ As Figure 2 shows, most attributes exhibit some correlation, but generally very weak ones. The only personality attribute that shows a strong correlation is masculinity. This result suggests that masculinity may be more salient, innate, or stereotyped than other attributes, such as aggressiveness, confidence, competence, attractiveness, intelligence, or trustworthiness.

Some of this data was previously analyzed in an earlier study by the authors (Chen et al., 2016). We expand on these earlier analyses in six directions. First, we include female lawyers. Second, we explore mechanisms behind the observed correlations, for example, by merging additional datasets for heterogeneity analyses to elucidate the mechanisms.

Recording 1 of 66

0:00

1. Please provide your impression of the voice recording in the matrix below:

Very Attractive	●	●	●	●	●	●	Very Unattractive
Very Masculine	●	●	●	●	●	●	Not At All Masculine
Not Intelligent	●	●	●	●	●	●	Intelligent
Very Unaggressive	●	●	●	●	●	●	Very Aggressive
Not Trustworthy	●	●	●	●	●	●	Trustworthy
Very Confident	●	●	●	●	●	●	Very Timid

2. Assuming that this is a lawyer arguing a case in front of a panel of judges, how likely do you think this lawyer will win the case?

Will Definitely Lose ●●●●●●●● Will Definitely Win

3. How good is the quality of the recording?

Very Bad ●●●●●●●● Very Good

Next

Figure 1. Screenshot of experiment. *Note.* This sample page illustrates how MTurk participants rated attorneys' vocal clips. Raters heard a brief audio segment ("Mr Chief Justice, may it please the Court?"), then indicated perceived masculinity, attractiveness, confidence, intelligence, trustworthiness, and aggressiveness on a 1-7 scale. They also guessed whether the lawyer won the case. The instructions, randomized question order, and forced minimum listening time are described in Section 3.1.

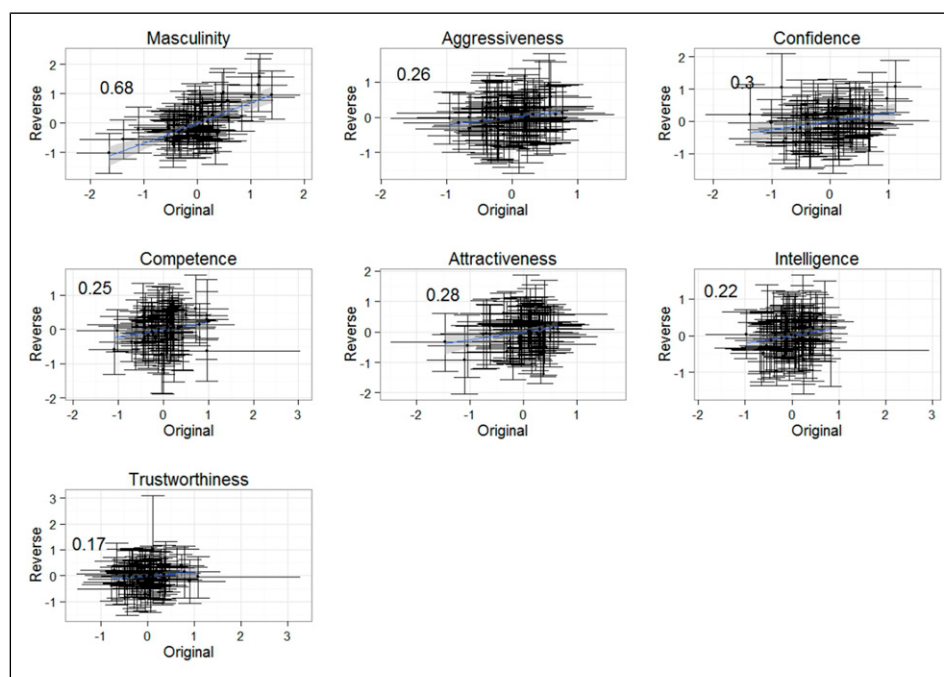


Figure 2. Correlation in voice perceptions across experiments (Forward vs. Reversed Clips). *Note.* We compare raters' impressions when the audio clip was played forward versus reversed (stripping English content). Each point represents a trait (masculinity, confidence, etc.) and shows the correlation of that trait across forward versus reversed conditions. See Section 3.1 for details on the reversed condition. Higher correlation suggests that perceived masculinity or femininity is robust to the language's intelligibility. The number represents the Pearson's r .

For example, the heterogeneity in judge votes assuages concerns that omitted case characteristics are associated with voice masculinity and win rates. Third, we examine how lawyers change their voice over time. Fourth, we implement a

de-biasing experiment. We use a factorial design. One factorial arm provides information to raters after each voice rating. More specifically, raters were asked whether they believed the lawyer would win. Immediately they are told

whether the lawyer actually won before the next voice clip is played for the rater. The other factorial arm provides incentives for raters to guess correctly whether the lawyer will win or not. Our de-biasing experiment complements the observational data by disentangling the roles of statistical discrimination and prejudice, revealing that both factors significantly contribute to the correlation between voice perception and court success. Fifth, we find that *whose* perception of masculinity matters. Sixth, we use machine learning to predict Supreme Court votes and find that our results hold controlling for the best prediction model of Supreme Court votes. When we introduce masculine perceptions as a feature, the model also selects masculinity as a feature.

3.2. Measurement

We measure perceived masculinity in three ways. Within-rater normalization entails adjusting for cross-subject variability in the cardinality and spread of ratings. Let $attribute_{itw}$ be subject w 's perception of a given attribute of advocate i in case t , where $attribute$ refers to any one of the six traits (i.e., attractiveness, confidence, etc.) and the perceived likelihood of winning. The normalized rating is given by

$$\widehat{attribute}_{itw} = \frac{attribute_{itw} - \overline{attribute}_w}{\sigma(attribute)_w}, \quad (3)$$

where $\overline{attribute}_w$ is the average perception of a given attribute across rater w 's advocate ratings and $\sigma(attribute)_w$ is the standard deviation of these ratings. As a result, for each rater w , $\widehat{attribute}_{itw}$ is a continuous measure with a mean of 0 and a standard deviation of 1.

Our second measure uses raw ratings, which give more weight to raters who provide more signal amid greater variance in their ratings.⁴

A third measure of voice attribute uses the average scores of each lawyer, matching only one voice measure to each oral argument. To do this, we take the average ratings across raters as follows:

$$\overline{\widehat{attribute}}_{it} = \frac{1}{|W_{it}|} \sum_w \widehat{attribute}_{itw}, \quad (4)$$

where W_{it} is the set of raters who rated the voice of advocate i in case t .

There are 33,666 individual ratings, with roughly 20 ratings per oral argument. The standard deviation of the raw ratings is about 1.5. For robustness, we present sets of analyses using all three measures in Section 4, but the raw ratings is our preferred measure because it weights more heavily the raters who provide more signal amid greater variance in their ratings. We use the raw ratings when conducting mechanism analyses in Section 5.

3.3. Empirical Strategy

We consider two empirical specifications to explore whether first impressions of lawyers' voices predict Supreme Court outcomes. In the micro-level strategy, we estimate the following equation:

$$win_{it} = \alpha + \widehat{attribute}_{itw}'\beta + \mathbf{x}_{itw}'\gamma + \varepsilon_{itw}, \quad (5)$$

where the dependent variable is an indicator for whether advocate i actually won case t , and the key independent variables denoted by the vector **attribute** are continuous measures of the advocate's set of attributes in case t as perceived by the MTurk raters. Given the regression equation, β represents the association between advocate traits and actual wins.

3.3.1. Control Variables. We further expand our analysis by including in some regressions additional covariates denoted by the vector $\mathbf{x}$. The vector $\mathbf{x}$ can include rater w 's age, gender, race, income, education, and state of residence, lawyer i 's characteristics, and characteristics of case t .

Attorney Characteristics: These comprise SCOTUS clerkships, age, number of clerkships, law review, other graduate degree, law school tier, masculinity of first name (the number of males divided by the number of females with that name in the census), number of previous SCOTUS oral arguments, years since graduation, number of admitted courts, number of practice areas, firm size, office size, and whether they are a law firm partner.

Case-Level Controls: These controls are case category, region of origin, and lower-court characteristics—reversal of trial court, opinion length, disagreement, political division, ideology, and number of self-certainty words that proxy for confidence.

Industry or Client Type: For some specifications, we track whether the advocate's client is a government entity, nonprofit, or private corporation, along with indicators for "masculine" industries (e.g., penal institutions, telecommunications).

In other specifications, we adopt a machine-learning approach (inspired by [Katz et al., 2015](#)) in which these controls feed into a random forest to create a high-quality "predicted probability of winning." We then include that prediction in a simpler regression with vocal masculinity as the variable of interest. This approach ensures that the entire information set—attorney plus case-level covariates—is accounted for, while enabling us to see whether vocal traits retain an independent correlation.

Because our main goal is to isolate the effect of voice net of attorney quality or case factors, we do not emphasize every control's individual coefficient in the main text. Instead, we show how adding these covariates affects the estimated coefficient on vocal traits—a standard approach in causal-inference frameworks (see [Oster, 2019](#)).

In certain specifications, we use attorney fixed effects rather than explicit variables, effectively removing all time-invariant differences across that attorney. In specifications with lawyer fixed effects, we effectively control for any stable personal attributes, including facial appearance. Thus, while [Chen \(2018\)](#) finds that face and voice both matter in cross-sectional predictions, our within-lawyer approach shows that mutable vocal traits can still affect outcomes beyond any time-invariant characteristics—such as facial features—associated with each attorney. Where feasible, we add year or case fixed effects, further netting out variations specific to a particular term or dispute.

We adjust the standard errors of the regression estimates for clustering at the case level (and also multi-way clustering at the lawyer level and, later, the judge level). Petitioners (the first lawyer to speak and arguing on behalf of the plaintiff) and respondents (the second lawyer to speak and arguing on behalf of the defendant) are presented separately.⁵

In the case-level strategy, we estimate the following equation:

$$win_{it} = \alpha + \widehat{\text{attribute}'_{it}\beta} + \mathbf{x}'_{it}\gamma + \varepsilon_{it}. \quad (6)$$

In these regressions we cannot control for intra-rater rating correlations nor rater characteristics. For these reasons, the aggregated regression is generally viewed as too conservative in terms of statistical precision ([Bertrand et al., 2004](#)). We provide regression results using both the micro- and case-level strategies. Importantly, the interpretation of the magnitudes of the association will differ depending on the level of aggregation.

3.4. Data Generating Process

The underlying data-generating process is halfway between the micro-level regression with 20 raters per advocate and the case-level regression with one rating per advocate. In the Supreme Court, each of nine Justices forms his or her own perception. The corresponding micro-level strategy is:

$$judge\ votes\ for\ advocate_{itw} = \alpha + \text{attribute}'_{itw}\beta + \mathbf{x}'_{itw}\gamma + \varepsilon_{itw}, \quad (7)$$

where w subscripts for judge.

The outcome of the case is a function of the nine separate impressions of the advocate.

$$win_{it} = f(judge\ impression) = \alpha + \widehat{\text{attribute}'_{it}\beta} + \mathbf{x}'_{it}\gamma + \varepsilon_{it}. \quad (8)$$

Differences between the case outcome and the judge votes would be suggestive of whether the swing voter's behavior is linked to perceived masculinity.

Data simulations of equation (8) and comparing β and $\tilde{\beta}$ yield two separate observations: First, β is more

significant than $\tilde{\beta}$ if $\text{attribute}'_{itw}$ is highly correlated across attributes at the w -level, which is consistent with [Bertrand et al. \(2004\)](#). Moreover, β is smaller than $\tilde{\beta}$ because $\sigma(\text{attribute}'_{itw}) > \sigma(\widehat{\text{attribute}'_{it}})$. The proper interpretation is in terms of standard deviation units.

4. SCOTUS Outcomes and Perceived Masculinity

We can discretize the continuous rating measures at the lawyer level and observe that a below-median masculinity rating corresponds to a roughly 7 percentage-point greater likelihood of winning ([Figure 3](#)). No association between perceived masculinity and court outcomes is found among the lawyers for the respondent, consistent with the primacy of first impressions.

Since female lawyers are coached to be more masculine ([Starecheski, 2014](#)), this raises the question of whether our findings extend to female lawyers. The anecdote from Starecheski's NPR piece (2014)—in which a lawyer deliberately sought to deepen her voice—demonstrates how legal professionals may feel compelled to modify their vocal presentation. A broader empirical literature shows that advocates, politicians, and other speakers consciously adjust pitch and resonance over time. For instance, [Kramer \(1987\)](#) and [Atkinson \(1984\)](#) document Margaret Thatcher's deliberate pitch-lowering exercises, while [Pemberton et al. \(1998\)](#) find that working women often shift toward lower pitch in response to workplace norms. Moreover, [Case \(1995\)](#) analyzes litigation over hiring decisions allegedly contingent on an applicant's vocal depth, indicating that deeper or more 'authoritative' tones can be rewarded in certain professional environments. Together, these works illustrate that while voice modification is not trivial, it is perceived as sufficiently feasible that attorneys, politicians, and other public figures invest in coaching and training.

Studies on voice-based social biases have repeatedly observed significant differences in how listeners react to voices of different (perceived) genders ([Babel et al., 2014](#)). We thus examine male and female lawyers separately for the association between perceived masculinity (femininity) and court outcomes.

The results extend to females: while an increase in perceived masculinity among male lawyers correlates with a decrease in the likelihood of winning, the same degree of increase in perceived femininity in female lawyers correlates with an increase in the likelihood of prevailing in a court case ([Figure 4](#)). If "masculine" were the opposite of "feminine," then the pooled results would be stronger.

Another interpretation of our results – where we discretize the continuous rating measures at the lawyer level – is that a below-median masculinity rating corresponds with a roughly 2 percentage-point greater likelihood of winning for males, but a below-median femininity rating corresponds with an

approximately 5 percentage-point lower likelihood of winning for females (Figure 5).

Interestingly, the correlation with outcomes is specific to perceived femininity. In Table 1 (using raw ratings) and A.2 (using normalized ratings), we present these findings. In Columns 1–4 we narrowly focus on the association between *Masculine (Feminine)* and the case outcome without adding any additional covariates. It appears that while sounding masculine is negatively correlated with winning the case for male advocates, female advocates who sound more feminine are more likely to win. These results continue to hold in Columns 5–8 with the inclusion of the full set of traits. Among petitioners, a one-standard-deviation increase in

perceived masculinity predicts a lower likelihood of winning for men; and, the correlation between femininity and outcomes for female petitioners is positive.

In Table 2, we examine the relationships at the case level, where the relationship with perceived femininity remains robust and specific to perceived femininity; it is also larger, and it appears primarily for the petitioner at the case level. We present regression results using the normalized data in Table A.4, which shows that the qualitative findings are consistent with the results using the raw data: perceptions of masculinity are negatively correlated with win probabilities for male advocates, and perceptions of femininity are positively correlated with win probabilities for female advocates.

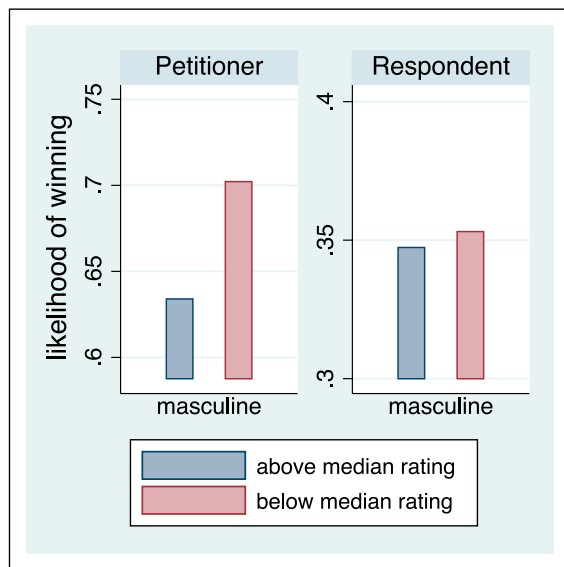


Figure 3. Perceived masculinity and win rate by petitioner–respondent status. *Note.* We group attorneys by above versus below median perceived masculinity (1–7 scale) and show the fraction of cases won. Petitioner lawyers argue first, Respondent lawyers second. Bars indicate mean outcomes. Source: 19,982,012 SCOTUS audio data. See Section 4 for a full discussion of sample construction.

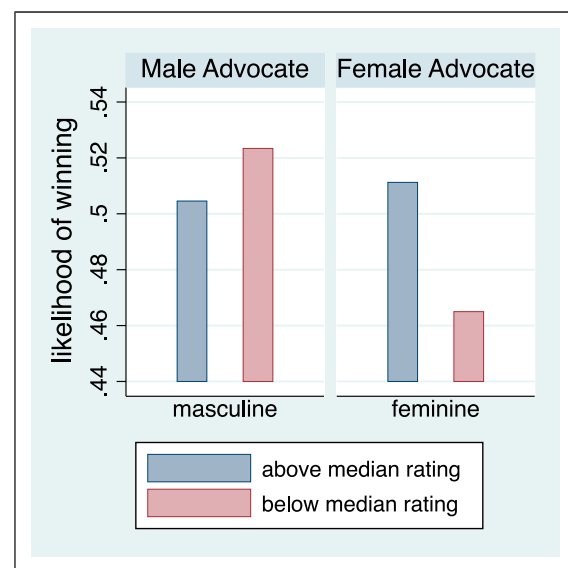


Figure 5. Perceptions of masculinity (Femininity) and advocate win rates. *Note.* We group male attorneys by above versus below median perceived masculinity (1–7 scale) and likewise for female attorneys by above versus below median perceived femininity. Bars indicate mean outcomes. Source: 1998–2012 SCOTUS audio data.

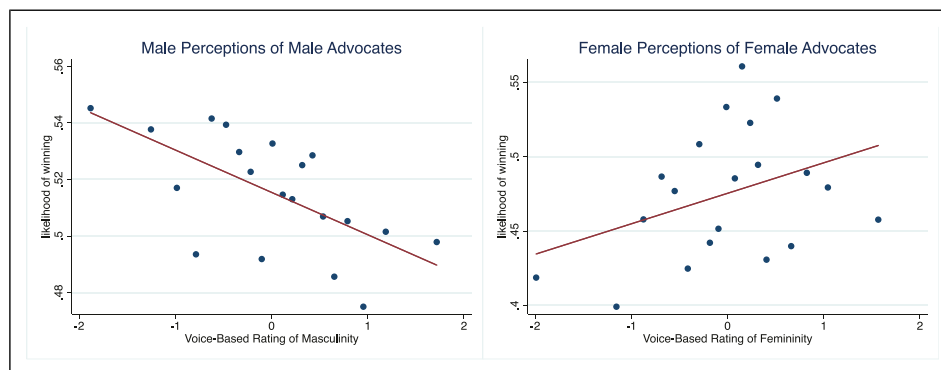


Figure 4. Graphical illustration of gender and predicted outcomes. *Note.* Illustrates how male versus female attorneys fare given different ranges of perceived vocal masculinity/femininity. X-axis = perceived masculinity/femininity (from lower to higher). Each bin on the X-axis represents an equal-sized subset of lawyers by rating. Y-axis = predicted probability of winning.

Table 1. Voice Perceptions and Gender (Raw Ratings, Micro Level).

Dependent Variable: Case Outcome (Win = 1, Lose = 0)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Male Advocate		Female Advocate		Male Advocate		Female Advocate	
	Petitioner	Both	Petitioner	Both	Petitioner	Both	Petitioner	Both
Masculine/Feminine	−0.0118** (0.00524)	−0.00707* (0.00413)	0.00991 (0.00952)	0.0153** (0.00727)	−0.0149*** (0.00565)	−0.00975** (0.00440)	0.0131 (0.00933)	0.0125* (0.00678)
Confident					0.00508 (0.00387)	0.00417 (0.00312)	−0.00821 (0.00992)	0.00701 (0.00730)
Attractive					0.0000377 (0.00445)	−0.000821 (0.00328)	0.00280 (0.0112)	−0.00587 (0.00743)
Intelligent					0.00244 (0.00385)	0.00634** (0.00292)	−0.00192 (0.0109)	−0.00578 (0.00847)
Trust					0.00356 (0.00344)	−0.00177 (0.00264)	−0.00439 (0.00973)	−0.00175 (0.00647)
Aggressive					−0.00134 (0.00345)	0.00182 (0.00266)	0.00501 (0.0102)	0.00243 (0.00786)
Likely winner					−0.000977 (0.00411)	−0.00270 (0.00315)	0.00251 (0.0108)	0.00361 (0.00785)
Observations	18543	35330	2543	5868	18542	35329	2543	5868

Note. Linear probability model (LPM) using raw (1–7) ratings of perceived masculinity (or femininity). The unit of analysis is a lawyer-by-rater observation; the dependent variable is an indicator for whether the advocates side won (1 = win, 0 = lose). Columns focus on male or female attorneys, and some specifications add controls. Column 1 tests whether perceived masculinity is associated with case outcome among male petitioners. Column 2 tests whether perceived masculinity is associated with case outcome among male advocates. Columns 3–4 do the same as Columns 1–2 but among female advocates. Columns 5–8 do the same as Columns 1–4 but also control for other perceptual ratings. Standard errors are clustered by case. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

To further examine the robustness of these advocate-by-gender findings, we gradually add MTurk rater controls and lawyer fixed effects in Tables 3 and A.3. The inclusion of rater characteristics in Columns 1–4 does not influence the estimated coefficients. Likewise, with the addition of lawyer fixed effects, the coefficient estimates for male advocates are very similar to those obtained in previous regressions using only within-advocate variation. That is, a one-standard-deviation increase in perceived masculinity predicts a lower likelihood of winning. However, the same does not hold for female advocates. In Columns 7 and 8, controlling for rater characteristics and lawyer fixed effects, coefficient estimates for female advocates are negative and decline significantly towards zero. This perhaps reflects the limited within-advocate variation among female advocates in our sample.

To benchmark our findings, the 0.86 percentage-point difference in court outcomes attributed to a one-standard-deviation change in our voice-based measure of perceived masculinity is equivalent to nearly one-quarter of the gender gap. When analyzed at the case level, a one-standard-deviation change corresponds to 5.2 percentage-point difference in court outcomes (when not including further controls).⁶

In Table 4, we find significant correlations at the outcome win level (Column 1) but not at the judge vote level (Column 2)

or for the number of judge votes (Column 3). If the votes of swing voters are associated with perceived masculinity in close cases (and the votes of Justices are not associated with perceived masculinity in easy cases), we would expect to find associations at the case level rather than vote level. This result is consistent with heuristics or biases being more likely to influence decisions in hard or close cases.

Our reference to heuristics or biases draws on the large body of behavioral research showing that people often rely on simplifying mental shortcuts—particularly when facing uncertainty or complexity (Tversky & Kahneman, 1974; Kahneman, 2011). In the context of Supreme Court adjudication, these cognitive shortcuts can become more influential precisely when a case is legally or ideologically “hard”: that is, when doctrinal precedent does not squarely resolve the issue or when the Justices are evenly divided in their ideological leanings. Without a clear rule or consensus, extralegal factors—such as an advocate’s vocal pitch or perceived confidence—can inadvertently shape which side appears “slightly more persuasive.”

Rachlinski et al., 2009b have documented how System 1 thinking (Kahneman, 2011)—fast, intuitive decision-making—can affect judges, even in relatively stable legal domains. In the Supreme Court, where cases often involve high-stakes or novel legal questions, these snap judgments or intuitive heuristics may loom larger when Justices find little

Table 2. Voice Perceptions and Gender (Raw Ratings, Case Level).

Dependent Variable: Case Outcome (Win = 1, Lose = 0)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Male Advocate	Female Advocate	Male Advocate	Female Advocate	Male Advocate	Female Advocate	Male Advocate	Female Advocate
	Petitioner	Both	Petitioner	Both	Petitioner	Both	Petitioner	Both
Masculine/ Feminine	−0.0498** (0.0225)	−0.0279* (0.0169)	0.0829 (0.0801)	0.125** (0.0580)	−0.0773** (0.0320)	−0.0369 (0.0239)	0.273* (0.154)	0.182 (0.113)
Confident					0.0816 (0.0540)	0.0502 (0.0431)	−0.247 (0.162)	0.0136 (0.104)
Attractive					−0.0233 (0.0484)	−0.0167 (0.0349)	0.0626 (0.150)	−0.0543 (0.103)
Intelligent					−0.00769 (0.0643)	0.0907* (0.0464)	−0.0209 (0.109)	0.00457 (0.0762)
Trust					0.0731 (0.0620)	−0.0185 (0.0461)	−0.112 (0.177)	−0.0566 (0.124)
Aggressive					−0.0222 (0.0469)	0.0160 (0.0378)	0.134 (0.125)	−0.0259 (0.0904)
Likely winner					−0.0324 (0.0754)	−0.0796 (0.0568)	0.00836 (0.194)	−0.00525 (0.129)
Observations	856	1634	116	267	856	1634	116	267

Note. Same raw ratings (1–7), but averaged by case (i.e., we take the mean rating for each attorney across all raters). Dependent variable: case outcome (win = 1). Columns focus on male or female attorneys, and some specifications add controls. Column 1 tests whether perceived masculinity is associated with case outcome among male petitioners. Column 2 tests whether perceived masculinity is associated with case outcome among male advocates. Columns 3–4 do the same as Columns 1–2 but among female advocates. Columns 5–8 do the same as Columns 1–4 but also control for other perceptual ratings. Standard errors are clustered by case. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

doctrinal clarity or experience near-even splits. In such scenarios, judges may—consciously or not—rely on “soft” cues like voice timbre, demeanor, or emotional tone (Schubert et al., 2002).

While the significance of gendered language and courtroom behavior is not new, prior scholarship has typically focused on broad differences between male and female attorneys rather than on acoustic details (e.g., pitch, resonance, or intonation) that vary even within gender categories. Our findings suggest these voice-based impressions may be especially potent in marginal or contested cases, contributing to the broader conversation about how cognitive heuristics operate when Justices lack unequivocal doctrinal signals or when ideological divisions are sharp. In other words, heuristics involving vocal attributes could serve as an additional tipping factor—not displacing legal reasoning, but accentuating or reducing the perceived credibility of each side’s argument when the legal question remains uncertain.

Hence, in “hard” cases, the Court’s decision-making may involve a blend of System 2 deliberation (analytical parsing of precedent, legal tests) and System 1 intuitions (rapid, sometimes unconscious assessments of the advocate’s presentation). By illuminating how perceived vocal cues correlate with votes and outcomes, this study underscores the subtle extralegal elements that can influence even the highest judicial forum. That influence, we propose, is most visible in

close cases, where small psychological nudges—including the advocate’s voice—may help Justices tip from one outcome to the other.

While the focus on language and gender in the courtroom is not new, previous studies have focused primarily on the gendered language performance of witnesses (O’Barr & Atkins, 1980) or the discursive practices in the courtroom (O’Barr, 1982). Scholars have long examined the significance of oral argument at the Supreme Court, tracing how the substance and style of advocacy can shape judicial decision-making. In particular, Chief Justice Roberts (2005) and Johnson et al. (2006) explore how advocate performance can persuade Justices and influence case outcomes, laying the groundwork for subsequent empirical studies on this topic. Other scholars find that Supreme Court outcomes are correlated with authors’ coding of emotional arousal in the behavior of Justices, lawyers, and their voices (Schubert et al., 1992); with the number of questions asked by Justices (Epstein et al., 2010); and with measurements of the emotional content of questions using linguistic dictionaries (Black et al., 2011). By building on these sources, we emphasize the continuity in research linking advocacy style and real-world appellate outcomes, setting the stage for our own focus on how voice characteristics might matter at the margin.

A wealth of scholarship already documents the ways gender may influence judicial outcomes, including analyses

Table 3. Gender, Lawyer, and MTurker Rater Controls (Raw Ratings, Micro Level).

Dependent Variable: Case Outcome (Win = 1, Lose = 0)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Male Advocate	Female Advocate	Male Advocate	Female Advocate	Male Advocate	Female Advocate	Male Advocate	Female Advocate
	Petitioner	Both	Petitioner	Both	Petitioner	Both	Petitioner	Both
Masculine/ Feminine	−0.0144** (0.00571)	−0.00963** (0.00444)	0.00991 (0.0103)	0.0133* (0.00743)	−0.00569** (0.00241)	−0.00668*** (0.00203)	−0.00153 (0.00341)	−0.000395 (0.00402)
Confident	0.00457 (0.00390)	0.00379 (0.00314)	−0.00950 (0.0105)	0.00814 (0.00802)	0.00314 (0.00194)	0.00233 (0.00179)	0.00365 (0.00532)	0.00132 (0.00376)
Attractive	−0.000428 (0.00447)	−0.00137 (0.00332)	0.00577 (0.0115)	−0.00550 (0.00758)	0.000228 (0.00185)	0.00152 (0.00169)	−0.00125 (0.00413)	−0.00110 (0.00351)
Intelligent	0.00323 (0.00390)	0.00662** (0.00298)	−0.00669 (0.0127)	−0.00799 (0.00954)	0.00187 (0.00177)	0.00201 (0.00157)	0.00252 (0.00418)	0.00354 (0.00421)
Trust	0.00279 (0.00351)	−0.00221 (0.00267)	−0.00737 (0.0105)	0.00239 (0.00686)	0.00105 (0.00189)	0.000673 (0.00156)	0.00570* (0.00326)	0.00521 (0.00350)
Aggressive	−0.00156 (0.00347)	0.00182 (0.00272)	0.00711 (0.0109)	0.00154 (0.00836)	0.00174 (0.00179)	0.000554 (0.00151)	−0.00220 (0.00291)	−0.00109 (0.00444)
Likely winner	−0.000989 (0.00415)	−0.00258 (0.00320)	0.00514 (0.0114)	0.00115 (0.00833)	0.00172 (0.00228)	0.00235 (0.00216)	−0.00944*** (0.00357)	−0.00376 (0.00374)
MTurker controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lawyer FE	No	No	No	No	Yes	Yes	Yes	Yes
Observations	18542	35329	2543	5868	18542	35329	2543	5868

Note. Linear probability model with the raw perceptual ratings for masculinity/femininity, plus demographic controls for the rater (gender, age, region) and lawyer fixed effects if indicated. The dependent variable is whether the advocate won the Supreme Court case. Column 1 tests whether perceived masculinity is associated with case outcome for male petitioners controlling for rater characteristics. Column 2 tests whether perceived masculinity is associated with case outcome for male advocates controlling for rater characteristics. Columns 3–4 do the same as Columns 1–2 but among female advocates. Columns 5–8 do the same as Columns 1–4 but also control for lawyer fixed effects. Standard errors are clustered by case. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4. Other Outcomes and Perceived Masculinity.

	(1)	(2)	(3)	(4)
	Outcome	Judge Votes for Lawyer	# Votes for Lawyer	Outcome
Masculine	−0.0211** (0.00928)	−0.00829 (0.00676)	−0.0722 (0.0605)	0.0523 (0.0793)
Number of Democrats				0.0215 (0.0380)
Masculine* Number of Democrats				−0.0249 (0.0237)
Other Ratings	Y	Y	Y	Y
Collapsed	N	N	N	Y
N	153545	150304	153545	3689
R-sq	0.002	0.001	0.001	0.006

Note. Explores alternative dependent variables, e.g., margin of victory. Key regressor: perceived masculinity (raw or aggregated rating). This table presents results analyzing the correlation between perceptions of masculinity and Supreme Court case outcomes (in Column 1, the win rate), Supreme Court Justices' votes for the lawyer (in Column 2), the number of Justices voting for the lawyer (in Column 3), and the win rate collapsed to the lawyer-case level (in Column 4). Column 4 also includes the interaction between perceived masculinity and the number of Democrats on the Supreme Court. Other perceptions are always included as controls. Standard errors are clustered by case. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

of whether female versus male attorneys fare differently before the Court (e.g., [Szmer et al., 2010](#)). However, much of that research examines aggregate gender categories—male versus female attorneys—without disentangling the precise mechanisms at work, such as differences in rhetorical style, experience levels, or vocal delivery.

By contrast, the design here isolates vocal perceptions specifically, holding lexical content constant through a set phrase—“Mr Chief Justice, may it please the Court?”—that every advocate utters. This approach is intended to filter out differences in argument quality or strategy and focus on how acoustic features of speech alone (e.g., pitch, resonance) might shape listeners’ impressions of “masculinity” or “femininity.” In effect, we are looking at one dimension of how gendered signals can emerge—even among attorneys of the same gender—beyond simply classifying a lawyer as male or female. We do not dispute the prior literature’s demonstration that gender matters at the Court. Instead, we contribute a more granular question: How much do subjective impressions of an advocate’s voice—when they utter the identical words—predict case outcomes, above and beyond the broader male/female attorney effect?

Hence, although existing studies have addressed gender’s role in Supreme Court advocacy, none, to our knowledge, have systematically examined how differences in perceived vocal traits (despite identical phrasing) correlate with real vote and win/loss patterns. By spotlighting that narrower factor, we capture a dimension of extralegal influence—voice-based snap judgments—that operates at a level of detail previous “direct effects” research has not isolated.

Masculinity is a quality or set of practices that is stereotypically, though not exclusively, connected with men. Women may engage in masculine practices equally as often, although such practices are usually either not noticed or censured. The performative nature of “masculinity” makes possible the existence of non-masculine men and masculine women ([Kiesling, 2007](#); [Eckert & McConnell-Ginet, 2003](#); [Butler, 1990](#); [Kessler & McKenna, 1978](#)). Different cultures may also construct different notions of masculinity that are reflected in stereotypical ways of talking and thinking about men and masculinity.

The following sections identify three channels for the presence of these correlations. The next section investigates the role of judicial ideology and further assuage concerns of omitted case characteristics. The section after investigates the role of firms and the potential channel of information and taste in selecting lawyers with masculine voices.

5. Why Does Perceived Masculinity Predict SCOTUS Outcomes: The Role of Justices

5.1. Is Voice Masculinity a Response to Case Weakness?

Are pre-argument characteristics correlated with perceived masculinity? Using the Supreme Court Database⁷, we find that perceived masculinity still predicts outcomes after controlling for all

available pre-argument data. Column 1 of [Table 5](#) presents the analysis without additional controls. Column 2 adds whether the lower court had a dissent, was reversing the trial court decision, or made a liberal decision. Column 3 adds the number of words in the lower-court opinion, the number of self-certainty words, whether the lower court had both Democrats and Republicans on the three-judge panel, along with fixed effects for the Circuit and the legal topic. This restricts the sample to cases with non-missing data that came from the Circuit Courts (the remainder would come from state courts). Columns 4–6 include lawyer fixed effects.

5.2. Do Justices Differ in How Their Votes Correlate with Masculine Voices?

Votes of Democrats, but not Republicans, are negatively associated with perceived masculinity ([Figure 6](#)). In the higher intercept for Republicans, we see that Republicans favor male petitioners more so than Democrats do ([Szmer et al. \(2010\)](#)). The slope on the left side indicates that Democrats disfavor masculine males relative to less-masculine males – and the lack of a gradient for Republicans indicates that the adverse outcome for masculine males can be entirely attributed to Justices appointed by Democrats. All analyses in this sub-section regress:

$$judge\ votes\ for\ advocate_{itwj} = \alpha + \mathbf{attribute}'_{itwj}\beta + \mathbf{x}'_{itwj}\gamma + \varepsilon_{itwj}, \quad (9)$$

where w subscripts for rater and j for Justice. Standard errors are clustered at the lawyer and Justice levels, and Justice fixed effects are included. Interactions between $\mathbf{attribute}'_{itwj}$ and indicators for whether the Justice is a Democrat or Female are included in some models.

The heterogeneity along the dimension of Justices’ politics is not due to their gender. When both are included as interaction terms, only the interaction between perceived masculinity and political party is statistically significant ([Table 6](#)). Nor are experience and age driving factors. Even though only 13 individuals served as Supreme Court Justices during our time frame, the differences between Justices based on party affiliation are highly significant ($p < .01$).

If we replace party with Justices’ ideology scores ([Martin & Quinn, 2002](#)) or include both together, we find that ideology ($p < .01$) is more important than party, which is rendered insignificant ([Figure 7](#)).

We also find that Republicans vote for feminine-sounding females more than Democrats do ([Figure 8](#)). The difference is significant at $p < .1$. The basic pattern is again robust to considering the Justice’s gender ([Figure 9](#)).

5.3. How Does Predictive Power Compare with Previous Best Multivariate Models?

Having shown that judicial votes vary in the response to voice masculinity, we largely assuage concerns of omitted case

Table 5. Pre-argument Characteristics and Case Outcomes.

Dependent Variable: Case Outcome (Win = 1, Lose = 0)						
Masculine	−0.0153*** (0.00576)	−0.0125** (0.00576)	−0.0166** (0.00756)	−0.00546** (0.00225)	−0.00554** (0.00228)	−0.00389* (0.00235)
Aggressive	−0.000563 (0.00352)	0.0000108 (0.00350)	0.00181 (0.00424)	0.00228 (0.00173)	0.00263 (0.00175)	0.00217 (0.00164)
Attractive	0.00189 (0.00454)	−0.000299 (0.00443)	0.00189 (0.00570)	0.00202 (0.00179)	0.00172 (0.00172)	0.00101 (0.00191)
Confident	0.00554 (0.00396)	0.00390 (0.00384)	0.00199 (0.00461)	0.00284 (0.00191)	0.00258 (0.00196)	−0.00221 (0.00192)
Intelligent	0.00250 (0.00391)	0.00260 (0.00389)	0.00191 (0.00536)	0.000816 (0.00175)	0.000983 (0.00178)	0.00250 (0.00179)
Trust	0.00391 (0.00352)	0.00560 (0.00345)	0.00421 (0.00428)	0.00232 (0.00179)	0.00242 (0.00178)	0.000647 (0.00172)
Likely winner	−0.00194 (0.00421)	−0.00230 (0.00426)	−0.00435 (0.00518)	0.000736 (0.00219)	0.000825 (0.00225)	0.000916 (0.00204)
Lower court disagreement		0.107*** (0.0366)	0.142*** (0.0518)		0.0323 (0.0537)	−0.0659 (0.0786)
Lower court reversing		−0.0218 (0.0385)	−0.00438 (0.0483)		−0.0909* (0.0543)	−0.0964 (0.0703)
Trial court						
Lower court liberal		0.0750** (0.0355)	0.0406 (0.0471)		0.0155 (0.0592)	0.0127 (0.0771)
Decision						
Lower court opinion			−0.00612 (0.0134)			0.0203 (0.0256)
Word count						
Lower court opinion			−0.00130 (0.402)			−0.647 (0.710)
Self-certainty words						
Lower court politically			0.0977* (0.0559)			0.139* (0.0801)
Divided						
Circuit and topic fixed effects	No	No	Yes	No	No	Yes
Lawyer fixed effects	No	No	No	Yes	Yes	Yes
Observations	17782	17358	9542	17782	17358	9542
Sample: Male Petitioners						

Note. Adds variables capturing lower-court or pre-argument details (e.g., whether the lower court was reversed, the presence of dissent, circuit fixed effects) to see if they weaken the perceived-masculinity effect. Linear probability models analyzing the correlation between perceptions of masculinity and Supreme Court case outcomes, controlling for other perceptions (of the lawyers' aggressiveness, attractiveness, confidence, intelligence, trustworthiness, and winningness), pre-argument characteristics of the case included in the Supreme Court Database, as well as Circuit of origin fixed effects, topic fixed effects, and lawyer fixed effects. Standard errors are clustered by case. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

weakness. Here, we further assuage that concern using machine learning. We consider predictions using random forest tree methods (Katz et al., 2017; Guimerà & Sales-Pardo, 2011; Ruger et al., 2004; Martin et al., 2004). The latest model correctly identifies 70% of the Court's overall affirm / reverse decisions and 71% of the votes of individual justices from 1953–2013. It uses pre-argument characteristics (which collectively received roughly 25% of the importance weight) as well as court and Justice historical trends specific to issue, parties, and lower courts (which collectively received roughly 75% of the importance weight). Feature weights are presented in Table 7 (from Katz et al. (2017)).

We use this prediction of Justices' votes. Intuitively, if the predictive model were perfect, i.e., we essentially control for the correct decision or true case quality, then no extraneous factors should matter. We therefore ask two questions: Does perceived masculinity have an explanatory effect above and

beyond this best predictor? If so, how much R-square improvement?

Table 8 shows that the predicted vote correlates strongly with the actual Justice vote ($p < .01$) in Column 1. Column 2 examines the separate role of perceived masculinity ($p < .05$). Column 3 includes both, where both coefficients and standard errors are essentially unchanged from Columns 1 and 2. Two results emerge. First, perceived masculinity is orthogonal to observable pre-argument case characteristics, as seen in Section 5, and also orthogonal to court and Justice trends. Second, the R-square increases from 0.061 in Column 1 to 0.063 in Column 3, meaning that an additional 3% of variance can be explained. At the case level in Columns 4–6, an additional 10% of variance is explained, and the coefficient on perceived masculinity becomes substantially larger, about one-third the size of the correlation with the predicted vote.

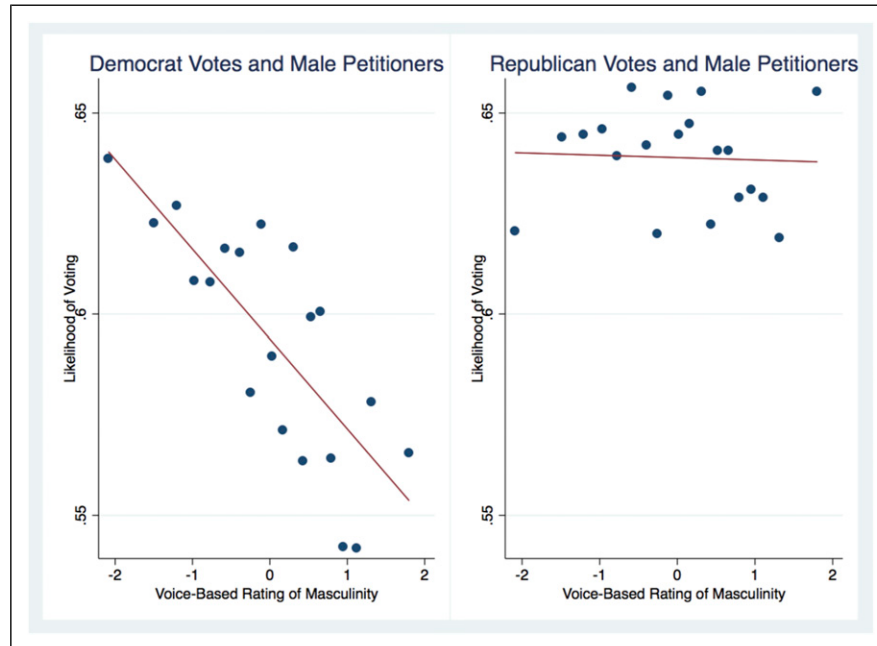


Figure 6. Political party and response to masculinity ($p < .01$). Note. Shows how Justices of different party of appointment respond to male petitioners' perceived masculinity (X-axis). Each bin on the X-axis represents an equal-sized subset of lawyers by rating. Y-axis measures predicted probability of Justice voting in favor.

Table 6. Political Party and Response to Male Petitioner Masculinity ($p < .01$).

	(I)
	Judge Votes for Lawyer
Masculinity	−0.000460 (0.00806)
Female Judge	−0.00554 (0.0194)
Masculinity * Female Judge	−0.00236 (0.00266)
Democrat	−0.0426* (0.0246)
Masculinity * Democrat Judge	−0.0203*** (0.00756)
Other Ratings	Y
Cluster	Lawyer and Judge
N	150304
R-sq	0.003

Note. Tests how Justices of different party affiliations (or appointment) respond to a male advocates perceived masculinity. Dependent variable: individual Justice votes. Model type (LPM or logit) indicated in columns. Interactions between the Justices' identity (female or male and Democrat or Republican) and masculinity are included. Standard errors are double-clustered by lawyer and by Justice. **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Why does perceived masculinity do so well relative to the best prediction model? Previous models perform best on cases on which the Court was in agreement (9–0) and performs worst on cases with high levels of disagreement

among members of the Court (5–4) (Katz et al. (2017)). In fact, in close cases affirming the lower court, the model predicts the outcome with only 25% accuracy.⁸ The interpretation that perceived masculinity improves prediction accuracy in hard cases is consistent with perceived masculinity being relevant with the swing voter.

5.4. How Much Do Perceptions Explain Beyond Objective Acoustic Data?

A contemporaneous study finds that lawyer's pitch is associated with Supreme Court outcomes (Dietrich et al. (2019)). Here, we analyze acoustic features of the clips using a set of many phonetic traits deemed as important by the phonology literature: (1) formant frequencies (F1, F2) for ve stressed vowels (/i, I, O, ej, 2/); (2) formant dispersion (average vowel distance from a central point per talker per sound clip); (3) spectral tilt (H1-H2, H1-A1, H1-A2, H1-A3); (4) sibilant's duration, four spectral moments, and peak frequency; (5) speaking rate (phonemes per second); (6) rhythm (Pairwise Variability Index); (7) F0 mean and standard deviation.

Table 9 presents our findings. Columns 1 and 2 present the association with perceived masculinity alone and with the additional perceptual variables. Column 3 adds all the acoustic controls. The association with perceived masculinity is basically invariant, despite perceived masculinity being correlated with many acoustic variables (Puts et al., 2011).

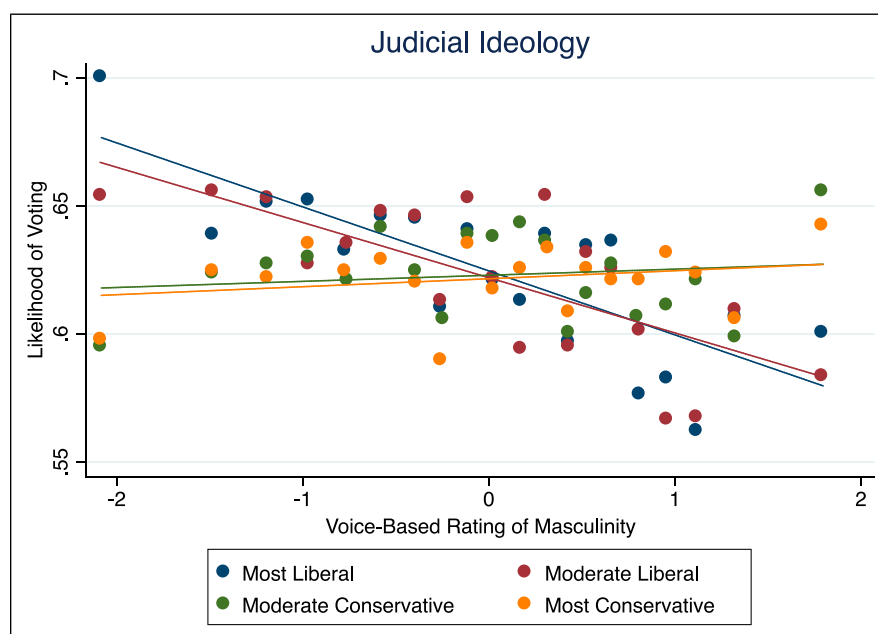


Figure 7. Judicial ideology and response to male petitioner masculinity ($p < .01$). *Note.* Shows how Justices of different ideological leanings respond to male petitioners' perceived masculinity (X-axis). Each bin on the X-axis represents an equal-sized subset of lawyers by rating. Y-axis measures predicted probability of Justice voting in favor. Judges are grouped by quartile of ideology.

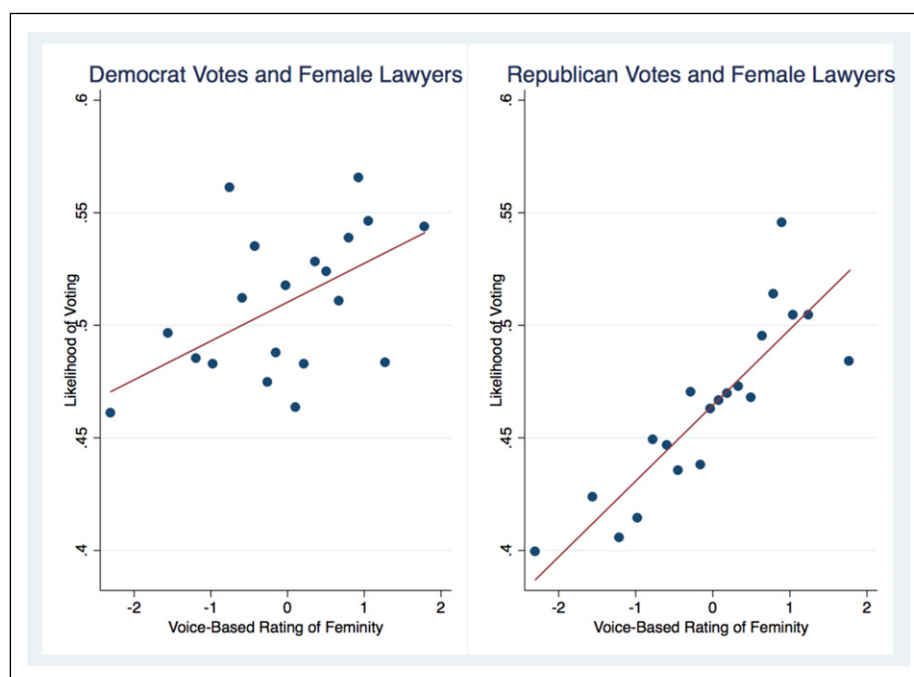


Figure 8. Political party and response to femininity ($p < .1$). *Note.* Shows how Justices of different party of appointment respond to female advocates' perceived femininity (X-axis). Each bin on the X-axis represents an equal-sized subset of lawyers by rating. Y-axis measures predicted probability of Justice voting in favor.

The invariance is possibly due to two reasons. First, while masculinity is correlated with the attributes we measure, like pitch, the same attributes are also correlated with many of the other perceptions (like confidence) that were not significantly

associated with case outcomes. Second, human perceptions matter beyond objective data: people can detect attributes and characteristics that computers cannot (White et al., 2015; Adler & Schuckers, 2007).

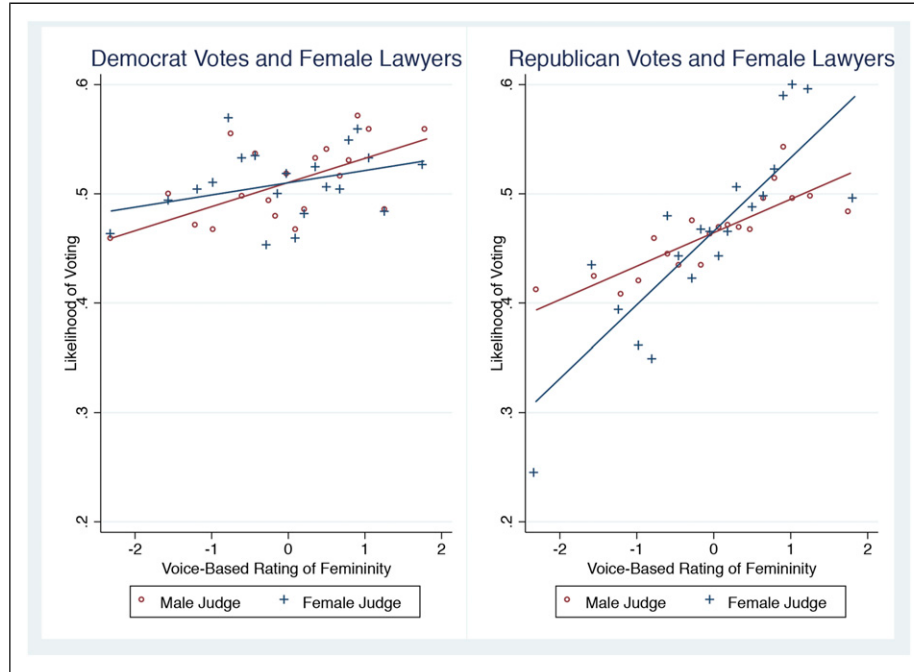


Figure 9. Political party and response to femininity ($p < .01$). Note. Shows how Justices of different gender and different party of appointment respond to female advocates' perceived femininity (X-axis). Each bin on the X-axis represents an equal-sized subset of lawyers by rating. Y-axis measures predicted probability of Justice voting in favor.

6. Why Does Perceived Masculinity Predict SCOTUS Outcomes: The Role of Firms

6.1. Is Masculinity of Voice Differentially Correlated with Court Outcomes Across Firms?

We now address the role of firm selection of lawyers with masculine voices more directly using the available data. In the model, industries that value $d > 0$ more highly should observe a steeper correlation between perceived masculinity and losses because these firms have a greater relative taste for M lawyers than for F lawyers.

Prior research finds that, in the private sector, CEOs with deeper voices manage larger companies (Mayew & Venkatachalam, 2012), which may lead private firms to prefer masculine-sounding lawyers. We begin by simply dividing the set of firms into two categories: not-for-profit (e.g., government) versus private. Not-for-profit firms' lawyers are slightly less masculine than private firm lawyers ($p < .1$) (Figure 11 and Table 10). We also observe stronger negative correlations between perceived masculinity and court outcomes among private firms (Figure 10), and in industries with more masculine voices ($p < .05$) (Figure 11), and essentially no correlation between perceived masculinity and win rates for public lawyers advocating on behalf of the government.

To be more systematic, we next construct an average masculinity rating for every class of petitioners. Perceived masculinity is negatively correlated with outcomes for

industries that have the most masculine-sounding lawyers. Figure 11 presents a gradient for four industry quartiles according to masculinity. The steepest gradient is for the most masculine and the flattest for the least masculine, which is consistent with differences in $d > 0$.⁹

6.2. Information and Incentives

We now turn to our de-biasing experiment. Let y_{ij} represent attitudes in treatment $j = 1, 2, 3, 4$ and recording i , F_j be an indicator variable that takes a value of 1 if raters receive information about whether the attorney on the recording won the case or not, and 0 otherwise. I_j indicates whether the raters were given an incentive to choose correctly ($I_j = 1$) or not ($I_j = 0$).

The estimation equation takes the following form:

$$y_{ij} = \alpha + \beta V_i + \gamma' T_j + \delta' T_j \times V_i + e_{ij},$$

where $V_i \in \{H, L\}$ is advocate i 's voice type and $T_j \in \{[F_j], [I_j]\}$ is a treatment vector (2×2 treatment). The incentives treatment entailed an additional \$0.5 for raters who get two-thirds or more of the outcomes correct.

Figure 12 shows that information reduces 40% of the correlation between perceived masculinity and perceived win ($p < .05$). In fact, all ratings' positive correlations with perceived wins are affected by information but not incentives. Recall that none of the other ratings predicted actual wins, but

Table 7. Final Feature Weights (Katz et al., 2015).

Justice and Court Background Information		Current Supreme Court Trends	
Justice [S]	0.00781	Mean Agreement Level of Current Court [FE]	0.00955
Justice Gender [FE]	0.00205	Std. Dev. of Agreement Level of Current Court [FE]	0.00936
Is Chief [FE]	0.00283	Mean Current Court Direction Circuit Origin [FE]	0.00789
Party President [FE]	0.00604	Std. Dev. Current Court Direction Circuit Origin [FE]	0.00850
Natural Court [S]	0.00764	Mean Current Court Direction Circuit Source [FE]	0.00945
Segal Cover Score [SC]	0.00971	Std. Dev. Current Court Direction Circuit Source [FE]	0.01021
Year of Birth [FE]	0.00793	Mean Current Court Direction Issue [FE]	0.01469
TOTAL	0.04403	Z-Score Current Court Direction Issue [FE]	0.00832
Case Information		Std. Dev. Current Court Direction Issue [FE]	0.01266
Admin Action [S]	0.00978	Mean Current Court Direction [FE]	0.00918
Case Origin [S]	0.00971	Std. Dev. Current Court Direction [FE]	0.00942
Case Origin Circuit [S]	0.00845	Mean Current Court Direction Petitioner [FE]	0.00863
Case Source [S]	0.00953	Std. Dev. Current Court Direction Petitioner [FE]	0.00894
Case Source Circuit [S]	0.01015	Mean Current Court Direction Respondent [FE]	0.00882
Law Type [S]	0.01370	Std. Dev. Current Court Direction Respondent [FE]	0.00888
Lower Court Disposition Direction [S]	0.01190	TOTAL	0.14456
Lower Court Disposition [S]	0.01125	Individual Supreme Court Justice Trends	
Lower Court Disagreement [S]	0.00706	Mean Justice Direction [FE]	0.01248
Issue [S]	0.01541	Mean Justice Direction 10 [FE]	0.01530
Issue Area [S]	0.01469	Mean Justice Direction Z Score [FE]	0.00826
Jurisdiction Manner [S]	0.00595	Mean Justice Direction Petitioner [FE]	0.00732
Month Argument [FE]	0.02014	Mean Justice Direction Petitioner 10 [FE]	0.01027
Month Decision [FE]	0.01349	Mean Justice Direction Respondent [FE]	0.00724
Petitioner [S]	0.01406	Mean Justice Direction Respondent 10 [FE]	0.01030
Petitioner Binned [FE]	0.01199	Mean Justice Direction for Circuit Origin [FE]	0.00792
Respondent [S]	0.01490	Mean Justice Direction for Circuit Origin 10 [FE]	0.00945
Respondent Binned [FE]	0.01179	Mean Justice Direction for Circuit Source [FE]	0.00891
Cert Reason [S]	0.01408	Mean Justice Direction for Circuit Source 10 [FE]	0.00970
TOTAL	0.22814	Mean Justice Direction by Issue [FE]	0.01881
Overall Historic Supreme Court Trends		Mean Justice Direction by Issue 10 [FE]	0.00950
Mean Court Direction [FE]	0.00988	Mean Justice Direction by Issue Z Score [FE]	0.00771
Mean Court Direction 10 [FE]	0.01997	TOTAL	0.14323
Mean Court Direction Issue [FE]	0.01546	Differences in Trends	
Mean Court Direction Issue 10 [FE]	0.00938	Difference Justice Court Direction [FE]	0.01210
Mean Court Direction Petitioner [FE]	0.00863	Abs. Difference Justice Court Direction [FE]	0.00929
Mean Court Direction Petitioner 10 [FE]	0.00904	Difference Justice Court Direction Issue [FE]	0.01167
Mean Court Direction Respondent [FE]	0.00875	Abs. Difference Justice Court Direction Issue [FE]	0.00968
Mean Court Direction Respondent 10 [FE]	0.00925	Z Score Difference Justice Court Direction Issue [FE]	0.01055
Mean Court Direction Circuit Origin [FE]	0.00791	Difference Justice Court Direction Petitioner [FE]	0.00705
Mean Court Direction Circuit Origin 10 [FE]	0.00864	Abs. Difference Justice Court Direction Petitioner [FE]	0.00708
Mean Court Direction Circuit Source [FE]	0.00951	Difference Justice Court Direction Respondent [FE]	0.00690
Mean Court Direction Circuit Source 10 [FE]	0.01017	Abs. Difference Justice Court Direction Respondent [FE]	0.00699
TOTAL	0.12663	Z Score Justice Court Direction Difference [FE]	0.01280
Lower Court Trends		Justice Lower Court Direction Difference [FE]	0.01922
Mean Lower Court Direction Circuit Source [FE]	0.00962	Justice Lower Court Direction Abs. Difference [FE]	0.02494
Mean Lower Court Direction Circuit Source 10 [FE]	0.01017	Justice Lower Court Direction Z Score [FE]	0.01126
Mean Lower Court Direction Issue [FE]	0.01334	Z Score Justice Lower Court Direction Difference [FE]	0.00992
Mean Lower Court Direction Issue 10 [FE]	0.00933	Agreement of Justice with Majority [FE]	0.00866
Mean Lower Court Direction Petitioner [FE]	0.00949	Agreement of Justice with Majority 10 [FE]	0.01483
Mean Lower Court Direction Petitioner 10 [FE]	0.00874	Difference Court and Lower Ct Direction [FE]	0.01522
Mean Lower Court Direction Respondent [FE]	0.00973	Abs. Difference Court and Lower Ct Direction [FE]	0.01199
Mean Lower Court Direction Respondent 10 [FE]	0.00900	Z-Score Difference Court and Lower Ct Direction [FE]	0.01217
TOTAL	0.07946	Z-Score Abs. Difference Court and Lower Ct Direction [FE]	0.01150
		TOTAL	0.23391

Note. These random-forest feature importances show how strongly each variable contributes to predicting the Supreme Courts overall affirm versus reverse decisions in Katz et al.'s model. Not a regression table. We adopt this "best predictor" approach as a comprehensive control (see Section 5.3). Features range from lower-court data to justice-level ideology.

ratars generally associated all the attributes with a greater likelihood of winning.

When we fully interact the treatments in Figure 13, we observe that incentives further reduce the correlation. Recall that incentives

to choose correctly erode the effect of taste on choices ($\pi_F - \pi_M > d/\alpha$). Providing incentives should bring $d \rightarrow 0$. The existence of the effect ($p < .05$, 30% of association with feedback) also means $d > 0$ (else d/α would be 0 regardless of the size of α).

Table 8. Best Prediction and Perceived Masculinity.

		Judge Votes for Lawyer			
Predicted vote from random forest	0.257*** (0.0486)		0.258*** (0.0487)	0.250*** (0.0485)	0.248*** (0.0489)
Masculine		−0.0223** (0.0101)	−0.0207** (0.0101)	−0.0852** (0.0359)	−0.0780** (0.0361)
Cluster		Lawyer and Judge			
Collapsed	No	No	No	Yes	Yes
Observations	26447	26391	26391	1229	1229
R-squared	0.061	0.002	0.063	0.058	0.008
Sample: Male petitioners, Democrat judges					

Note. Using the raw perceptual ratings, this table tests whether perceived masculinity is associated with Democrat Justices' votes for the lawyer controlling for a prediction of Supreme Court votes using the [Katz et al. \(2015\)](#) model. Analyses are conducted at both the raw perceptual rating level and collapsed at the advocate-case level. Standard errors are double-clustered by lawyer and by Justice. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 9. Case Outcomes and Perceived Masculinity.

Dependent Variable: Case Outcome (Win = 1, Lose = 0)			
Masculine	−0.0875** (0.0369)	−0.0972** (0.0364)	−0.0858** (0.0348)
Confident		0.0258 (0.0247)	0.0360 (0.0220)
Attractive		−0.0171 (0.0181)	−0.0197 (0.0144)
Educated		0.0158* (0.00878)	0.0146 (0.00932)
Intelligent		0.00549 (0.00893)	0.00635 (0.00783)
Trust		−0.00512 (0.00979)	−0.00528 (0.00786)
Likely winner		−0.00355 (0.00793)	−0.00132 (0.00729)
Acoustic Controls	No	No	Yes
Observations	10920	10080	10080

Note. Summarizes regressions linking perceived masculinity to Supreme Court success, controlling for other acoustic factors. We analyze acoustic features of the clips, i.e., (1) formant frequencies (F1, F2) for ve stressed vowels, (2) formant dispersion (average vowel distance from a central point per talker per sound clip), (3) spectral tilt (H1-H2, H1-A1, H1-A2, H1-A3), (4) sibilants duration, four spectral moments, and peak frequency, (5) speaking rate (phonemes per second), (6) rhythm (Pairwise Variability Index) (7) F0 mean and standard deviation. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

6.3. Whose Perceptions of Masculinity Predict Outcomes?

Regardless of the reason for Justices to vote differently according to voice masculinity, lawyers should adjust, unless another audience, like employers, matters. We now investigate two issues regarding whether perceptions of voice might differ by rater characteristic. First, do certain individuals' perceptions of masculinity (possibly those of law firm employers) predict court outcomes? Second, do certain individuals (possibly law firm employers) perceive masculine voices to be more likely to win?

Table 10. Public versus Private Firms and Response to Masculinity.

	Masculine Rating
Public	−0.0480* (0.0287)
Observations	1580

Note. Splits the sample by public (government, nonprofit) versus private (corporations). This table tests whether perceived masculinity of the advocate is associated with whether the lawyer is advocating on behalf of a public or private entity. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

We begin by analyzing the raters' gender to benchmark our findings. Previous studies find that members of the same sex may use certain criteria to rate each other, which differ from those used to rate members of the opposite sex ([Babel et al., 2014](#)). [Table A.5](#) shows that partial correlations among perceived attributes of men and women differ when rated by men or by women.

The associations stem more (but not significantly more) from the judgments made by male participants rating male lawyers and female participants rating female lawyers ([Figure 14](#)).

[Tables 11](#) and [A.6](#) show that the patterns discussed thus far are somewhat accentuated when the genders of the rater and advocate match. Male ratings of the masculinity (femininity) of male (female) advocates' voices are significantly negatively correlated with whether the male advocate wins, while female ratings of the femininity of female advocates' voices are positively correlated with whether the female advocate wins.¹⁰ Among petitioners, a one-standard-deviation increase in perceived masculinity measured using male perceptions predicts a 2.3 percentage-point lower likelihood of winning the case if the advocate is male. Using female perceptions, a similar change in perceived femininity predicts a 3.2 percentage-point higher likelihood of winning the case if the advocate is female. These associations are displayed in [Figure 4](#). The results for unmatched gender specification have substantially smaller coefficient estimates and are less

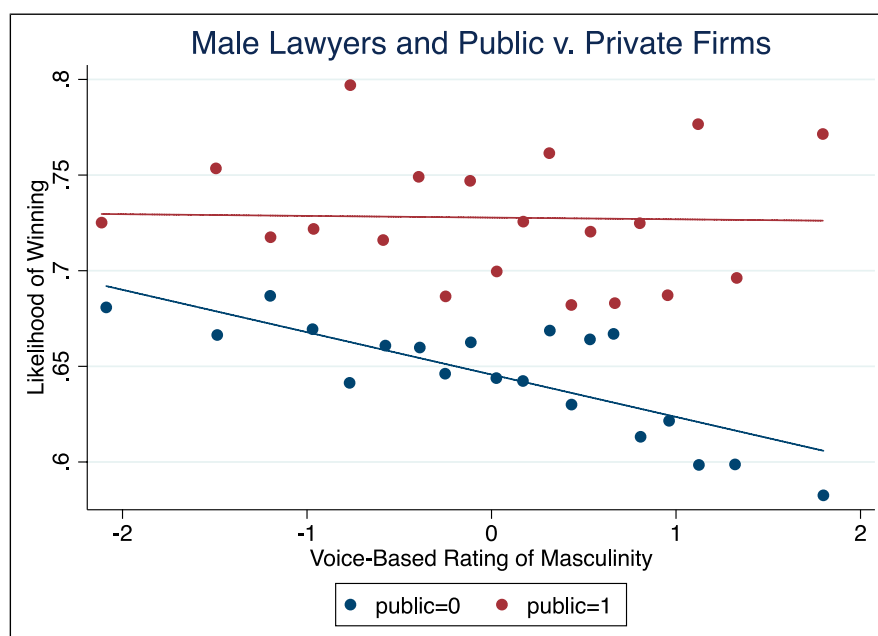


Figure 10. Public versus private firms and response to masculinity. *Note.* Illustrates how attorneys arguing on behalf of the government or a private firm fare given different ranges of perceived vocal masculinity. X-axis = perceived masculinity/femininity (from lower to higher). Each bin on the X-axis represents an equal-sized subset of lawyers by rating. Y-axis = predicted probability of winning.

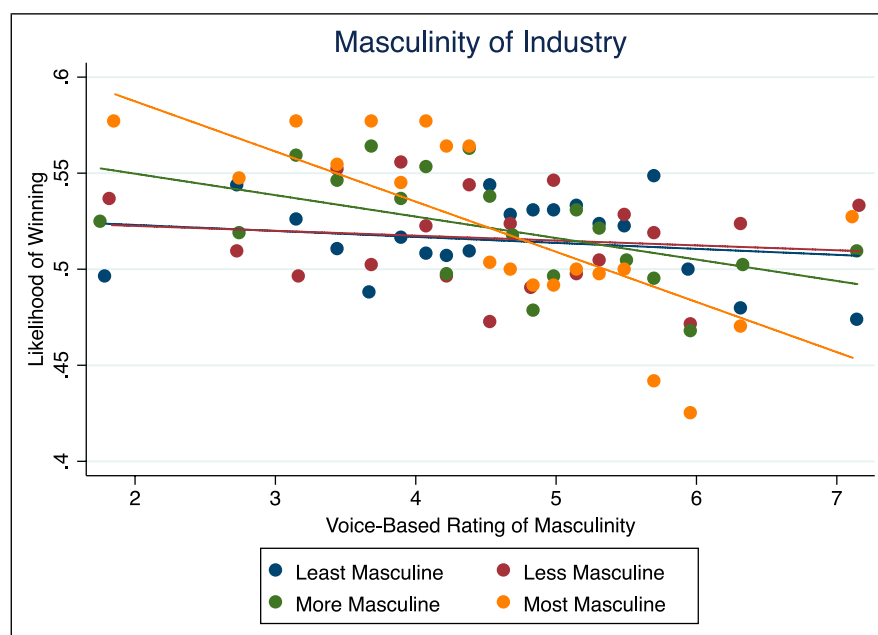


Figure 11. Masculinity of industry and response to masculinity ($p < .05$). *Note.* We classify petitioner industries by their average perceived masculinity and plot the correlation between perceived masculinity (X-axis) with case win rates (Y-axis). Each bin on the X-axis represents an equal-sized subset of lawyers by rating. Data suggest that in industries with the highest overall "masculine-sounding" lawyers, perceived masculinity negatively correlates with success. Section 6.1 details these groupings.

statistically distinguishable from 0, which can be seen in Figure 14.

Next, we investigate other demographic characteristics that play a bigger role than gender. Specifically, we

observe greater heterogeneity according to the rater's self-identified race and region (Figures 15 and 16). White ($p < .05$) and non-Southerners' ($p < .05$) perceptions of masculinity are more significantly predictive of court

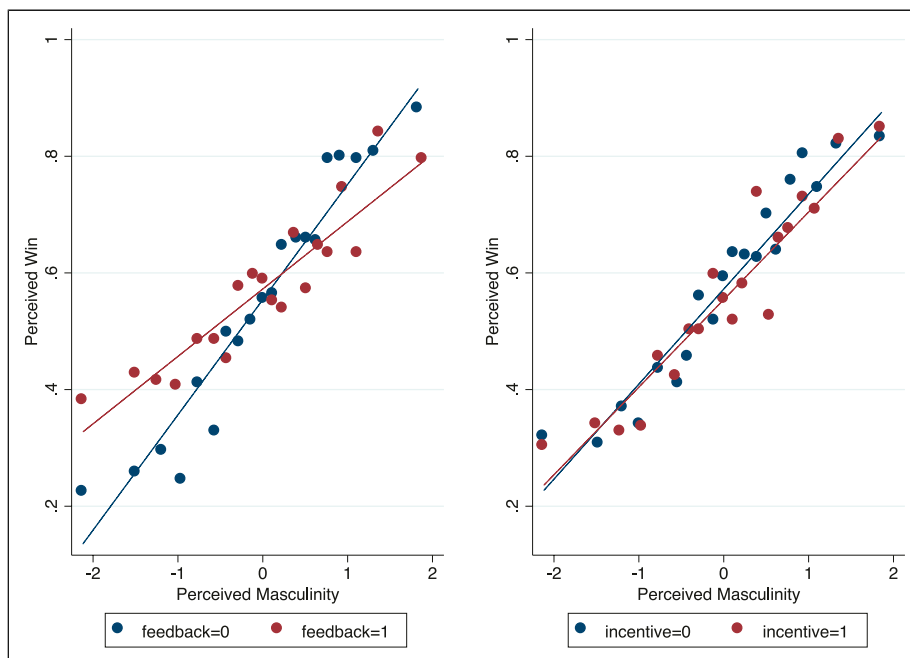


Figure 12. Feedback ($p < .01$), incentives. *Note.* De-biasing Experiment: Incentives and Feedback. Each binscatter line depicts the change in correlation between perceived masculinity and perceived win rate under different experimental conditions: (1) No Feedback (regardless of incentive), (2) Feedback (regardless of incentive), (3) No Incentive (regardless of feedback), (4) Incentive (regardless of feedback). Incentives = small bonus for accurate guesses. Feedback = telling participants the actual outcome after each rating. See Section 6.2 for the 22 design.

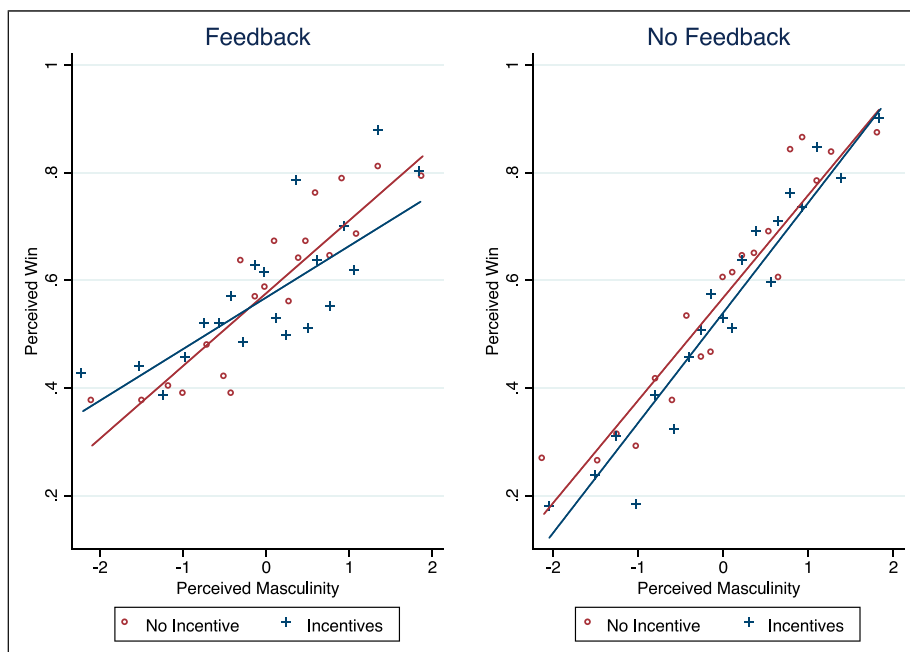


Figure 13. Incentives ($p < .05$) with feedback. *Note.* De-biasing Experiment: Incentives and Feedback. Each binscatter line depicts the change in correlation between perceived masculinity and perceived win rate under different experimental conditions: (1) No Feedback + No Incentive, (2) Feedback + No incentive, (3) No Feedback + Incentive, (4) Feedback + Incentive. Incentives = small bonus for accurate guesses. Feedback = telling participants the actual outcome after each rating. See Section 6.2 for the 2×2 design.

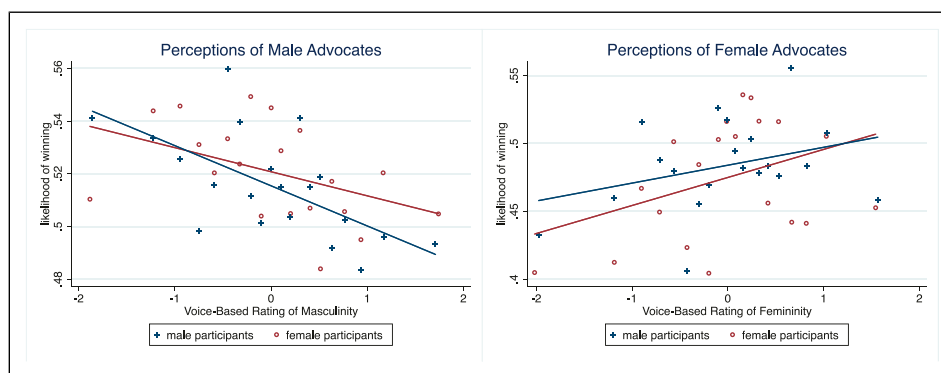


Figure 14. Voice-based perceptions and court outcomes by advocate and participant gender. *Note.* Shows how male versus female raters perceive male versus female advocates, and how these differing perceptions correlate with Court outcomes. For instance, male raters of male advocates versus female raters of female advocates. Y-axis = average share of cases won. Section 6.3 describes the gender match analysis in detail.

Table 11. Voice Perceptions and Gender Match (Raw Ratings, Micro Level).

Dependent Variable: Case Outcome (Win = 1, Lose = 0)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Male MTurk Workers				Female MTurk Workers			
	Male Advocate		Female Advocate		Male Advocate		Female Advocate	
	Petitioner	Both	Petitioner	Both	Petitioner	Both	Petitioner	Both
Masculine/Feminine	−0.0156** (0.00629)	−0.0122** (0.00486)	0.00332 (0.0131)	0.00958 (0.00934)	−0.0130** (0.00648)	−0.00704 (0.00496)	0.0200 (0.0121)	0.0146* (0.00809)
Confident	0.00477 (0.00512)	0.00465 (0.00394)	−0.0112 (0.0153)	0.00514 (0.00973)	0.00513 (0.00500)	0.00340 (0.00391)	−0.00579 (0.0106)	0.00861 (0.00807)
Attractive	0.00564 (0.00515)	0.00302 (0.00385)	0.00872 (0.0152)	−0.00561 (0.0109)	−0.00493 (0.00559)	−0.00445 (0.00421)	−0.00153 (0.0126)	−0.00597 (0.00858)
Intelligent	0.00797 (0.00490)	0.00843** (0.00377)	−0.0128 (0.00985)	−0.00789 (0.00895)	−0.00354 (0.00530)	0.00345 (0.00397)	0.00785 (0.0156)	−0.00390 (0.0107)
Trust	0.000508 (0.00478)	−0.00396 (0.00366)	0.00345 (0.0153)	0.00306 (0.00961)	0.00696 (0.00491)	0.000976 (0.00364)	−0.0120 (0.0141)	−0.00494 (0.00936)
Aggressive	−0.00156 (0.00457)	−0.000222 (0.00351)	0.00797 (0.0156)	−0.000861 (0.0112)	−0.00111 (0.00424)	0.00411 (0.00333)	0.00356 (0.0103)	0.00435 (0.00794)
Likely winner	−0.000237 (0.00545)	0.000112 (0.00411)	−0.00109 (0.0158)	0.000111 (0.0111)	−0.00184 (0.00554)	−0.00559 (0.00442)	0.00496 (0.0127)	0.00579 (0.00931)
Observations	9179	17435	1198	2714	9363	17894	1345	3154

Note. Tests how the rater's gender matching the advocate's gender shapes the correlation between masculinity/femininity and Supreme Court win probability. Data at the rater-lawyer level. Column 1 tests whether perceived masculinity is associated with case outcomes for male petitioners rated by male raters. Column 2 tests whether perceived masculinity is associated with case outcomes for male advocates rated by male raters. Columns 3–4 do the same as Columns 1–2 but for female advocates. Columns 5–8 do the same as Columns 1–4 but for ratings by female raters. Standard errors are clustered by case. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

outcomes than other demographic groups' perceptions.¹¹ If White non-Southerners are more involved in the process of selecting oral advocates, the stronger correlations would be consistent with the important role of the firm in explaining the negative correlation between masculinity and court outcomes.

Notably, White and non-Southerners' perceptions of femininity are also more predictive of court outcomes, though the difference is less salient than for masculinity (Figure 17).

When we examine raters' perceptions of winning and masculinity, we see that poor individuals ($p < .05$) and

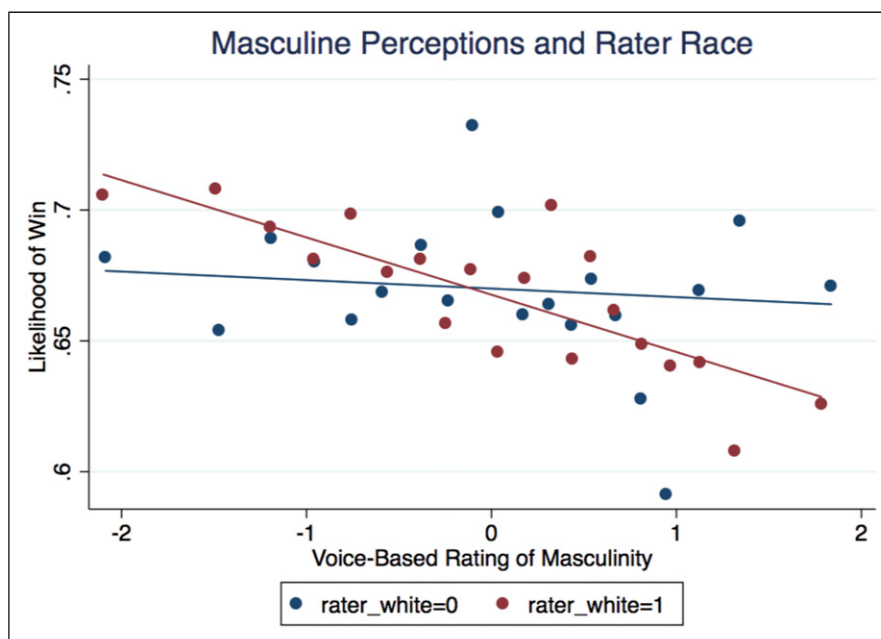


Figure 15. White raters' perceptions of masculinity predicted court outcomes ($p < .05$). *Note.* Shows how White versus non-White raters perceive male advocates, and how these differing perceptions correlate with Court outcomes. Y-axis = average share of cases won. For more on rater demographic splits, see Section 6.3.

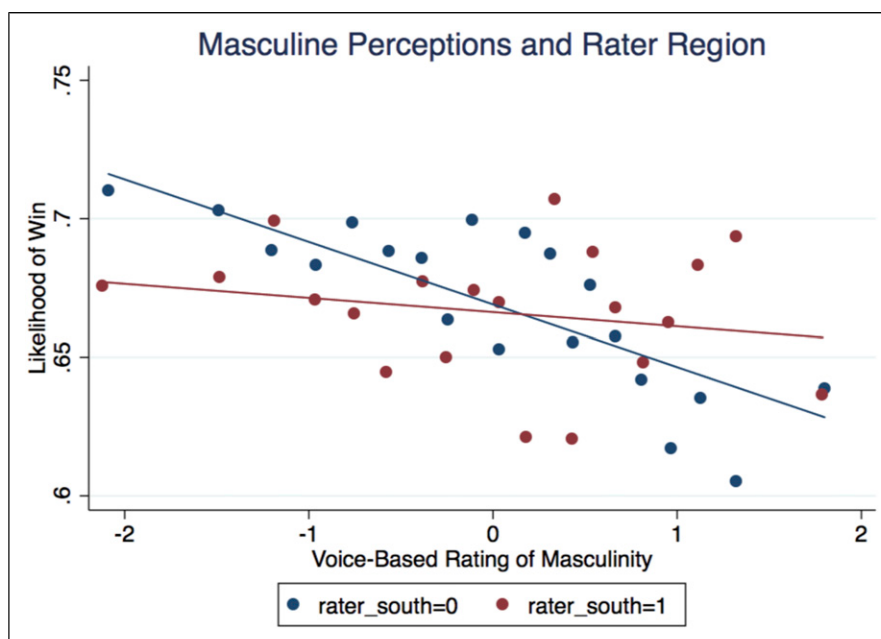


Figure 16. Non-Southerners' perceptions of masculinity predicted court outcomes ($p < .05$). *Note.* Shows how Southerner versus non-Southerner raters perceive male advocates, and how these differing perceptions correlate with Court outcomes. Y-axis = average share of cases won. For more on rater demographic splits, see Section 6.3.

non-whites ($p < .05$) are significantly less likely to correlate masculinity with winning. Poor individuals ($p < .05$), males ($p < .1$), and the less-educated ($p < .05$), are significantly less likely to correlate femininity with

winning. If these individuals are less likely to be selecting oral advocates, the results are consistent with the role of firms in selecting masculine-sounding males as Supreme Court oral advocates. Perceptions of

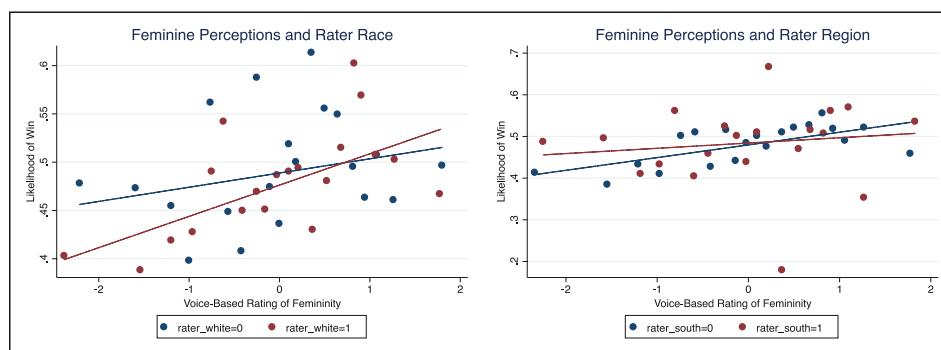


Figure 17. White non-Southerners' perceptions of femininity also predict court outcomes ($p > .1$). *Note.* Shows how White versus non-White raters and Southerner versus non-Southerner raters perceive female advocates, and how these differing perceptions correlate with Court outcomes. Y-axis = average share of cases won. For more on rater demographic splits, see Section 6.3.

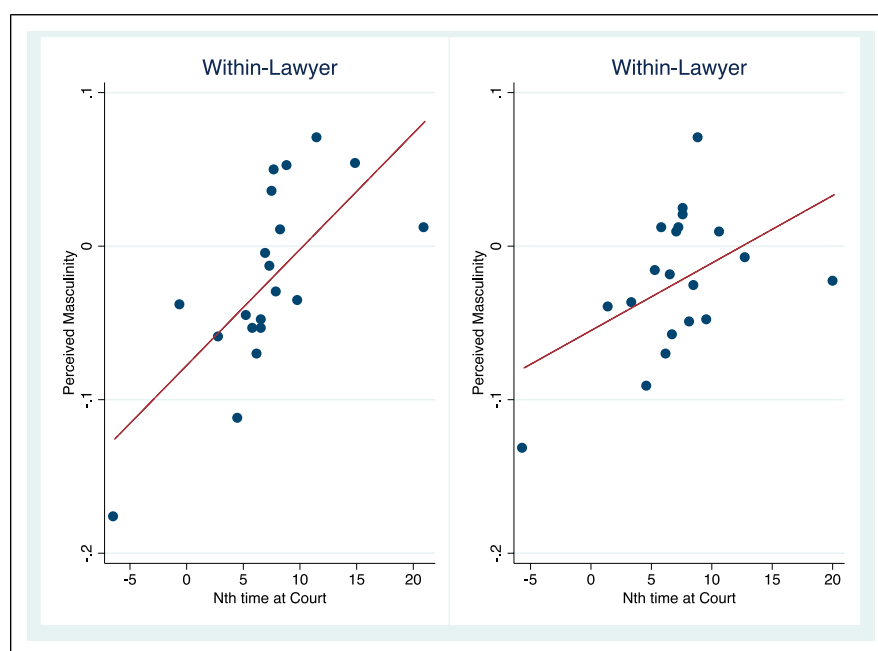


Figure 18. Masculine ratings and order of appearance. *Note.* We track how advocates' perceived masculinity changes across multiple appearances before the SCOTUS. X-axis = ordinal index of each argument (first, second, third appearance, etc.). Y-axis = average perceived masculinity. These graphics are residualized on lawyer fixed effects. The right-side graphic also residualizes for years since law school graduation. Section 6.4 explores whether lawyers become more or less masculine-sounding over time.

masculinity are subjective and not everyone perceives the same voice as masculine. Different stereotypes persist, which may drive behavioral responses.

6.4. What Are Lawyers Doing Over Time?

Voice appears to be mutable, and perceived masculinity is negatively correlated with winning in the Supreme Court. Do lawyers learn over time to make their voice less masculine? On the contrary, lawyers' perceived masculinity increase over time as they appear before the SCOTUS more often. This is surprising since it would seem to suggest that lawyers are not

learning and instead, adapt their voice in the opposite direction. Figure 18 visualizes how individual lawyers become more masculine over time. The left side presents the raw correlation, residualizing only on lawyer fixed effects. The right side also controls for years since law school graduation. Figure 19 presents the cross-sectional variation without lawyer fixed effects.

Table 12 rejects the hypothesis of lawyer learning from losses to decrease their voice's perceived masculinity in the following case. Column 1 repeats the negative association between masculinity rating and court outcome. Column 2 presents the correlation between a loss in the previous

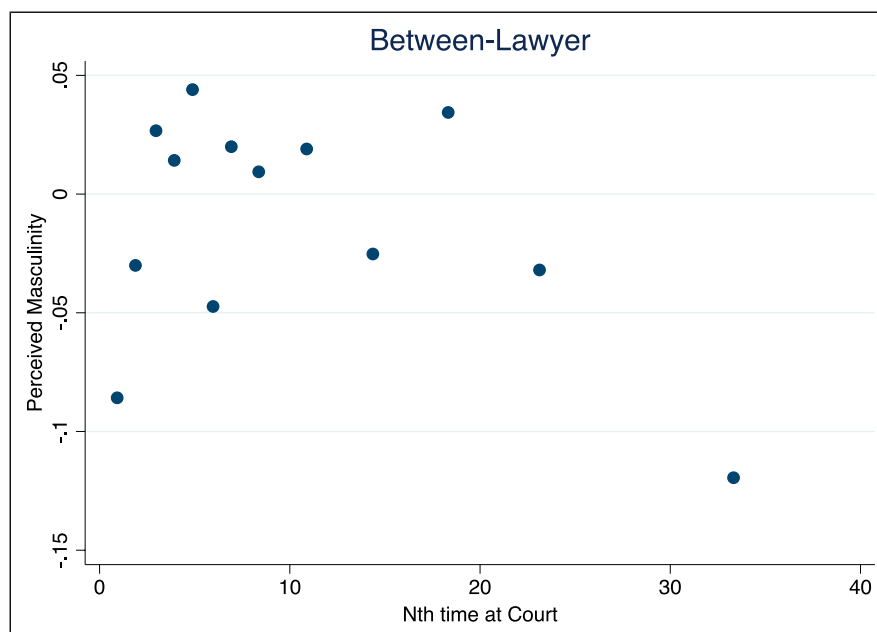


Figure 19. Cross-sectional variation. *Note.* We track how advocates’ perceived masculinity changes across multiple appearances before the SCOTUS. X-axis = ordinal index of each argument (first, second, third appearance, etc.). Y-axis = average perceived masculinity. These graphics are not residualized on lawyer fixed effects to illustrate the cross-sectional variation. Section 6.4 discusses how different cohorts or experience levels show variation in average ratings.

Supreme Court appearance and the masculinity rating in the current Supreme Court appearance. The positive association means a significant rejection of learning/modifying one’s voice to be less masculine in response to a loss. Column 3 presents the correlation between a previous loss and a change in the masculinity rating from the previous appearance to the current appearance. It is negative, but not significantly so. The contrast between Columns 2 and 3 is consistent with the selection for masculinity, among lawyers who appear repeatedly at the Supreme Court.

7. Discussion

Our approach integrates dual-process theories of cognition (Kahneman, 2011) and existing research on judicial behavior (Rachlinski et al., 2009a), proposing that vocal attributes act as an extralegal “System 1” heuristic in situations of uncertainty.

Even when Justices are extensively briefed, the sheer complexity of many cases—plus the limited time for oral argument—can encourage “on the spot” intuitions. A brief, deeply pitched utterance might subconsciously convey authority or confidence, while a higher, more “feminine” pitch might signal approachability or clarity (Klofstad et al., 2012). Our data show that these impressions, formed in seconds, correlate with final votes and case outcomes.

We hypothesize that voice may be a proxy for attorney quality (confidence, competence) or a stand-in for agreement with the Justice’s worldview. If a Justice is uncertain about the

legal merits—or perceives both sides’ doctrinal arguments as roughly equal—vocal cues might tilt the balance. To test this, our empirical design includes controls for actual attorney credentials (e.g., prior SCOTUS arguments, clerkship experience), perceptions of confidence, intelligence, among other traits, and ideological salience (e.g., whether the case involves politically charged subject matter). If voice retains a significant effect even after these controls, it suggests the heuristic is not merely substituting for measured attorney skill or ideological alignment.

Our analysis also splits the sample by case complexity or level of disagreement (e.g., 5–4 vs. 9–0 decisions) to see whether the voice-based correlation is stronger in ambiguous or closely contested scenarios. Consistent with Guthrie et al. (2007) and Schubert et al. (2002), the data indicate that extralegal signals become more salient when the law does not provide a clear resolution or the Court is split ideologically. In these moments, “System 2” deliberation may be insufficient to entirely override the immediate, intuitive impressions formed as an advocate begins to speak.

While major ideological commitments or well-settled precedent can dominate a Justice’s vote, System 1 impressions can shape the trajectory of oral argument—e.g., by influencing the tone or aggressiveness of questioning. This dynamic can subtly color how the Court processes each side’s argument, even if, by opinion-writing stage, the Justice attempts a more deliberative approach.

We include appointing-president partisanship as one measure of judicial attitude, consistent with long-standing

Table 12. Are Lawyers Learning?

	Outcome	Masculinity Rating	Δ Masculinity Rating
Masculinity Rating	−0.0605** (0.0308)		
Previous Loss		0.0445 (0.0551)	−0.0463 (0.0359)
Collapsed	Y	Y	Y
Observations	862	389	389
R-squared	0.004	0.002	0.004
Sample: Male Petitioners			

Note. Investigates whether lawyers change their vocal masculinity after a loss. Unit of analysis: lawyer across multiple SCOTUS appearances. Key regressor: lost last case (t-1). Dependent variable: difference in perceived masculinity from t-1 to t or raw masculinity at time t. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

empirical practice in political science (Segal & Spaeth, 2002; Sunstein et al., 2006). The party of appointment has historically served as an imperfect yet widely accepted proxy for a Justice's underlying ideological orientation—especially across decades. To refine this approach, we supplement the appointing-president variable with Martin–Quinn scores (Martin & Quinn, 2002), yielding a continuous measure of ideology over our entire sample.

Regarding the ideologies of the court of origin, we draw on the feature set developed by Katz et al. (2017), which leverages the Supreme Court Database (SCDB) and encodes each vote and case through a rich collection of variables. These include (i) basic categorical fields—such as petitioner, respondent, issue, case origin, and lower-court disposition—converted into numerous binary (dummy) indicators; (ii) chronological markers (term, month of argument, days from argument to decision); (iii) engineered features capturing each Justice's historical behavior (e.g., reversal rate, proportion of liberal/conservative votes, dissent rate) and the difference between a Justice's tendencies and the Court's overall patterns; and (iv) court-of-origin groupings (federal circuits vs. state or administrative courts). By binarizing SCDB variables and adding these summary statistics, Katz et al. track both the 'who, what, and where' of a case and each Justice's prior voting trajectory. Although they do not incorporate a specialized metric for state supreme court ideology, their random forest accommodates case origin, circuit, or state court category, and it infers broader patterns of judicial ideology via historical vote distributions. By including court of origin dummy indicators and year in the feature set, the random forest automatically incorporates time-varying ideological measures of the court of origin. This design ensures a comprehensive set of controls for case-level, judge-level, and procedural factors when predicting or explaining Supreme Court outcomes.

Our study relies on audio from 1998–2012 primarily because, during that period, Oyez.org provided high-quality, consistently labeled recordings that reliably matched transcripts at the individual speaker level. While Oyez originated at Chicago-Kent, it has since migrated hosting and consolidated its archive; we accessed the site under its historical affiliation at the time of data collection. Although Oyez now offers more recent material, that

content was not fully synchronized or standardized at our project's inception. Looking forward, additional analyses already support the generality of our findings. Dietrich et al. (2019), for instance, adopt a broader window without restricting to a single sentence or relying on human-subject impressions, and still find that higher-pitched advocates are likelier to succeed. Similarly, Chen (2018) leverages the entire archive from 1955 onward, combining facial and vocal cues to improve prediction of Supreme Court outcomes. These replications suggest that, despite the narrower 1998–2012 range here, pitch and other acoustic qualities are robustly correlated with judicial decision-making across a wider historical sweep.

Overall, our findings highlight how voice-based perceptions, though often overlooked, can exert a measurable influence on Supreme Court case outcomes. By embedding this inquiry within established judicial-politics research (e.g., McGuire, 1993; 1995; Segal & Spaeth, 2002; Johnson et al., 2006) and enhancing our empirical approach through random-forest controls (Katz et al., 2017) and ideological proxies (Martin & Quinn, 2002), we illustrate that even in a domain dominated by legal doctrine and precedent, extralegal cues—such as pitch or perceived vocal masculinity—may subtly “tip the scales,” especially in close or uncertain cases. We emphasize that these voice characteristics represent a mutable dimension of professional identity that can shape success at the highest judicial level. Future work might explore whether attorneys consciously adjust their vocal presentation in realtime, how these influences interact with visual or emotional signals, or whether additional covering pressures (Yoshino, 2006) arise for advocates who deviate from prevailing norms. We hope these insights spur broader conversation on the interplay between social perceptions and legal outcomes in courts, encouraging further interdisciplinary study of how small but significant heuristics operate within even the most formal adjudicative environments.

8. Conclusion

In light of current policy debates, our study underscores the significant yet often understudied impact of mutable

characteristics, aligning with recent economic analysis (Bertrand & Duflo, 2017). Our study reveals that snap judgments based on the perceived masculinity of advocate voices are predictive of outcomes in the U.S. Supreme Court (SCOTUS). Notably, male lawyers perceived as less masculine have a higher probability of winning their cases—by up to 7 percentage points—a result that challenges the idea that more masculine-sounding males necessarily enjoy an advantage, especially in high-stakes litigation (Crawford & Malik, 2016).

The response of Justices to voice characteristics suggests that the association we observe is not merely a proxy for case weakness. The preferences of Democratic and Republican Justices for less masculine-sounding males and more feminine-sounding females, respectively, point to underlying ideological biases. Moreover, these patterns are especially pronounced among private firms and in industries with predominantly masculine voices, suggesting that both justices' preferences and firm-level selection decisions may shape advocacy outcomes, though we acknowledge that our data do not directly test the selection pipeline.

Our debiasing experiment provides crucial insights, indicating that providing information reduces the correlation between perceived masculinity and perceived win rates by 40%. Additionally, incentives further diminish this correlation by 20% in the presence of information. These findings help distinguish between prejudice and statistical discrimination. Contrary to a learning-over-time hypothesis, we find that lawyers do not typically modulate their voices to sound less masculine following a loss, and experience does not correspond to increased masculinity in voice.

The relevance of our findings extends beyond the legal domain, touching on broader societal issues of affirmative action and gender stereotypes. While affirmative action has traditionally focused on immutable characteristics, our results raise questions about supporting broader gender diversity, beyond simply aiding women who conform to masculine norms, in breaking the glass ceiling. This aligns with the argument that achieving diversity milestones might ease the pressure to continue diversifying (Zweigenhaft, 1987; Zweigenhaft & Domhoff, 1998). Furthermore, our study illuminates the significant role of social and mental stereotypes in shaping behavioral responses to speakers, resonating with theories in philosophy of language (Putnam, 1973) and recent economic decision-making models (Bordalo et al., 2014). In conclusion, our findings highlight the nuanced interplay between voice perception, assimilation, and judicial outcomes, emphasizing the need for deeper exploration of mutable characteristics in policy and economic research.

Acknowledgments

We thank Michael Boutros and Dennis Luo for invaluable research assistance. We thank seminar participants at the Amsterdam Center for Research in Experimental Economics and Political Decision-Making,

Association for Society, Religion, Economics, and Culture Conference, Behavioral Economics and Experimental Research Summer School, the European University Institute, Université de Paris II, Stockholm University, Toulouse School of Economics, University of Chicago, University of Glasgow, University of Michigan, University of Toronto, Canadian Economic Association, Linguistics Society Annual Meeting, and the Tokyo Laboratory Phonology Conference.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was partially supported by the Social Sciences and Humanities Research Council of Canada, the European Research Council (Grant No. 614708), Humanities Division of the University of Chicago, the Connaught Fund at the University of Toronto, the Swiss National Science Foundation, and Agence Nationale de la Recherche.

ORCID iD

Daniel L. Chen  <https://orcid.org/0000-0002-5774-2211>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. The event study of Katz et al. (2015) looks at one stock per case whereas Crawford and Malik (2016) makes an estimate for an entire industry, so these estimates of stakes are not strictly comparable.
2. Further details of the experimental instructions are in the Appendix.
3. Samples can be found at: forward (<https://goo.gl/3gCMTx>) and reversed (<https://goo.gl/zhzzhc>) clip.
4. One concern with an online setting like MTurk for conducting surveys is that subjects can try to do the task as quickly as possible so they can move onto the next paid task, and occasionally vary their responses to make it look like they are carrying out the task properly. If some raters only occasionally change their numbers to give the appearance that they are doing work, normalization puts more weight on those individuals who are providing less signal in their ratings, which would yield less precise associations with outcomes.
5. The results are robust to using linear probability, probit, or logistic regressions.
6. 2.79 divided by 0.54.
7. <https://scdb.wustl.edu/data.php>.
8. Figure 6 (Katz et al. (2017)).
9. The most represented petitioner categories in the highest quartile of masculine voices are: penal institutions,

telecommunications, and drug manufacturers. The most represented petitioner categories in the lowest quartile of masculine industries are: United States, city or municipality, and employee or job applicant.

10. The average of Column 1 and Column 5 of Table A.6 is Column 5 in Table A.2 (that is, the data used is pooled but the model remains the same: no MTurk controls and no lawyer fixed effects).
11. We constructed five dummy indicators for MTurk rater demographics and two of the five are significant at the 5% level. The five dummies are indicators for male (49% of the sample), poor (less than US\$20,000 income per year) (20% of the sample), White (74% of the sample), at-most high school education (11% of the sample), and being from the U.S. South (30% of the sample). We did not construct or test any other dummy indicators or interactions of indicators.

References

- Adler, A., & Schuckers, M. E. (2007). Comparing human and automatic face recognition performance. *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics: A Publication of the IEEE Systems, Man, and Cybernetics Society*, 37(5), 1248–1255. <https://doi.org/10.1109/tsmcb.2007.907036>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111(2), 256–274. <https://doi.org/10.1037//0033-2909.111.2.256>
- Apple, W., Streeter, L. A., & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37(5), 715–727. <https://doi.org/10.1037//0022-3514.37.5.715>
- Atkinson, M. (1984). *Our Masters' Voices: The language and body language of politics*. Methuen & Co. Ltd.
- Austen-Smith, D., & Fryer, R. G., Jr. (2005). "An economic analysis of" acting white. *Quarterly Journal of Economics*, 120(2), 551–583. <https://doi.org/10.1162/0033553053970205>
- Babel, M., McGuire, G., & Cruz, S. (2013). Perceived vocal attractiveness across dialects is similar but not uniform. In *Proceedings of interspeech* (pp. 426–430).
- Babel, M., McGuire, G., & King, J. (2014). Towards a more nuanced view of vocal attractiveness. *PLoS One*, 9(2), Article e88616. <https://doi.org/10.1371/journal.pone.0088616>
- Becker, G. S. (1957). *Economics of discrimination*. University of Chicago Press.
- Benjamin, D. J., & Shapiro, J. M. (2009). Thin-slice forecasts of gubernatorial elections. *The Review of Economics and Statistics*, 91(3), 523–536. <https://doi.org/10.1162/rest.91.3.523>
- Berggren, N., Jordahl, H., & Poutvaara, P. (2010). The looks of a winner: Beauty and electoral success. *Journal of Public Economics*, 94(1), 8–15. <https://doi.org/10.1016/j.jpubeco.2009.11.002>
- Bertrand, M., & Duflo, E. (2017). *Field experiments on discrimination*. *Handbook of economic field experiments*, 1, 309–393. <https://doi.org/10.106/bs.hefe.2016.08.004>
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1), 249–275. <https://doi.org/10.1162/003355304772839588>
- Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *The American Economic Review*, 91(2), 67–72. <https://doi.org/10.1257/aer.91.2.67>
- Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *The American Economic Review*, 94(4), 991–1013. <https://doi.org/10.1257/0002828042002561>
- Black, R. C., Roberts, S. A., Johnson, T. R., & Goldman, J. (2011). Emotions, oral arguments, and supreme court decision making. *The Journal of Politics*, 73(2), 572–581.
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2014). *Stereotypes*. Technical report. National Bureau of Economic Research.
- Buller, D. B., Burgoon, J. K., Woodall, W. G., & communications, N. (1996). *The unspoken dialogue*. McGraw-Hill.
- Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity*. Routledge.
- Cameron, C. M. (1993). New avenues for modeling judicial politics. In *Conference on the political economy of public law*. W. Allen Wallis Institute of Political Economy, University of Rochester.
- Campbell-Kibler, K. (2010). Sociolinguistics and perception. *Language and Linguistics Compass*, 4(6), 377–389. <https://doi.org/10.1111/j.1749-818X.2010.00201.x>
- Case, M. A. C. (1995). Disaggregating gender from sex and sexual orientation: The effeminate man in the law and feminist jurisprudence. *The Yale Law Journal*, 105(1), 1. <https://doi.org/10.2307/797140>
- Charles, K. K., & Guryan, J. (2008). Prejudice and Wages: An empirical assessment of Becker's the economics of discrimination. *Journal of Political Economy*, 116(5), 773–809. <https://doi.org/10.1086/593073>
- Chen, D., Halberstam, Y., & Yu, A. C. (2016). Perceived masculinity predicts US supreme court outcomes. *PLoS One*, 11(10), Article e0164324. <https://doi.org/10.1371/journal.pone.0164324>
- Chen, D. L. (2018). Is justice really blind? And is it also deaf?.
- Chen, D. L., & Jess, E. (2017). "Can machine learning help predict the outcomes of asylum adjudications?" Proceedings of the 16th edition of the International Conference on Artificial Intelligence and Law.
- Crawford, J., & Malik, N. (2016). Billions of dollars at stake in supreme court power market fight. *Bloomberg News*.
- Danescu-Niculescu-Mizil, C., Lee, L., Pang, B., & Kleinberg, J. (2012). Echoes of power: Language effects and power differences in social interaction. In Proceedings of the 21st international conference on world wide web, New York, NY, USA, 2012, pp. 699–708.
- Dietrich, B. J., Enos, R. D., & Sen, M. (2019). Emotional arousal predicts voting on the US supreme court. *Political Analysis*, 27(2), 237–243. <https://doi.org/10.1017/pan.2018.47>
- Eckert, P. (2008). Variation and the indexical field. *Journal of SocioLinguistics*, 12(4), 453–476. <https://doi.org/10.1111/j.1467-9841.2008.00374.x>

- Eckert, P., & McConnell-Ginet, S. (2003). *Language and gender*. Cambridge University Press.
- Epstein, L., & Knight, J. (1997). *The choices justices make*. Sage.
- Epstein, L., Landes, W. M., & Posner, R. A. (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. <https://doi.org/10.1086/651511>
- Epstein, L., Landes, W. M., & Posner, R. A. (2013). *The behavior of federal judges: A theoretical and empirical study of rational choice*. Harvard University Press.
- Epstein, L., Parker, C. M., & Segal, J. A. (2018). Do justices defend the speech they hate? An analysis of in-group bias on the US supreme court. *Journal of Law and Courts*, 6(2), 237–262. <https://doi.org/10.1086/697118>
- Fisher, J. L. (2013). A clinic's place in the Supreme Court bar. *Stanford Law Review*, 65(137), 137.
- Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Prentice-Hall.
- Gregory, S. W., Jr., & Gallagher, T. J. (2002). Spectral analysis of candidates' nonverbal vocal communication: Predicting US presidential election outcomes. *Social Psychology Quarterly*, 65(3), 298–308. <https://doi.org/10.2307/3090125>
- Guimerà, R., & Sales-Pardo, M. (2011). Justice blocks and predictability of us supreme court votes. *PLoS One*, 6(11), Article e27188. <https://doi.org/10.1371/journal.pone.0027188>
- Guthrie, C., Rachlinski, J. J., & Wistrich, A. J. (2007). Blinking on the bench: How judges decide cases. *Cornell Law Review*, 93(1), 1–44.
- Hazelton, M. L. W., & Hinkle, R. K. (2022). *Persuading the supreme court: The significance of briefs in judicial decision-making*. University Press of Kansas.
- Heckman, J. J., & Siegelman, P. (1993). *The urban Institute audit studies: Their methods and findings clear and convincing evidence: Measurement of discrimination in America*. Urban Institute Press.
- Hodges-Simeon, C. R., Gaulin, S. J. C., & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, 21(4), 406–427. <https://doi.org/10.1007/s12110-010-9101-5>
- Johnson, T. R., Spriggs, J. F., & Wahlbeck, P. J. (2006). The influence of oral arguments on the US Supreme Court. *American Political Science Review*, 100(1), 99–113. <https://doi.org/10.1017/s0003055406062034>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PLoS One*, 12(4), Article e0174698. <https://doi.org/10.1371/journal.pone.0174698>
- Katz, D. M., Bommarito, M. J., Soellinger, T., & Chen, J. M. (2015). Law on the market? Evaluating the securities market impact of supreme court decisions. *arXiv preprint arXiv:1508.05751*.
- Kessler, S., & McKenna, W. (1978). *Gender: An ethno-methodological approach*. John Wiley and Sons.
- Kiesling, S. (2007). Men, masculinities, and language. *Language and Linguistics Compass*, 1(7), 653–673. <https://doi.org/10.1111/j.1749-818x.2007.00035.x>
- Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings. Biological sciences*, 279(1738), 2698–2704. <https://doi.org/10.1098/rspb.2012.0311>
- Knowles, J., Persico, N., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, 109(1), 203–229. <https://doi.org/10.1086/318603>
- Kornhauser, L. A. (1999). Judicial organization and administration. In W. S. Chris (Ed.), *Encyclopedia of law and economics* (5, pp. 27–44). Chapter 11.
- Kramer, M. (1987). Saying bye-bye to the wimp factor. *US News and World Rep.*
- Lazarus, R. J. (2008). Advocacy matters before and within the supreme court: Transforming the court by transforming the bar. *The Georgetown Law Journal*, 1487.
- List, J. A. (2004). The nature and extent of discrimination in the marketplace: Evidence from the field. *Quarterly Journal of Economics*, 119(1), 49–89. <https://doi.org/10.1162/003355304772839524>
- Martin, A. D., & Quinn, K. M. (2002). Dynamic ideal point estimation via Markov chain Monte Carlo for the US supreme court. *Political Analysis*, 10(2), 134–153. <https://doi.org/10.1093/pan/10.2.134>
- Martin, A. D., Quinn, K. M., Ruger, T. W., & Kim, P. T. (2004). Competing approaches to predicting supreme court decision making. *Perspectives on Politics*, 2(04), 761–767. <https://doi.org/10.1017/s1537592704040502>
- Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, 67(1), 1–43. <https://doi.org/10.1111/j.1540-6261.2011.01705.x>
- McAleer, P., Todorov, A., & Belin, P. (2014). How do you say hello? Personality impressions from brief novel voices. *PLoS One*, 9(3), Article e90779. <https://doi.org/10.1371/journal.pone.0090779>
- McGuire, K. T. (1993). *The Supreme Court bar: Legal elites in the Washington community*. University of Virginia Press.
- McGuire, K. T. (1995). Repeat Players in the Supreme Court: The role of experienced lawyers in litigation success. *The Journal of Politics*, 57(01), 187–196. <https://doi.org/10.2307/2960277>
- Mobius, M. M., & Rosenblat, T. S. (2006). Why beauty matters. *The American Economic Review*, 96(1), 222–235. <https://doi.org/10.1257/000282806776157515>
- Morawetz, N. (2011). Counterbalancing distorted incentives in supreme court Pro Bono practice: Recommendations for the new supreme court Pro Bono bar and public interest practice communities. *NYU Law Review*, 86, 131.
- Nallon, S. (2014). The thatcher years - voice analysis of Margaret thatcher for US TV.
- Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal*

- of *Experimental Psychology: Applied*, 7(3), 171–181. <https://doi.org/10.1037/1076-898X.7.3.171>
- Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56(3), 799–866. <https://doi.org/10.1257/jel.20161309>
- O'Barr, W., & Atkins, B. (1980). "Women's Language" or "powerless language". In S. M. Ginet, R. Borker, & N. Furman (eds), *Women and languages in literature and society* (pp. 93–110). Praeger.
- O'Barr, W. M. (1982). *Linguistic evidence: Language, power, and strategy in the courtroom studies on law and social control*. Academic Press.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2), 187–204. <https://doi.org/10.1080/07350015.2016.1227711>
- Pemberton, C., McCormack, P., & Russell, A. (1998). Have women's voices lowered across time? A cross sectional study of Australian women's voices. *Journal of Voice*, 12(2), 208–213. [https://doi.org/10.1016/s0892-1997\(98\)80040-4](https://doi.org/10.1016/s0892-1997(98)80040-4)
- Phelps, E. S. (1972). The statistical theory of racism and sexism. *The American Economic Review*, 62(4), 659–661.
- Posner, R. A. (1973). An economic approach to legal procedure and judicial administration. *The Journal of Legal Studies*, 2(2), 399–458. <https://doi.org/10.1086/467503>
- Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, 18(10), 10–30. <https://doi.org/10.1177/0261927x99018001002>
- Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy*, 70(19), 699–711. <https://doi.org/10.2307/2025079>
- Puts, D. A., Barndt, J. L., Welling, L. L. M., Dawood, K., & Burriss, R. P. (2011). Intrasexual competition among women: Vocal femininity affects perceptions of attractiveness and flirtatiousness. *Personality and Individual Differences*, 50(1), 111–115. <https://doi.org/10.1016/j.paid.2010.09.011>
- Rachlinski, J. J., Lynn Johnson, S., Wistrich, A. J., & Guthrie, C. (2009a). Does unconscious bias affect trial judges? *The Notre Dame Law Review*, 84(3), 1195–1246.
- Rachlinski, J. J., Lynn Johnson, S., Wistrich, A. J., & Guthrie, C. (2009b). Does unconscious racial bias affect trial judges? *The Notre Dame Law Review*, 84(3), 1195–1246.
- Rao, G. (2014). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools." *American Economic Review*, 109(3): 774–809.
- Roberts, J. G. (2005). Oral advocacy and the Re-emergence of a supreme court bar. *Journal of Supreme Court History*, 30(1), 68–81. <https://doi.org/10.1353/sch.2005.0017>
- Ruger, T. W., Kim, P. T., Martin, A. D., & Quinn, K. M. (2004). The Supreme Court forecasting project: Legal and political science approaches to predicting Supreme Court decisionmaking. *Columbia Law Review*, 104(4), 1150–1210. <https://doi.org/10.2307/4099370>
- Scherer, K. R. (1979). Voice and speech correlates of perceived social influence in simulated juries. In K. R. Scherer, H. Giles, & R. St. Clair (eds), *The social psychology of language* (pp. 88–120). Blackwell.
- Schroeder, J., & Epley, N. (2015). The sound of intellect speech reveals a thoughtful mind, increasing a job candidate's appeal. *Psychological Science*, 26(6), 877–891. <https://doi.org/10.1177/0956797615572906>
- Schubert, J. N., Peterson, S. A., Schubert, G., & Wasby, S. (1992). Observing supreme court oral argument: A biosocial approach. *Politics and the Life Sciences*, 11(1), 35–51. <https://doi.org/10.1017/s0730938400017196>
- Schubert, J. N., Peterson, S. A., Schubert, G., & Wasby, S. L. (2002). Dialect, sex and risk effects on judges questioning of counsel in supreme court oral argument. In F. K. Salter (Ed.), *Risky transactions: Trust, kinship, and ethnicity* (p. 304). Bergham Books. chapter 10.
- Segal, J. A., & Spaeth, H. J. (2002). *The supreme court and the attitudinal model revisited*. Cambridge University Press.
- Spriggs, II., James, F., & Wahlbeck, P. J. (1995). Calling it quits: Strategic retirement on the federal courts of appeals, 1893–1991. *Political Research Quarterly*, 48(3), 573–597. <https://doi.org/10.1177/106591299504800306>
- Starecheski, L. (2014). *Can changing how you sound help you find your voice?* National Public Radio.
- Sunstein, C. R., Schkade, D., Ellman, L. M., & Sawicki, A. (2006). *Are judges political? An empirical analysis of the federal judiciary*. Brookings Institution Press.
- Szmer, J. J., Sarver, T. A., & Kaheny, E. B. (2010). Have we come a long way, baby? The influence of attorney gender on supreme court decision making. *Politics and Gender*, 6(01), 1–36. <https://doi.org/10.1017/s1743923x09990493>
- Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, 33(3), 210–216. <https://doi.org/10.1016/j.evolhumbehav.2011.09.004>
- Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, 308(5728), 1623–1626. <https://doi.org/10.1126/science.1110589>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Waterbury, N. W. (2024). Justice Isn't blind: Attorney attractiveness and success in US federal court. *Journal of Law and Courts*, 12(2), 1–22. <https://doi.org/10.1017/jlc.2024.2>
- White, D., Jonathon Phillips, P., Hahn, C. A., Hill, M., & O'Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proceedings. Biological Sciences*, 282(1814), 20151292. <https://doi.org/10.1098/rspb.2015.1292>
- Yoshino, K. (2000). The epistemic contract of bisexual erasure. *Stanford Law Review*, 52(2), 353–461. <https://doi.org/10.2307/1229482>
- Yoshino, K. (2006). *The hidden assault on our civil rights*. Random House Incorporated.
- Zweigenhaft, R. L. (1987). Minorities and women of the corporation: Will they attain seats of power. In: *Power elites and organizations*, 37, (pp. 62).
- Zweigenhaft, R. L., & Domhoff, G. W. (1998). *Diversity in the power elite: Have women and minorities reached the top?* Yale University Press.