

COVERING: MUTABLE CHARACTERISTICS AND PERCEPTIONS OF VOICE IN THE U.S. SUPREME COURT

DANIEL L. CHEN, YOSH HALBERSTAM, AND ALAN YU*

Abstract The emphasis on “fit” as a hiring criterion has raised the spectrum of a new form of subtle discrimination (Bertrand and Duflo 2016). Under complete markets, correlations between malleable characteristics and outcomes should not persist (Becker 1957). Yet using data on 1,901 U.S. Supreme Court oral arguments between 1998 and 2012, we document that voice-based snap judgments based on lawyers’ identical introductory sentences, “Mr. Chief Justice, (and) may it please the Court?”, predict court outcomes. The connection between vocal characteristics and court outcomes is specific only to perceptions of masculinity and not other characteristics, even when judgment is based on less than three seconds of exposure to a lawyer’s speech sample. Consistent with employers mistakenly favoring lawyers with masculine voices, perceived masculinity is negatively correlated with winning and the negative correlation is larger in more masculine-sounding industries. The first lawyer to speak is the main driver. Among these petitioners, males below median in masculinity are 7 percentage points more likely to win in the Supreme Court. Female lawyers are also coached to be more masculine and women’s perceived femininity predict court outcomes. Republicans, more than Democrats, vote for more feminine-sounding females, while Democrats, but not Republicans, vote for less masculine-sounding men. A de-biasing experiment finds that information reduces 40% of the correlation between perceived masculinity and perceived win, and incentives reduce another 20%. A model shows how the information treatment identifies statistical discrimination and the incentives treatment identifies taste. Perceived masculinity explains additional variance relative to and is orthogonal to the best random forest prediction model of Supreme Court votes.

Keywords: Gender, identity, speech, judicial decision making, **JEL codes:** J15, J78, K41

*Daniel L. Chen, daniel.chen@iast.fr, Toulouse School of Economics, Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France; dchen@law.harvard.edu, LWP, Harvard Law School; Yosh Halberstam, yosh.halberstam@utoronto.ca, University of Toronto, Economics Department; Alan Yu, alyu@uchicago.edu, University of Chicago, Linguistics Department. First draft: October 2014. Current draft: August 2016. The most recent version of this article is available at: <http://users.nber.org/~dlchen/papers/Covering.pdf>. This work was partially supported by the Social Sciences and Humanities Research Council of Canada, the European Research Council, 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. 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.

1 Introduction

Civil rights law is populated with plaintiffs who were punished for being different. 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*, which means to assimilate to dominant norms (Goffman 1963). Mutable characteristics have recently entered economic models of identity formation (Austen-Smith and Fryer Jr 2005). One sentence—but no data—is devoted to malleable characteristics in two summaries of the empirical literature (Bertrand and Duflo 2016; Neumark 2016): “Recently, the emphasis on ‘fit’ [of] a prospective employee ... as a hiring criterion has raised the spectrum of a new form of subtle discrimination.” (Whether this is viewed as a bad thing, of course, depends on one’s attitude towards fit.) For example, African-Americans cannot be fired for their skin color, but they could be for wearing cornrows. Potential jurors cannot be struck for their ethnicity, but can be for speaking a foreign language. Women cannot be fired for their gender, but can be penalized for becoming mothers. Jews and Muslims cannot be discharged from the US military or suspended from French public schools for their religion, but they can be for covering their head. Female Indonesian police applicants cannot be rejected for being female, but can be for failing virginity tests (Chen 2004).

Throughout history, when dominant groups ask subordinate groups to cover, or when courts make employment contingent on covering, it is perceived to legitimize second-class citizenship for the subordinated group (Balkin 2011). Should we understand, as constitutional scholars argue, such legally sanctioned differential treatment to be harmful—are individuals “punished” for not conforming when it comes to mutable characteristics (Yoshino 2000)? This distinction between *being* and *doing* incentivizes assimilation, a topic that is relevant in a wide variety of contexts: EU refugee policy’s focus on integration, the tragic outcomes of non-assimilation when it pertains to ethnic-religious conflict, and contemporary discussions of trauma, micro-aggression, and safe spaces, for example in the topic of sexual harassment, where it was once perceived that women needed to assimilate to the dominant machismo culture of workplace environments (Chen and Sethi 2011).

Economists have recognized the importance of assimilation in identity formation (Fryer and Torelli 2010) and documented its psychological costs using proxies in the National Longitudinal Survey of Youth (Mueller-Smith 2013). When it comes to how one speaks, minorities’ choice of diction have been found to be associated with long-run labor market outcomes (Grogger 2011). Female lawyers routinely pay coaches to sound more masculine (Starecheski 2014), while men with non-masculine voices are disadvantaged in the labor market (Case 1995). Vocal similarity is important for maintaining social relationships in both animals and humans (Toft and Wright 2015; Giles and Powesland 1997). But is it homophily—speaking in the particular

manner of certain identity groups—that engenders trust? It is hard to know the mechanism behind differential treatment of voice, much less whether there is even differential treatment, as the content of speech may change.

The roots of discrimination are perceived to originate in communication difficulties across groups (Lang 1986; Lazear 1999), which can lead to mistrust (Alesina and La Ferrara 2005). Ethnolinguistic fractionalization has been linked to lower economic growth and higher inequality (Mauro 1995). During the process of fractionalization, individuals arguably begin to speak differently. Does this mitosis of sounds lead to mistrust of individuals who are audibly different? We generally distinguish between economic or statistical discrimination (when people are treated differently because they are believed to be less productive (Phelps 1972; Arrow 1973)) and prejudice (when people are treated differently because of animus—a distaste despite its economic consequences (Becker 1957)). While statistical discrimination focuses on information as the source of discrimination, prejudice focuses on preferences. Modern interpretations blur the boundary between information and preferences. Endogenous information acquisition can be due to preferences (Brewer 1988), and what appear to be preferences can be due to Type I (automatic, cognitive, unconscious) instead of Type II (reflective, motivational, conscious) thinking (Uhlmann and Cohen 2005). Complete markets can eliminate misbeliefs, so correlations persist only if there exists animus from the marginal employer (Becker 1957). The prior literature has focused on immutable characteristics, where inequality is determined by the tastes of the potentially smaller set of employers who are on the margin of hiring one individual or another. With mutable characteristics, more employers can be marginal; entire industries were deemed to be systematically more harassing when it came to hostility to female employment (Hersch 2011).

A classical interpretation of Becker (1957) would predict under complete markets that prejudice would also not survive as prejudiced employers in a competitive market will be forced to shut down in the long run because they sacrifice profits by discriminating. However, a large number of correspondence studies find evidence for discrimination (e.g., Bertrand and Mullainathan 2004), and labor market inequality is associated with prejudicial attitudes (Charles and Guryan 2008). Yet it is hard to know the drivers of inequality: separating prejudice from statistical discrimination is challenging, since we typically lack data on productivity (Heckman and Siegelman 1993; Neumark 2016). Innovative experiments with sports card collectors (List 2004), maze solvers (Mobius and Rosenblat 2006), and Indian schoolchildren (Rao 2014) (to name a few) disentangle these two types of discrimination, but it is hard to do so in a real-world, high-skill, high-stakes setting. We analyze oral arguments before the Supreme Court of the United States (SCOTUS), in which winning is the productivity measure, using snap judgments of voice samples (e.g., [Sample 1](https://goo.gl/ZPdCkU) (<https://goo.gl/ZPdCkU>) and [Sample 2](https://goo.gl/mbhuLF) (<https://goo.gl/mbhuLF>)) from 1,901 U.S. Supreme Court petitioner and respondent oral arguments, in which lawyers use identical introductory sentences: “Mr. Chief Justice, (and)

may it please the Court?” The oral arguments are arguably high stakes. According to Justices, “[y]our oral argument [is the] the most effective weapon you have got” (Chief Justice Roberts 2005) and “[it is like] building a cathedral” (Justice Jackson 1951). In his personal diaries, Justice Blackmun graded the oral arguments of each attorney and the grades are correlated with outcomes (Johnson et al. 2006; McAtee and McGuire 2007). A recent study finds that Supreme Court cases from 1999–2013 collectively moved \$140 billion in stock prices (Katz et al. 2015).

Deep-voiced individuals are perceived to have greater leadership capacity and perceived to be more competent, persuasive, confident and trustworthy (Klofstad et al. 2012; Tigue et al. 2012; Apple et al. 1979; Buller et al. 1996). Based on these perceptions, we present a simple model where firms favor masculine voices because they believe a lawyer with a masculine voice is more likely to win (statistical discrimination) or because they prefer lawyers with masculine voices regardless of the economic consequences (prejudice). We use a de-biasing experiment to distinguish between prejudice and statistical discrimination.

We find that first impressions of lawyers’ voices predict Supreme Court outcomes, and this finding is specific to perceived masculinity: males are more likely to win when their voices are perceived to be less masculine (Figure 1). The magnitude of the association remains stable with controls for i) lawyer’s personal characteristics, ii) case history, iii) acoustics features, iv) demographics of the raters, and v) other perceptual ratings of the voices, e.g., perceived aggressiveness, confidence, competence, attractiveness, intelligence, trustworthiness, attitude, respect, and empathy. No other perception is consistently predictive of Supreme Court outcomes. Furthermore, in a mechanism experiment using reversed clips, perceived masculinity, but not other perceptions, are strongly correlated across forward and reverse clips. Masculinity perhaps may be an attribute of voice that is more salient, innate, or stereotyped than other attributes.

When we analyze the petitioners and respondents separately, the petitioner—who is the party who petitioned the Supreme Court to review the case—and speaks first—while the respondent speaks in the second-half of the oral argument—is the main driver of the results, which is consistent with the role of anchoring (a cognitive bias that describes the common human tendency to rely too heavily on the first piece of information offered) (Kahneman 2011). Among these petitioners, being above or below median in masculine rating would be equivalent to 7 percentage points in the likelihood of winning. This result along with the magnitude of the stock market response to Supreme Court outcomes (Katz et al. 2015) suggests that it is unlikely for salaries to offset this difference. We lack data on salaries paid to Supreme Court lawyers with more masculine voices. In another setting, salaries are higher for more attractive lawyers (Hamermesh and Biddle 1994; Biddle and Hamermesh 1998). If the same is true for Supreme Court lawyers, and perceived masculinity of voice and physical attractiveness are correlated, then the salary differential goes in the opposite direction of productivity, measured

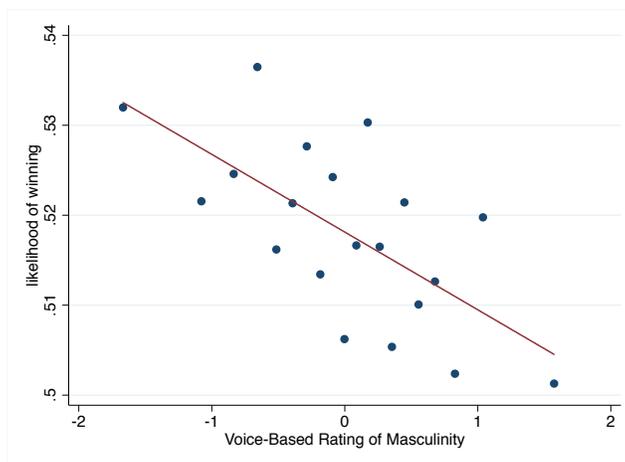


Figure 1: Voice-Based Masculinity Rating and Court Outcomes ($p < 0.01$)

Notes: Binned scatterplots illustrating the association between voice-based masculinity rating and court outcomes. Binned scatterplots are a non-parametric method of plotting the conditional expectation function (which describes the average y-value for each x-value).

as actual win rates.

Masculine voice is mutable to some extent. Both within- and between-lawyer variation drive the association between perceived masculinity and case outcomes. Variation within the same lawyers' voices is found to account for roughly one-third of the total correlation while variation between lawyers drives the remaining two-thirds. To be sure, lawyers may consciously or subconsciously perceive their cases to be weaker or stronger, which could lead their voices to change, so we will be including controls for pre-trial characteristics and the best existing prediction model of advocate wins.

Female lawyers are also coached to have their voices sound more masculine (Starecheski 2014; Kramer 1987; Orvice 1993; Atkinson 1984; Nallon 2014), and we find that the negative correlation between court success and perceived masculinity extends to women. Less-masculine sounding men and women (i.e., more-feminine sounding women) are more likely to win, and this association is specific to perceived femininity for females and perceived masculinity for males. Being above or below the median in perceived femininity corresponds to 5 percentage points in likelihood of winning for females. If the male and female results are pooled together, assuming that masculinity can be treated as the polar opposite of femininity (a controversial assumption to be sure), the pooled results would be stronger. These results are consistent with a model where less-masculine sounding lawyers need to be substantially better than their more-masculine counterparts in order to be hired or retained by discriminatory firms.

Next, we document substantial heterogeneity in how judges' votes correlate with perceived masculinity. Democrats, but not Republicans, vote for less-masculine sounding men (Figure 2), while Republicans, more than Democrats, vote for more feminine-sounding females. (In the higher intercept for Republicans, we see that Republicans favor males, as documented previously by Szmer et al. (2010)). We verify that this heterogeneity is due to ideology rather

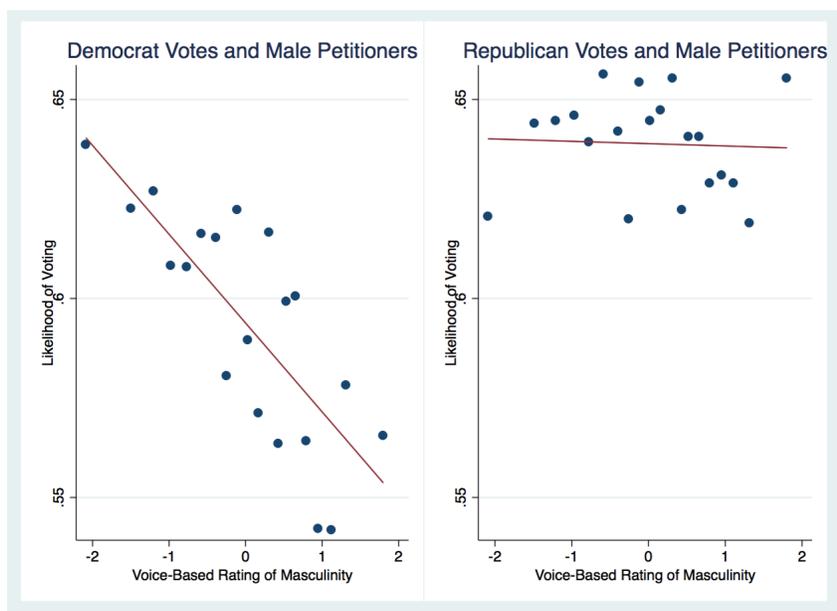


Figure 2: Political Party and Response to Masculinity ($p < 0.01$)

than gender of the judge, though the number of judges is limited. A full exploration of the reason for the ideological heterogeneity is beyond the scope of this paper. For instance, liberals may wish to signal a preference for gender-neutral norms—and conservatives for traditional gender norms—or gender-neutral sounding voices may signal liberal ideology and engender trust (and vice versa for conservative ideology). Data on different types of cases suggests that at least some of the time, Democrats and Republicans vote together for less-masculine sounding males and more-feminine sounding females in technical areas such as taxation, which rarely divide 5-4, where dividing 5-4 is a proxy for whether an area of law is overtly ideological.

Under complete markets, lawyers would have incentives to make their voice sound less masculine. However, lawyers might not adjust if firms have the misbelief that masculine-sounding lawyers are better or have a preference for masculine-sounding lawyers. Preference for masculine-sounding lawyers may vary across industry. For example, prior research finds that, in the private sector, CEOs with deeper voices manage larger companies (Mayew and Venkatachalam 2012). In our setting, we observe stronger negative correlations between perceived masculinity and court outcomes among private firms and stronger negative correlations in industries with more masculine voices ($p < 0.05$) (Figure 20). There is essentially no correlation between perceived masculinity and win rates for public lawyers advocating on behalf of the government (Figure 3), which would seem to render less likely an explanation based on ideological profiling. (Throughout the paper, the terms “firm” and “industry” are used to refer to the category of the petitioner or respondent. 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

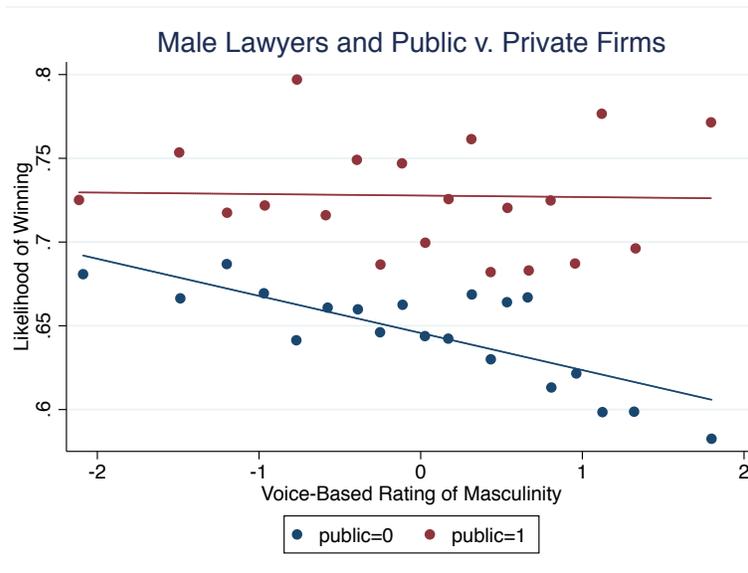


Figure 3: Public vs. Private Firms and Response to Masculinity

industries are: United States, city or municipality, and employee or job applicant.)

These differences are consistent with a model where greater economic costs come with having a preference for masculine-sounding lawyers and some industries may have stronger preferences for masculine characteristics. These industries and firms would give up more in lost court wins in order to employ masculine-sounding oral advocates.

The Supreme Court Bar is a heavily vetted and elite group of oral advocates, and firms typically exercise great care in choosing who will argue their case. We find that the more lawyers appear at the Supreme Court, the more masculine they sound regardless of whether they won or lost the previous case. This pattern is inconsistent with lawyers learning to sound less masculine, but may be consistent with Austen-Smith and Fryer's (2005) ethnic identity model, adapted by Grogger (2011) to study the linguistic choices of minorities. An individual has a "two-audience" signaling problem, and in our application, the second audience is potentially the client. We lack information about the selectors, but it is reasonable to think the top of elite law firms with Supreme Court practices would be predominantly White and non-Southern, and if so, their stereotypes would be more predictive of outcomes. Consistent with this, we observe that White raters' perceptions of masculinity were more predictive of court outcomes ($p < 0.05$) and non-Southerners' perceptions were more predictive of court outcomes ($p < 0.05$). This difference suggests the perception of masculinity can differ based on the stereotypes associated with the sound of a speaker's words. Moreover, poor and non-White individuals were less likely to correlate masculinity with winning ($p < 0.05$ for both), which would be consistent with the behavioral response to stereotypes differing across individuals if perceptions of winning precipitate choices.

The behavioral response to stereotypes may be due to tastes (prejudice) or information

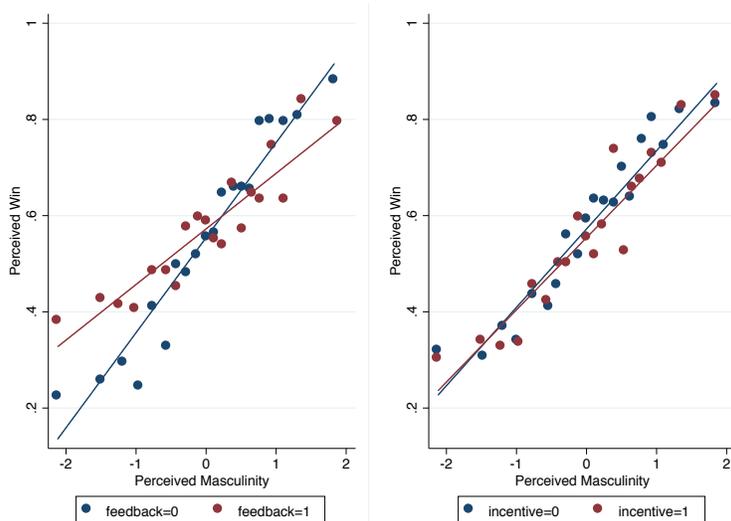


Figure 4: Feedback ($p < 0.01$), Incentives

(statistical discrimination). We execute a de-biasing experiment that provides information and/or incentives in a 2x2 design to distinguish the two channels. We find that information reduces 40% of the correlation between perceived masculinity and perceived win (Figure 4), identifying the presence of statistical discrimination. Incentives further reduce another 20% of the correlation when there is information, identifying the presence of prejudice. Incentives, which erode the effect of taste, would have no effect if there is no differential taste (prejudice).

Finally, we further assess the robustness of the association between perceived masculinity and court outcomes. To be sure, we cannot make strong causal inferences, and it is infeasible to conduct an experiment in the Supreme Court that modifies the sound of a voice after a lawyer speaks but before the Justice hears the words. However, we can assess the degree to which omitted variables drive our results by comparing with the previous best prediction models of Supreme Court votes (Katz et al. 2014). Random forest trees—a weighted non-parametric model that forms weighted predictions based on nearest neighbors—can predict how the judges will vote with 70% accuracy. Two results emerge from this comparison. First, perceived masculinity is essentially orthogonal to the best prediction model. This indicates that perceived masculinity is uncorrelated with pre-trial case characteristics and judge and court trends used in the random forest model, even when these characteristics are combined in a non-linear manner. As to why perceived masculinity is orthogonal to the random forest model, the prediction model performs well in 9-0 and 0-9 cases but performs poorly in close cases, i.e., when decision makers are more indifferent. In contrast, perceived masculinity’s role emerges at the case level rather than the vote level. This result suggests that perceived masculinity is influential in close cases and for the swing voter, and is consistent with heuristics or biases being more likely to influence decisions in hard or close cases. Second, in trying to predict the Democratic votes for male petitioners, the previous best predictive model explains roughly

6% of variance—a small amount, because it is a binary outcome—and including perceived masculinity renders 6.3-6.4% of variance explained, which is 0.2-0.6 percentage points or 3-10% more than the baseline.

The remainder of this paper proceeds as follows. Section 2 provides background and theory. Section 3 describes the experimental methods and data. Section 4 presents the baseline association between first impressions of lawyers’ voices and U.S. Supreme Court outcomes. It also demonstrates the primacy of the first speaker, assesses the role of within- vs. between-lawyer variation in accounting for the association, and explores whether the results extend to female lawyers. Section 5 investigates why first impressions of voice predict SCOTUS outcomes considering lawyer characteristics, pre-trial case characteristics, judicial heterogeneity, firm heterogeneity, and rater demographics. Section 6 documents that lawyers sound more masculine the more they appear before the SCOTUS. It also reports the results of a de-biasing experiment. Section 7 decomposes the explanation of variance using the best predictive models of Supreme Court votes and data on acoustic features. Section 8 concludes.

2 Background and Theory

In theory of language, meanings of words are found in the context in which concepts are encountered. Semantic competence, the ability to determine meaning, depends on one’s stereotypes associated with a speaker’s words (Putnam 1973).

A large body of work examines how people speak – their vowels, pitch, diction, and intonation – but there is relatively little evidence that speech variation beyond lexical choices (fluctuations in the way one speaks holding the words fixed) matters in real-world behavior, and this is where our paper comes in. Are vocal cues relevant in high-stakes policy-making settings such as the U.S. Supreme Court even with advocates using identical sentences?

First impressions of voice are associated with the subsequent behavior of the perceiver in mate selection, leader selection, housing options, consumer choices, and even stock market outcomes in a recent vocal analysis of earnings conference calls (Nass and Lee 2001; Klofstad et al. 2012; Purnell et al. 1999; Scherer 1979; Tigue et al. 2012; Mayew and Venkatachalam 2012). These findings are consistent with a broader behavioral literature on Type I vs. Type II thinking – perhaps most famously, finding that facial cues predict electoral outcomes (Todorov et al. 2005; Benjamin et al. 2009; Berggren et al. 2010) and that nonverbal behavior predicts teacher evaluations in an experiment eliciting student ratings of silent clips of professors at the beginning of the semester (Ambady and Rosenthal 1992) (to name a few). A recent study finds that even when visual cues are present, potential employers rely more on voice-based impressions of a job applicant than on visual cues (Schroeder and Epley 2015). The reliance on

voice may have an evolutionary basis: infants use vocal cues as the first marker of in-group vs. out-group membership (Kahneman 2011; Kinzler et al. 2007). The gendered differentiation of masculine and feminine language has been argued to have a different evolutionary basis—males would be selected for greater aggressivity and higher variance (Locke 2011), which in this case would mean a higher probability of losing.

There is a variety of reasons why voice impressions should not matter. From a rational perspective, information should override first impressions (Posner 1973). From an ideological perspective, court outcomes are largely political and predetermined outcomes (Cameron 1993). From a legal perspective, legal decisions should be based on the legal content of the argument (Kornhauser 1999). From an economic perspective, correlations between voice and outcome should not persist as law firms and advocates adjust their behavior to eliminate such correlations (Becker 1957; Knowles et al. 2001), whereas facial features may be less mutable than voice and teaching evaluations lower stakes than SCOTUS outcomes. But from a behavioral perspective, the way one speaks reveals a lot about one’s personality and identity (Babel et al. 2013; Hodges-Simeon et al. 2010; McAleer et al. 2014; Bordalo et al. 2014), and previous studies have found evidence suggesting that vocal cues can matter in settings as diverse as the court room (Schubert et al. 2002) and U.S. Presidential debates (Gregory Jr and Gallagher 2002).

By and large, judges believe there is no such thing as bias (Edwards and Livermore 2008). As Supreme Court Justice Samuel Alito articulated, “[i]t’s not like we’re judging a moot court: which lawyer is better? ... It’s the case, not the lawyer” (Biskupic et al. 2014). However, the degree to which a lawyer’s words converge to a SCOTUS judge predicts the judge’s vote (Danescu-Niculescu-Mizil et al. 2012). Our study of the Supreme Court differs from previous economic studies of ideological bias: while market pressures may drive identity effects in economic behavior towards the rational model, behavioral anomalies in judicial decision-making can have more permanent consequences (Gennaioli and Shleifer 2007; Baker and Mezzetti 2012). Our study also differs from other studies of courtroom bias in that we focus on three actors: the judge, the lawyer, and the firm.

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 such rainmakers. 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 time when oral advocates are selected by firms is when perceived masculinity may *positively* affect selection.

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; Orvick 1993). Via humming exercises, Thatcher made her voice more masculine by the amount equivalent to [half](https://goo.gl/8bMkut) (https://goo.gl/8bMkut) the male-female difference (Atkinson 1984), though her natural voice occasionally [slipped out](https://goo.gl/WNDgr0) (https://goo.gl/WNDgr0) (Nallon 2014). As women have entered the work force and positions of authority, their voices have moved closer to a masculine standard (Pemberton et al. 1998) and have been rewarded for doing so (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:

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

where $\pi_v = Prob(v \text{ wins})$ 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. This taste can be based on a sense of shared or opposing identity or it could be ideological (Chen and Schonger 2013; 2015; Chen 2010; 2013). 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 stakes in Supreme Court power market fight” (Crawford and Malik 2016). To be sure, 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.

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 :

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

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 (informa-

tion, $\pi_F < \pi_M$) 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 ($\pi_F - \pi_M > \frac{d}{\alpha}$). 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,

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

Analogous to a recent paper that structurally estimates the taste parameter (Rao 2014), if we assume that only beliefs, taste, and stakes enter utility, then as $\alpha \rightarrow \infty$, only 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 de-biasing 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 reduces 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 $\frac{d}{\alpha}$ would be 0 regardless of the size of α .

3 Experiment

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 (<http://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 not solicited by any of the parties 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 U.S. to evaluate 1,901 audio clips comprising 1,085 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.

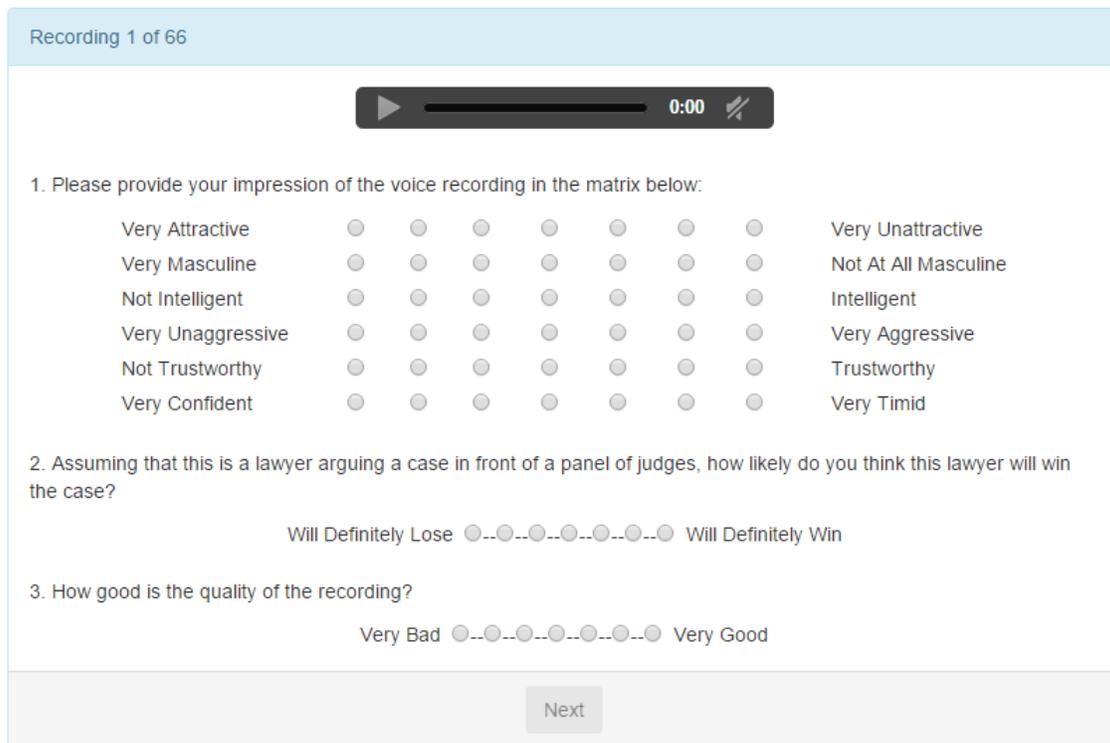
A little over half (382) of the 748 MTurk raters were female. Two-thirds were aged between 20 and 35 years old, and one-third were older than 35. Likewise, one-third indicated they had some college education, whereas one-third claimed to have a bachelor’s degree. The median income of those who completed the survey was about US\$40,000. Their racial and geographical distribution broadly reflects that of the U.S. population. In a separate paper documenting only the correlation and not the mechanisms presented in this paper, Chen et al. (2016e) shows that correlation between the share of participants from a given state and the state share of US population is 0.9588.

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

Figure 5: Screenshot of Experiment



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, which if raters failed they would be 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 5. No information regarding the identity of the lawyer or the nature of the case was given to the participants. Other screenshots of the experimental instructions are in the Appendix.

To control for the possibility of within-voice modeling by raters, instead of the basic design (in which the listener was presented with one voice sample and rates the sample on all scales), we also employed a design with only one question, randomly selected for each voice sample. This is a design perturbation based on earlier pilots with only 60 clips and fewer subjects. Each

Figure 6: Correlation in Average Voice Perceptions across Experiments (Many vs. One Attribute)



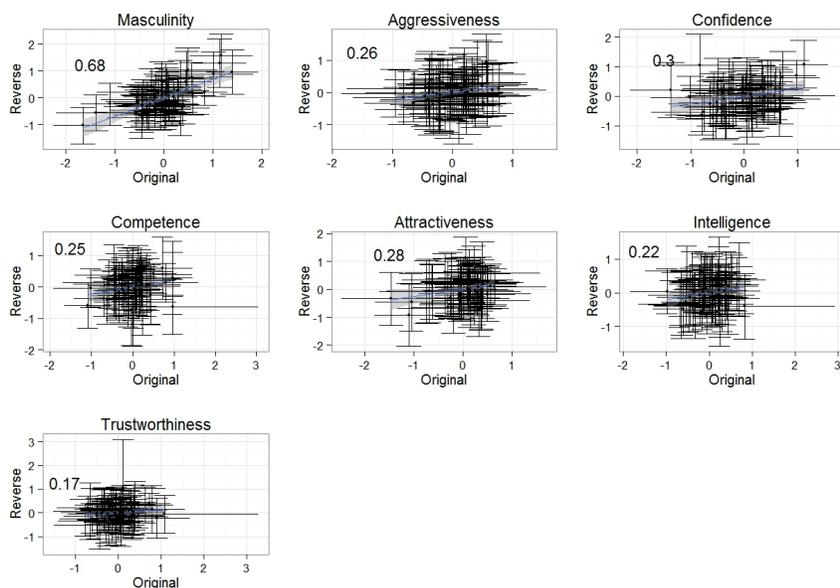
Notes: This figure plots the mean rating of sixty voice samples in the pilot, where the x-axis reflects mean ratings obtained from raters who were asked to rate each advocate on the full set of attributes, whereas the y-axis reflects the mean ratings obtained from raters who were randomly assigned to rate each advocate on only one attribute.

voice clip was played aloud only once, in order to capture the respondents' first impressions – and to avoid them over thinking their responses (Ballew and Todorov 2007). We added and removed a few additional questions, such as perceived age and perceived standard english, and varied the way we asked respondents whether the lawyer *should win* the case versus how likely do they think the lawyer *will win* the case (Question 2 of Figure 5). We find a high degree of correlation of individual perceptions across experimental designs and stimulus presentation methods (see Figure 6).

Figure 6 plots the mean rating for each of the advocates using the two approaches discussed above. The x-axis reflects the the mean ratings obtained from raters who were asked to rate each advocate on the full set of attributes, and the y-axis reflects the mean ratings obtained from raters who were randomly assigned to rate each advocate on only one attribute. The ratings are highly correlated across these experimental designs, suggesting that trait judgments obtained from listening to a voice are quite stable.

We also reversed the voice clips (for a sample: [forward](https://goo.gl/3gCMTx) (<https://goo.gl/3gCMTx>) and [reversed](https://goo.gl/zhzzhc) (<https://goo.gl/zhzzhc>) clip) so the sentence no longer sounded anything like English. The speech features are also different, but some acoustic information is similar. This is a mechanism experiment based on the pilot sample with only 60 clips and 10 ratings per reversed clip. As Figure 7 shows, most attributes exhibit some correlation, but generally very weak ones. The only personality attribute that shows a strong correlation is masculinity. This result sug-

Figure 7: Correlation in Voice Perceptions across Experiments (Forward vs. Reversed Clips)



Notes: This figure plots the rating, where the x-axis reflects ratings obtained from raters who were asked to rate each advocate on the forward version of the clip, whereas the y-axis reflects the ratings obtained from raters who were assigned to rate each advocate on the reversed version of the clip. The number represents the Pearson’s r .

gests that masculinity may be more salient, innate, or stereotyped than other attributes, such as aggressiveness, confidence, competence, attractiveness, intelligence, or trustworthiness.

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., attractive, confidence, etc.) and the perceived likelihood of winning. The normalized rating is given by

$$(1) \quad \widehat{attribute}_{itw} = \frac{attribute_{itw} - \overline{attribute}_w}{\sigma(attribute)_w},$$

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. We provide summary statistics of these measures for male advocates in Table Ia.

One concern with an online setting like MTurk for conducting surveys is that subjects can

TABLE I
SUMMARY STATISTICS

(a) Rating-by-Audio Data (normalized)

Variable	Obs	Mean	SD	Min	Max
Outcome	33666	0.518	0.500	0.000	1.000
Masculine	33666	0.014	0.989	-6.308	4.031
Confident	33666	0.002	0.993	-4.641	4.172
Attractive	33666	-0.005	0.992	-8.001	5.701
Intelligent	33666	-0.006	0.999	-8.001	8.001
Trustworthy	33666	-0.007	0.996	-8.001	8.001
Aggressive	33666	0.002	0.994	-7.261	8.001
Likely winner	33666	0.000	0.995	-5.787	8.001

(b) Audio-Level Data (average of normalized)

Variable	Obs	Mean	SD	Min	Max
Outcome	1634	0.517	0.500	0.000	1.000
Masculine	1634	0.015	0.540	-2.108	1.537
Confident	1634	0.003	0.465	-2.165	1.192
Attractive	1634	-0.006	0.425	-1.513	1.326
Intelligent	1634	-0.006	0.370	-1.739	0.893
Trustworthy	1634	-0.006	0.323	-1.484	1.048
Aggressive	1634	0.004	0.421	-1.733	1.408
Likely winner	1634	0.001	0.441	-2.402	1.216

Notes: This table presents summary statistics of normalized perceptual ratings.

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

Chen et al. (2016e), which focused on normalized ratings and does not explore any of the mechanisms, addresses this issue in a different manner. It begins by examining rater inconsistency using the six clips that were randomly chosen to be played again. Then, when excluding survey participants who score in the bottom quintile of rating consistency, it reports that the results strengthen. It also reports results are robust after excluding outliers, defined by exceeding the Mahalanobis distance threshold. Then, when excluding participants who exceeded this threshold on more than one third of their ratings, the results strengthen. In this paper, we have decided to use the more conservative approach and include these ratings in our analysis since we might prefer to avoid selection of ratings. Even outliers might represent the overall distribution of perceptions. What appear to be outliers may be the legitimate perceptions of a minority demographic. Our goal is to document a relationship between these perceptions and outcomes. The fact that the results using normalized ratings strengthened when restricting to

subjects who displayed greater intra-rater reliability suggests that our reason for preferring the unnormalized measure may be justified.

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

$$(2) \quad \overline{\widehat{attribute}_{it}} = \frac{1}{|W_{it}|} \sum_w \widehat{attribute}_{itw},$$

where W_{it} is the set of raters who rated the voice of advocate i in case t . In Table Ib, we report summary statistics of these average measures. Notice that the data is now transformed such that each observation represents a single measure of voice for each oral argument, whereas observations in the micro-level data consist of multiple ratings. As expected, the variation in trait judgments in the averaged data is reduced significantly. In particular, the standard deviation of the masculine measure is 0.54. This also means that the coefficients in tables need to be divided by 0.54 in order to be interpreted. The use of average scores assumes that all perceptions are the same across individuals and that differences are due to classical measurement error. However, this may not be a reasonable assumption; as Section 5 shows, perception is very subjective.

There are 33,666 individual ratings, with roughly 20 ratings per oral argument. The standard deviation of the raw ratings is about 1.5 (so coefficients need to be divided by 1.5 in order to interpret the coefficients displayed in the relevant tables). There are more observations in the analyses using raw ratings because a small minority of raters had no variation in their ratings, so we exclude their data when calculating the normalized rating. Lack of variation in ratings across voices is consistent with some raters disregarding the instructions. 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 will use the raw ratings when conducting mechanism analyses in Section 5.

Sample clips were at the **10%-ile** (<https://goo.gl/ZPdCkU>) and **90%-ile** (<https://goo.gl/mbhuLF>) in masculinity ratings (-0.7 and $+0.65$). Sample clips of the same lawyer is consistent with the voice being mutable: **10%-ile** (<https://goo.gl/Gy4Ik7>) and **90%-ile** (<https://goo.gl/NECVV1>) in masculinity ratings of the same lawyer (-0.54 and $+0.2$). Female sample clips also suggest audible differentiation at the **10%-ile** (<https://goo.gl/tvxHFz>) and **90%-ile** (<https://goo.gl/uRbRci>) in femininity ratings (-0.6 and $+0.48$).

Strictly speaking, we want to highlight a distinction we want to make between mutable and concealable, which is studied elsewhere and arguably pertain to small populations (Mueller-

Smith 2013; Chen et al. 2014b). The sample clips suggest that a broad part of the population are subject to stereotypes, which was recently formally modeled in Bordalo et al. (2014).

3.3 Empirical Strategy

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

$$(3) \quad win_{it} = \alpha + \widehat{\mathbf{attribute}}'_{itw}\beta + \mathbf{x}'_{itw}\gamma + \varepsilon_{itw},$$

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. We further expand our analysis by including in some regressions additional covariates denoted by the vector \mathbf{x} .

To explore whether heterogeneity of the MTurk raters influence our findings, \mathbf{x} includes rater w ’s age, gender, race, income, education, and state of residence. Further analyses also include lawyer i ’s characteristics, such as 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. We also control for pre-trial characteristics of case t : 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, ideological perfectionism, or fundamentalism. 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 and respondents are presented separately. In the correlational analysis, Chen et al. (2016e) shows that the baseline estimates are similar whether using linear probability, probit, or logistic regressions, so we omit presenting that here.

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

$$(4) \quad win_{it} = \alpha + \widehat{\mathbf{attribute}}'_{it}\beta + \mathbf{x}'_{it}\gamma + \varepsilon_{it}.$$

In these regressions we cannot control for intra-rater rating correlations nor rater character-

istics. 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:

$$(5) \quad \textit{judge votes for advocate}_{itw} = \alpha + \mathbf{attribute}'_{itw}\beta + \mathbf{x}'_{itw}\gamma + \varepsilon_{itw},$$

where w subscripts for judge.

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

$$(6) \quad \textit{win}_{it} = f(\textit{judge impression}) = \alpha + \overline{\mathbf{attribute}'_{it}}\tilde{\beta} + \mathbf{x}'_{it}\gamma + \varepsilon_{it}.$$

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

Simulations of the data-generating process in Equation (7) and comparing β and $\tilde{\beta}$ yield two separate observations. First, β is more significant than $\tilde{\beta}$ if $\mathbf{attribute}'_{itw}$ 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 $\sigma(\mathbf{attribute}'_{itw}) > \sigma(\overline{\mathbf{attribute}'_{it}})$. The proper interpretation is in terms of standard deviation units.

As a diagnostic, we directly examine the correlation across traits. In Table II, we show the correlation in trait judgments in the micro-level strategy. The table shows that judgments are highly correlated across traits. Principal components analysis shows that the voice judgments clustered in one factor that explains 66 percent of the variation in the data. A separate regression tree analysis yields masculinity and confidence in the final pruned tree.

TABLE II
CORRELATION IN TRAIT JUDGMENTS (MICRO-LEVEL)

Variable	Masculine	Confident	Attractive	Intelligent	Trustworthy	Aggressive	Likely winner
Masculine	1.0000						
Confident	0.4418	1.0000					
Attractive	0.3381	0.3604	1.0000				
Intelligent	0.2332	0.4006	0.3477	1.0000			
Trustworthy	0.2002	0.2658	0.3546	0.3678	1.0000		
Aggressive	0.3448	0.4974	0.2296	0.2349	0.1017	1.0000	
Likely winner	0.4133	0.5593	0.4387	0.4770	0.3968	0.3917	1.0000

4 SCOTUS Outcomes and Perceived Masculinity

4.1 Baseline

We examine the relationship between how people perceive certain voice attributes and whether these perceptions can predict real outcomes. Focusing on male advocates, we find a robust negative association between the perceived masculinity of male advocates and winning in the Supreme Court (Figure 1). The result of a simple regression of the case outcomes (Win) on $Masculine$ indicates a negative association (-0.0098) between advocates who are perceived as sounding masculine and their likelihood of winning their argument before the SCOTUS. This estimate is halved with the inclusion of lawyer fixed effects in Column 2, so we see that both within-lawyer changes and between-lawyer variance in perceived masculinity predict outcomes. With the inclusion of the full set of trait judgments in Column 3, the coefficient estimates increase to -0.012 . In Column 4, the regression with lawyer fixed effects, the estimate is also statistically significant at the 1% level. Specifically, the likelihood of winning a case is 0.86 percentage points lower with every one-standard-deviation increase of the masculine measure ($p < 0.01$; Table III Column 4). This is roughly one-quarter of the gender gap: i.e., males are 3.7 percentage points more likely to win than females. Table III Columns 5–8 present the analogous specifications for the case-level strategy using the average normalized ratings, where one standard deviation of masculinity score is associated with roughly 36 percentage points in Column 8 ($p < 0.1$). (The standard deviation in average masculinity score is 0.54, which divides the coefficient displayed in the table.) Petitioners win about 70% of cases, so the case-level effect is sizable. The micro-level effect is smaller for the reasons previously described. The effect size is subject to the usual caveats in the literature—they are not necessary conditions for an outcome (Deaton 2010).¹

Figure 1 displays the association between the perceived masculinity of male advocates and Supreme Court wins using only within-advocate variation. In other words, this correlation

¹A defendant who shares the same first initial as a judge receives 8% longer sentences, but this effect only explains 0.03% variance (Chen and Prescott 2016). See also Angus Deaton’s NYU “Debates in Development” lecture on the topic, where he describes causal effects as Insufficient but Non-redundant parts of a condition which is Unnecessary but Sufficient (INUS).

TABLE III
 BASELINE RESULTS (NORMALIZED RATINGS)
 Dependent Variable: Case Outcome (win = 1, lose = 0)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Micro level				Case level			
Masculine	-0.00975 (0.00684)	-0.00518* (0.00298)	-0.0122* (0.00695)	-0.00864*** (0.00313)	-0.0314 (0.0228)	-0.101 (0.0888)	-0.0357 (0.0334)	-0.195* (0.109)
Confident			0.00738 (0.00465)	0.00278 (0.00269)			0.106 (0.0666)	0.0316 (0.150)
Attractive			-0.000228 (0.00481)	0.00113 (0.00242)			-0.00770 (0.0457)	0.00688 (0.133)
Intelligent			0.00747* (0.00401)	0.00333 (0.00215)			0.108* (0.0583)	0.125 (0.136)
Trustworthy			-0.00305 (0.00343)	0.000746 (0.00201)			-0.0582 (0.0567)	-0.00510 (0.123)
Aggressive			-0.000718 (0.00415)	-0.000359 (0.00227)			-0.0209 (0.0546)	-0.0375 (0.128)
Likely winner			-0.00451 (0.00419)	0.00406 (0.00275)			-0.105 (0.0708)	0.0885 (0.144)
Lawyer FE	No	Yes	No	Yes	No	Yes	No	Yes
Observations	35276	35276	33666	33666	1634	1634	1634	1634

Notes: Using normalized perceptual ratings, Column 1 tests whether perceived masculinity is unconditionally associated with case outcome. Column 2 tests whether perceived masculinity is associated with case outcome controlling for lawyer fixed effects. Column 3 tests whether perceived masculinity is associated with case outcome controlling for other perceptual ratings. Column 4 tests whether perceived masculinity is associated with case outcome controlling for other perceptual ratings and lawyer fixed effects. Columns 5-8 repeat the same tests but uses the average normalized rating per advocate. Standard errors are clustered by case. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

is driven entirely by changes across cases in the perceived masculinity of a given advocate holding the other voice ratings (i.e., confidence, trustworthiness, etc.) constant. The figure reflects the model presented in Column 4 of Table III and shows that the results are not driven by outliers. No other personality dimension consistently predicts court outcomes, and the perceptions are jointly significant at $p < 0.05$.

Next, we present our baseline results of estimating Equation (3), where Table IV uses raw ratings. Point estimates are similar. The likelihood of winning a case is lower with every one-standard-deviation increase of the masculine measure ($p < 0.01$), as we see in Column 4, and slightly more significant than using the normalized ratings.

TABLE IV
 BASELINE RESULTS (RAW RATINGS, MICRO LEVEL)
 Dependent Variable: Case Outcome (win = 1, lose = 0)

	(1)	(2)	(3)	(4)
Masculine	-0.00707*	-0.00346*	-0.00975**	-0.00674***
	(0.00413)	(0.00184)	(0.00440)	(0.00201)
Confident			0.00417	0.00248
			(0.00312)	(0.00178)
Attractive			-0.000821	0.00198
			(0.00328)	(0.00167)
Intelligent			0.00634**	0.00116
			(0.00292)	(0.00155)
Trust			-0.00177	0.000510
			(0.00264)	(0.00155)
Aggressive			0.00182	0.000421
			(0.00266)	(0.00149)
Likely winner			-0.00270	0.00265
			(0.00315)	(0.00211)
Lawyer FE	No	Yes	No	Yes
Observations	35330	35330	35329	35329

Notes: Using the raw perceptual ratings, Column 1 tests whether perceived masculinity is unconditionally associated with case outcome. Column 2 tests whether perceived masculinity is associated with case outcome controlling for lawyer fixed effects. Column 3 tests whether perceived masculinity is associated with case outcome controlling for other perceptual ratings. Column 4 tests whether perceived masculinity is associated with case outcome controlling for other perceptual ratings and lawyer fixed effects. Standard errors are clustered by case. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

The association between the perceived masculinity of male advocates and Supreme Court wins is robust to a participant-level analysis that controls for how participants correlate ratings of different characteristics, and to a case-level analysis that uses only the average rating per voice clip, as we see in Table V, where one standard deviation of the masculinity measure has correspondingly larger associations in Column 1 and in Column 4. Moreover, with the case-level data, in both Table V and in Columns 5-8 of Table III, we find that estimates produced by variation in within-advocate voice ratings are significantly larger in magnitude than when

using all advocates, including those who appear only once in our sample of oral arguments.

TABLE V
BASELINE RESULTS (RAW RATINGS, CASE LEVEL)

Dependent Variable: Case Outcome (win = 1, lose = 0)				
	(1)	(2)	(3)	(4)
Masculine	-0.0279*	-0.0767	-0.0369	-0.166**
	(0.0169)	(0.0646)	(0.0239)	(0.0778)
Confident			0.0502	0.0358
			(0.0431)	(0.0970)
Attractive			-0.0167	0.0562
			(0.0349)	(0.0933)
Intelligent			0.0907*	0.0398
			(0.0464)	(0.105)
Trust			-0.0185	0.0305
			(0.0461)	(0.0948)
Aggressive			0.0160	-0.00136
			(0.0378)	(0.0836)
Likely winner			-0.0796	0.0551
			(0.0568)	(0.119)
Lawyer FE	No	Yes	No	Yes
Observations	1634	1634	1634	1634

Notes: Using the average raw rating per advocate, Column 1 tests whether perceived masculinity is unconditionally associated with case outcome. Column 2 tests whether perceived masculinity is associated with case outcome controlling for lawyer fixed effects. Column 3 tests whether perceived masculinity is associated with case outcome controlling for other perceptual ratings. Column 4 tests whether perceived masculinity is associated with case outcome controlling for other perceptual ratings and lawyer fixed effects. Standard errors are clustered by case. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Next, we examine the results from design and model perturbations. First, the predictive effect of perceived masculinity is robust across several experimental designs. To address any concerns regarding framing or the number of ratings, we present the results using the same 30 cases from our pilot with 2008 data, using different experimental designs to elicit perceptions from the raters. In Columns 1 and 2 of Table VI, we report regression results using 60 ratings per voice trait where following each recording, the rater was asked to rate the advocate on the full set of traits. Columns 3 and 4 show the results using 60 ratings per voice trait where the rater was randomly assigned to provide a perception of only one trait following each recording. And finally, Columns 5 and 6 show the results using the method we ultimately adopted to elicit perceptions for our full sample of oral arguments (i.e., the same selected 30 cases pulled out of the full data used in the previous tables), but with only 20 ratings per voice trait. While there are slight differences in magnitude across experiments, the results are very similar, further demonstrating the robustness of our key findings on the connection between trait judgments

TABLE VI
 ROBUSTNESS TO EXPERIMENTAL VARIATION (NORMALIZED RATINGS, CASE LEVEL)
 Dependent Variable: Case Outcome (win = 1, lose = 0)

	(1)	(2)	(3)	(4)	(5)	(6)
Masculine	-0.224** (0.0846)	-0.324** (0.131)	-0.219* (0.124)	-0.317 (0.197)	-0.197 (0.140)	-0.434** (0.200)
Attractive		-0.0654 (0.257)		-0.0616 (0.242)		0.111 (0.221)
Confident		0.404 (0.239)		0.346 (0.342)		0.167 (0.253)
Trustworthy		-0.436 (0.574)		-0.347 (0.488)		0.285 (0.288)
Ratings per voice	60	60	60	60	20	20
Cross-lawyer ratings	No	No	Yes	Yes	Yes	Yes
Observations	60	60	60	60	58	58

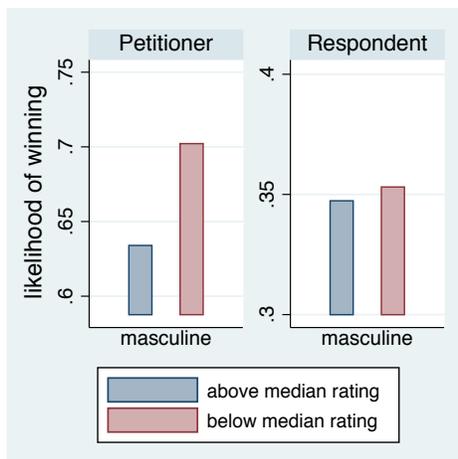
Notes: Using the average normalized rating per advocate, Columns 1-2 test whether perceived masculinity is associated with case outcomes using 60 ratings per voice trait where following each recording, the rater was asked to rate the advocate on the full set of traits. Columns 3-4 do the same, but the rater was randomly assigned to provide a perception of only one trait following each recording. Columns 5-6 do the same using the method ultimately adopted to elicit perceptions for the entire sample of oral arguments, but with only 20 ratings per voice trait. Standard errors are clustered by case. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

from advocates' voices and SCOTUS outcomes.² As such, our results are unlikely to be driven by any specific choice of the number of ratings or framing.

Lawyers who are perceived as more masculine might be construed as being more dominant and authoritative. It is noteworthy that the gendered differentiation of masculine and feminine language has been argued to have a different evolutionary basis (Locke 2011). In the present context, the fact that court outcome is negatively associated with masculinity points to a possible connection with the discourse of dominance. Males are seen as being selected over generations to be aggressive and dominant, but this selective pressure might be a double-edged sword since these types of behaviors are more likely to lead to lethal confrontation. In this context, the dominant and aggressive behavior of masculine-sounding lawyers might have invited an adverse response from the court. To be sure, this is only one interpretation as to explain why masculinity of voice may matter and in the Appendix we explore the extent to which these constructs play a role in the decision process as judges deliberate court decisions. In particular, we include additional elicitations where we control for perceived competence, empathy, respect, and attitude. The results become larger and more statistically significant at both the micro level ($p < 0.05$) and case level ($p < 0.01$), with an R-square of 0.02 and 0.17, respectively. The point estimate is 0.55 for the analog to Column 6 of Table VI. We also confirm that the results are robust (and if anything, more significant) when we model the court outcome in terms of mixed-effects logistic regression, which is often used in the

²Columns 5 and 6 have fewer observations than in Columns 1 to 4 because the final study dropped two clips due to implementation error.

Figure 8: Perceived Masculinity and Win Rate by Petitioner-Respondent Status



linguistics literature, where the fixed effects include the attorney’s role and the random effects are the case.

4.2 Petitioner

If we assume the primacy of first impressions on court decisions, the first person to argue before the Justices should exhibit a stronger vocal first-impression effect. That is, the first speaker may have a longer-lasting impact on the court and subsequent outcomes due to the “anchoring” effect (Tversky and Kahneman 1992), in which individuals rely on an initial piece of information to make subsequent judgments. The respondent speaks 30 minutes later.

Our analysis shows that the petitioner is the main driver of the correlation between perceived masculinity and court outcomes. Similar to the general findings above, a one-standard-deviation increase in the masculinity score of the petitioner is associated with a lower likelihood of winning the case ($p < 0.05$; Table VII Column 4). For ease of interpretation, 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 8). No association between perceived masculinity and court outcomes is found among the lawyers for the respondent. This finding supports the hypothesis that first impressions – in this case, of the first lawyer to argue before the Justices – are disproportionately associated with judicial decisions. We are unable to determine if the literature on Supreme Court oral arguments always observes the petitioners’ characteristics to matter more (e.g., on the pattern of questioning or on Justice Blackmun’s grading of advocates), as the published papers do not make their data publicly available as far as we are aware, and none of the published tables or figures allow an inference.

As a benchmark for comparison, Columns 3 and 6 in Table VII show the results using the full sample of arguments; Column 6 includes lawyer fixed effects in addition to MTurk rater

TABLE VII
 MTURK RATER CHARACTERISTICS AND VOICE PERCEPTION (NORMALIZED RATINGS, MICRO LEVEL)
 Dependent Variable: Case Outcome (win = 1, lose = 0)

	Petitioner (1)	Respondent (2)	Both (3)	Petitioner (4)	Respondent (5)	Both (6)
Masculine	-0.0198** (0.00913)	0.00742 (0.00972)	-0.0123* (0.00695)	-0.00766** (0.00372)	0.00136 (0.00364)	-0.00879*** (0.00312)
Confident	0.00919 (0.00591)	-0.00968 (0.00671)	0.00747 (0.00463)	0.00396 (0.00301)	-0.00642* (0.00344)	0.00276 (0.00270)
Attractive	-0.00134 (0.00641)	-0.000508 (0.00679)	-0.000239 (0.00481)	-0.000188 (0.00275)	0.00300 (0.00326)	0.00103 (0.00242)
Intelligent	0.00744 (0.00531)	0.00371 (0.00564)	0.00743* (0.00400)	0.00291 (0.00240)	-0.000169 (0.00271)	0.00336 (0.00215)
Trustworthy	0.00142 (0.00448)	-0.00949* (0.00505)	-0.00314 (0.00342)	0.00166 (0.00237)	-0.00270 (0.00266)	0.000696 (0.00201)
Aggressive	-0.00493 (0.00533)	0.000942 (0.00584)	-0.000701 (0.00414)	0.00280 (0.00265)	-0.00429 (0.00299)	-0.000389 (0.00227)
Likely winner	-0.00237 (0.00542)	-0.0145** (0.00599)	-0.00441 (0.00419)	0.00294 (0.00293)	-0.000927 (0.00348)	0.00414 (0.00276)
MTurker controls	Yes	Yes	Yes	Yes	Yes	Yes
Lawyer FE	No	No	No	Yes	Yes	Yes
Observations	17665	16001	33666	17665	16001	33666

Notes: Using the normalized perceptual ratings, Column 1 tests whether perceived masculinity is associated with case outcome among petitioners, controlling for other perceptual ratings and for rater characteristics. Column 2 tests whether perceived masculinity is associated with case outcome among respondents, controlling for other perceptual ratings and for rater characteristics. Column 3 tests the same for both sets of advocates. Columns 4-6 do the same as Columns 1-3 and control for lawyer fixed effects. Standard errors are clustered by case. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

controls. As expected, the coefficient estimates on *Masculine* are very similar to those obtained in the baseline regression. The most interesting observation we obtain from Tables VII (using normalized ratings) and VIII (using raw ratings) is that the predictive power stems from the voice judgments of the petitioner, not the respondent, corroborating our conjecture that the first speaker may have a long-lasting impact on the Court and subsequent outcomes. Further, comparing Column 1 to Column 4, we see that nearly 40% of the magnitude of β_1 stems from within petitioner variation, whereas the other 60% stems from cross-petitioner variation in trait judgments of masculinity. Overall, a one-standard-deviation increase in perceived masculinity is associated with a 1.98 percentage-point lower likelihood of the petitioner winning the case.

TABLE VIII
MTURK RATER CHARACTERISTICS AND VOICE PERCEPTION (RAW RATINGS, MICRO LEVEL)

	Dependent Variable: Case Outcome (win = 1, lose = 0)					
	Petitioner	Respondent	Both	Petitioner	Respondent	Both
	(1)	(2)	(3)	(4)	(5)	(6)
Masculine	-0.0144** (0.00571)	0.00491 (0.00630)	-0.00963** (0.00444)	-0.00569** (0.00241)	0.000420 (0.00232)	-0.00668*** (0.00203)
Confident	0.00457 (0.00390)	-0.00924** (0.00438)	0.00379 (0.00314)	0.00314 (0.00194)	-0.00520** (0.00219)	0.00233 (0.00179)
Attractive	-0.000428 (0.00447)	-0.00271 (0.00479)	-0.00137 (0.00332)	0.000228 (0.00185)	0.00148 (0.00235)	0.00152 (0.00169)
Intelligent	0.00323 (0.00390)	0.00652 (0.00425)	0.00662** (0.00298)	0.00187 (0.00177)	0.00108 (0.00204)	0.00201 (0.00157)
Trust	0.00279 (0.00351)	-0.00475 (0.00399)	-0.00221 (0.00267)	0.00105 (0.00189)	-0.000211 (0.00213)	0.000673 (0.00156)
Aggressive	-0.00156 (0.00347)	0.00522 (0.00372)	0.00182 (0.00272)	0.00174 (0.00179)	-0.000667 (0.00199)	0.000554 (0.00151)
Likely winner	-0.000989 (0.00415)	-0.0101** (0.00452)	-0.00258 (0.00320)	0.00172 (0.00228)	-0.00112 (0.00261)	0.00235 (0.00216)
MTurker controls	Yes	Yes	Yes	Yes	Yes	Yes
Lawyer FE	No	No	No	Yes	Yes	Yes
Observations	18542	16787	35329	18542	16787	35329

Notes: Using the raw perceptual ratings, Column 1 tests whether perceived masculinity is associated with case outcome among petitioners, controlling for other perceptual ratings and for rater characteristics. Column 2 tests whether perceived masculinity is associated with case outcome among respondents, controlling for other perceptual ratings and for rater characteristics. Column 3 tests the same for both sets of advocates. Columns 4-6 do the same as Columns 1-3 and control for lawyer fixed effects. Standard errors are clustered by case. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

As described earlier, we have a large set of rater covariates. The rater characteristics are given as categorical covariates, and we include indicator variables for each value the categorical covariate takes. Specifically, the fully saturated model includes rater characteristics: age and dummies for each racial group, income cohort, gender group, education level, and state of

residence. In Tables VII and VIII, we present regression results with this set of controls.

4.3 Within vs. Between Variation

Thus far, we have found that first impressions of lawyers' voices (specifically related to perceived masculinity) predict U.S. Supreme Court outcomes. Males are more likely to win when they are perceived as less masculine. Petitioners (the first lawyer to argue) are the main driver. Since our findings persist removing cross-advocate variation, we now look more closely at between vs. within variation among repeat advocates.

We partition the sample of advocates into two groups – those who argued only once during the 15 years of our study and those who argued multiple times – which results in splitting the voices roughly in half. We present the results in Tables IX (using raw ratings) and A.1 (using normalized ratings). The regression results, which do not control for lawyer fixed effects, suggest that experienced petitioners are the primary driver of the correlation between perceived masculinity and outcomes, with a reduction in the likelihood of winning as the result of an experienced petitioner perceived as sounding one-standard-deviation more masculine. Note that since this table does not include lawyer fixed effects, the average of Columns 1 and Column 4 in Table A.1 would roughly approximate Column 1 from Table VII. Comparing Column 4 of Table VII and Column 4 of Table A.1 yields the inference that roughly 30% of the magnitude of β_1 stems from within-advocate variation among advocates who argued more than once. Among experienced petitioners, roughly 30% of the association between perceived masculinity and court outcomes comes from within-lawyer variation, while 70% comes from between-lawyer variation.

TABLE IX
ADVOCATE EXPERIENCE AT THE SCOTUS AND OUTCOMES (RAW RATINGS, MICRO LEVEL)
Dependent Variable: Case Outcome (win = 1, lose = 0)

	Petitioner (1)	Respondent (2)	Both (3)	Petitioner (4)	Respondent (5)	Both (6)
	argued only once			argued more than once		
Masculine	-0.00189 (0.00861)	0.00755 (0.00850)	-0.000210 (0.00630)	-0.0190*** (0.00726)	0.00516 (0.00872)	-0.0128** (0.00598)
Confident	-0.000625 (0.00650)	-0.00543 (0.00659)	0.000428 (0.00480)	0.00709 (0.00443)	-0.0111* (0.00569)	0.00599 (0.00392)
Attractive	-0.00295 (0.00728)	-0.00947 (0.00654)	-0.00676 (0.00507)	-0.000407 (0.00531)	0.00260 (0.00653)	0.00167 (0.00442)
Intelligent	0.00297 (0.00604)	0.00822 (0.00646)	0.00805* (0.00454)	-0.00000342 (0.00508)	0.00461 (0.00525)	0.00245 (0.00379)
Trust	0.00315 (0.00573)	-0.00437 (0.00533)	-0.00215 (0.00397)	0.00348 (0.00418)	-0.00474 (0.00574)	-0.00149 (0.00354)
Aggressive	0.0000306 (0.00590)	0.00685 (0.00511)	0.00285 (0.00408)	-0.00134 (0.00409)	0.00445 (0.00507)	0.00235 (0.00348)
Likely winner	-0.00938 (0.00621)	-0.0205*** (0.00650)	-0.0131*** (0.00450)	0.00367 (0.00540)	-0.00228 (0.00622)	0.00440 (0.00444)
Observations	8016	7950	15966	10526	8837	19363

Notes: Using the raw perceptual ratings, Column 1 tests whether perceived masculinity is associated with case outcome among petitioners who argued only once in the Supreme Court from 1998-2012. Column 2 tests whether perceived masculinity is associated with case outcome among respondents who argued only once in the Supreme Court from 1998-2012. Column 3 tests the same for both sets of advocates. Columns 4-6 do the same as Columns 1-3 but among advocates who argued more than once. Standard errors are clustered by case. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

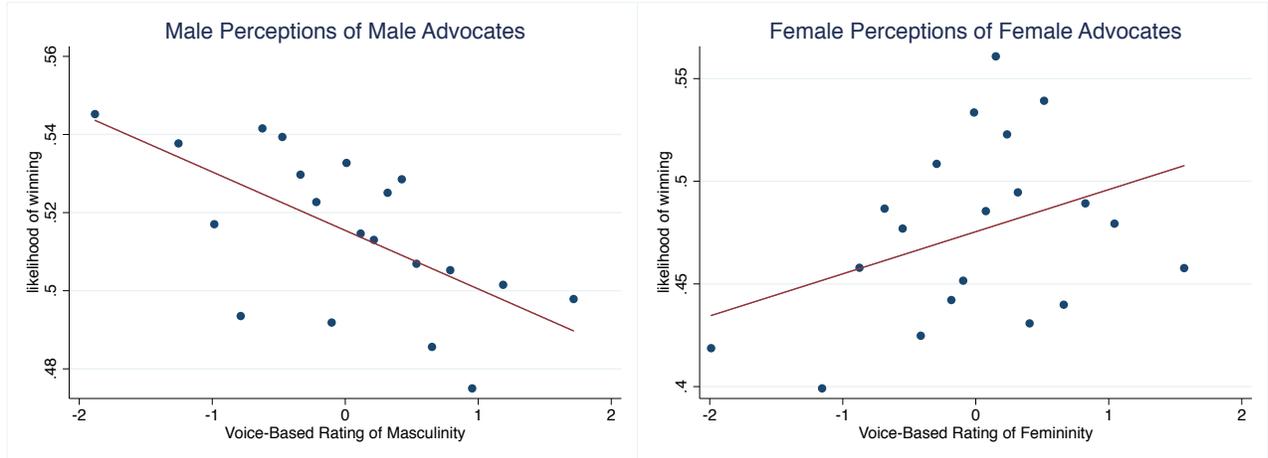
4.4 Gender

Since female lawyers are also coached to be more masculine (Starecheski 2014), this raises the question of whether our findings are restricted to male advocates or whether they also extend to female lawyers. 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.

Our 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 9). If “masculine” were the opposite of “feminine,” then the pooled results would be stronger.

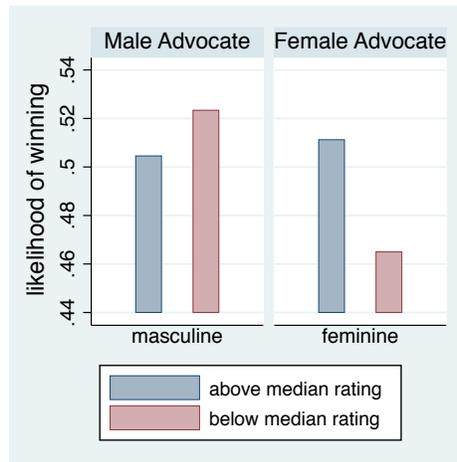
Another interpretation of our results—where we discretize the continuous rating measures at the lawyer level—is that a below-median masculinity rating corresponds with a roughly 2 percentage-point greater likelihood of winning for males, but a below-median femininity rating corresponds with an approximately 5 percentage-point lower likelihood of winning for females

Figure 9: Graphical Illustration of Gender and Predicted Outcomes



Notes: These figures present binned scatterplots illustrating the association between voice-based masculinity (femininity) rating and court outcomes.

Figure 10: Perceptions of Masculinity (Femininity) and Advocate Win Rates



(Figure 10).

Interestingly, the correlation with outcomes is specific to perceived femininity. In Tables X (using raw ratings) and A.2 (using normalized ratings), we present our initial 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.

TABLE X
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

Notes: Using the raw perceptual ratings, 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 percent, 5 percent, and 1 percent levels, respectively.

In Tables XI, 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.

TABLE XI
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	
	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

Notes: Using the average raw rating per advocate, 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 percent, 5 percent, and 1 percent 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 XII 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.

TABLE XII
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	
		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	Yes
Lawyer FE	No	No	No	No	No	Yes	Yes	Yes	Yes
Observations		18542	35329	2543	5868	18542	35329	2543	5868

Notes: Using the raw perceptual ratings, 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 percent, 5 percent, and 1 percent levels, respectively.

4.5 Discussion

Our results are robust to dropping lawyers with large caseloads; to a permutation (placebo) test in which we randomly reassign voice ratings to another lawyer and verify null effects; to day-of-week, seasonality, and year fixed effects; and to using the other subjects' ratings as instrumental variables. The focus on petitioners precludes case fixed effects, but a regression with case fixed effects suggests that respondents may be imitating the petitioner to some extent or simply have more noise. In Table XIII, we consider the potential role of the swing voter. We report significant correlations at the outcome 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 judges 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.

To our knowledge, this is the first study that documents an association between voice-based impressionistic judgments and judicial decisions. To benchmark our findings, the 0.86 percentage-point difference in court outcomes attributed to a one-standard-deviation change

TABLE XIII
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 *				-0.0249 (0.0237)
Number of Democrats				
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

in our voice-based measure of perceived masculinity is equivalent to nearly one-quarter of the gender gap (i.e., male lawyers are 3.7 percentage points more likely to win a court case than female lawyers). When analyzed at the case level, one-standard-deviation change corresponds to 5.2 percentage-point difference in court outcomes (when not including further controls).³

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 and Atkins 1980) or the discursive practices in the courtroom (O’Barr 1982). We also build on a literature using limited samples of Supreme Court oral arguments that finds, for example, 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). To the best of our knowledge, no studies have focused on perceptions of the lawyers’ voices where lawyers are using identical words.

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 and McConnell-Ginet 2003; Butler 1990; Kessler and McKenna 1978). Different cultures may also construct different notions of masculinity that are reflected in the stereotypical ways of talking and thinking about men and masculinity. In the United States, there are four main cultural discourses of masculinity (Kiesling 2007): gender difference, which pertains to categorical difference in biology and behavior between men and women; heterosexism, which sees being masculine as sexually desiring women and not men; dominance, which links masculinity with notions of authority or power; and male solidarity, which assumes a given bond among men.

³2.79 divided by 0.54.

Sections 5-7 address the nature of the association between voice judgments and court outcomes – i.e., why are court outcomes correlated with perceived masculinity but not other attributes? Given that the attributes are positively correlated with each other, the fact that only perceived masculinity is found to correlate with court outcomes suggests that masculinity captures a particular variance that is not captured by the other ratings.

Our findings thus far do not allow us to conclude whether the Justices were engaging in a form of linguistic profiling when making their judicial decisions nor whether law firms were engaging in a form of linguistic profiling when choosing their oral advocates. However, we will document that the associations are robust to controlling for pre-trial case characteristics, lawyer characteristics, acoustic features, and the best prediction of Supreme Court votes based on non-linear combinations of these characteristics, judge trends, and court trends.

Under complete markets, no extraneous factor should predict stock returns, since the market should incorporate all information immediately. Similarly, no extraneous factor should predict win rates, since the lawyers and law firms should adjust. Observing effects within lawyer precludes fixed unobservable lawyer characteristics from driving the entire result. One interpretation may be that even the highly skilled cannot conceal their tone – an unconscious response to weak cases. Another is that macho men select hard cases – a conscious response to weak cases. However, observing substantial judge heterogeneity precludes fixed unobservable case characteristics from driving the entire result. The pattern of results for women precludes judge constraints from driving the entire result. As we will see, Democrat votes correlate more positively with less masculine-sounding males, while Republican votes correlate more positively with feminine-sounding females.

5 Why Does Perceived Masculinity Predict SCOTUS Outcomes?

There are at least four possible sources for why first impressions of lawyers' voices predict U.S. Supreme Court outcomes: lawyers, justices, firms, and clients. Lawyers perceived as masculine may be different along relevant dimensions other than masculinity, and within-lawyer voice masculinity may be a response to case weakness. Justices may be linguistically profiling, which may be strategic, sincere, or homophily. Firms may be linguistically profiling in choosing their oral advocates. The “perceived” or performative nature of masculinity may involve lawyers performing for another audience or client. In Austen-Smith and Fryer's (2005) ethnic identity model, an individual has a "two-audience" signaling problem. Grogger (2011) adapted this model to the linguistic choices of minorities, in which one audience is the labor market and another is the peer group. In our context, one audience is the SCOTUS and another is the firm. What we can do to investigate these hypotheses is to examine whether the coefficient on

perceived masculinity is affected by observable lawyer controls and pre-trial controls, whether masculinity ratings are correlated with votes by some judges but not others (and for some industries but not others), and whose ratings of masculinity predict outcomes.

5.1 Are Lawyers Perceived as Masculine Different?

We confirmed the robustness of the association between perceived masculinity and court outcomes in a study using a subsample of 30 cases (i.e., 60 male voice clips) argued in 2008 and included in the analysis hand-collected lawyer covariates, such as age, number of clerkships, whether the lawyer was a member of his school’s law review, and whether the lawyer graduated from a top-5 law school. Our objective is to examine whether the connection between perceptions of voices as masculine and real outcomes reflects other underlying characteristics of advocates that may influence their success rates at the SCOTUS. We present the results using the case-level data on these advocates in Table XIV. In Column 1, we show that the negative correlation between *Masculine* and win probabilities continues to hold. In particular, with the addition of the full set of voice traits, the coefficient estimate on *Masculine* is significantly larger than the ones we obtained in earlier regressions. Nonetheless, the results are striking. In Column 3, we add the full set of advocate covariates we have managed to obtain. These include a dummy for whether the advocate was previously a SCOTUS clerk, the lawyer’s age, the number of clerkships the lawyer had, whether he was a member of the law review, a dummy for whether he had an additional graduate degree, and whether he is a graduate of Harvard, Yale, Stanford, Columbia, or the University of Chicago (‘Top-5’). The coefficient on *Masculine* becomes further negative and is statistically significant at the 1% level. Thus, at least for this set of cases, it does not appear that the predictive power of perceptions of masculinity is driven by advocate characteristics. Finally, in Column 4 we add case fixed effects to explore whether the results on masculinity are relative (i.e., petitioner relative to respondent). The estimates suggests that when using only within-case variation, the correlation is further amplified. Since this difference does not appear in the 15-year sample for the reasons described above, we do not explore it further.

Table XIV is unable to make far-reaching conclusions, given the small sample size. The next two tables include all the other lawyer covariates described in the methodology section (number of previous SCOTUS oral arguments, years since graduation, number of admitted courts, number of practice areas, firm size, office size, being a law firm partner, and having graduated from a top-5 law school), for the entire sample of lawyers with biographical data on Westlaw or Lexis.⁴ About 50% of the lawyers could be located (missing lawyers’ data was dummied out). The perceived masculinity correlation is highly robust (Table XV). Column 1 begins at the micro level without lawyer controls, Column 2 adds masculinity of name,

⁴The biographical pages were manually saved by research assistants, then a computer code digitized the hand-collected files.

TABLE XIV
PERCEPTIONS OF VOICE AND LAWYER CHARACTERISTICS (NORMALIZED RATINGS, CASE LEVEL)

Dependent Variable: Case Outcome (win = 1, lose = 0)				
	(1)	(2)	(3)	(4)
Masculine	-0.197 (0.140)	-0.434** (0.200)	-0.719*** (0.206)	-1.203* (0.618)
Attractive		0.111 (0.221)	0.281 (0.294)	0.384 (0.850)
Confident		0.167 (0.253)	0.277 (0.202)	0.545 (0.554)
Trustworthy		0.285 (0.288)	0.279 (0.317)	0.559 (1.419)
SCOTUS clerk			-0.0257 (0.227)	0.134 (0.817)
Lawyer's age			0.0121* (0.00646)	0.0203 (0.0261)
Number of clerkships			-0.114 (0.116)	-0.182 (0.465)
Law review			0.415*** (0.131)	0.815** (0.363)
Other graduate degree			0.115 (0.158)	-0.00364 (0.499)
Top-5 law school			0.0687 (0.153)	0.0956 (0.397)
Case fixed effects	No	No	No	Yes
Observations	58	58	58	58
R-squared	0.037	0.096	0.258	0.411

Notes: Using the average normalized rating per advocate, Columns 1-4 test whether perceived masculinity is associated with case outcomes while controlling for an increasing number of advocate characteristics. Columns 1-2 repeat the prior analyses for the subset of advocates for which we have data on age, number of clerkships, law school, graduate degrees, and having served on law review or as a Supreme Court clerk. Column 3 includes these characteristics as controls. Column 4 includes case fixed effects. Standard errors are clustered by case. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

Column 3 adds SCOTUS experience, and Column 4 includes all covariates. The coefficient magnitude is highly stable. In Column 5, we examine the relationship at the case level, where the coefficient is still significant at $p < 0.05$. Notably, the relationship remains specific to perceived masculinity. If perceptions of masculinity were simply reflecting other important advocate covariates, then the coefficient estimates on *Masculine* should be driven to zero. That this is not the case suggests that the channel of how trait judgments stemming from an extremely brief voice clip predict outcomes may not be as simple as one might expect. We do not have data on body size, but researchers articulate no robust association between voice pitch and body size (Fitch 1997; Künzel 1989).

Table XVI presents the analogous study for female lawyers. The point estimates remain similar across Columns 1-4, but become statistically less significant. Notably, feminine voices are perceived as more attractive (Puts et al. 2011; Feinberg et al. 2005), but we have controlled for attractiveness and its association is far smaller and less significant than for femininity.

5.2 Is Voice Masculinity a Response to Case Weakness?

Are pre-trial characteristics correlated with perceived masculinity? Using the Supreme Court Database⁵, we find that perceived masculinity still predicts outcomes after controlling for all available pre-trial data in the database, which is described in the Methodology section. Column 1 of Table XVII 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. Note that lower-court dissent is positively associated with the petitioner winning, which is noted in the large literature on the role of judicial dissents. Yet the association is only eight times larger than the association for perceived masculinity unit change. Since the standard deviation of the raw masculinity ratings was 1.5, this means the association with lower-court dissent is 5.5 times larger than the association with one-standard-deviation decrease in perceived masculinity. 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). Political division in the lower court is associated with a win by the petitioner, which is consistent with theoretical models of judicial decision-making (Beim et al. 2016; Chen et al. 2015), and the magnitude of the coefficient on masculinity becomes slightly larger relative to Column 2. Columns 3–6 include lawyer fixed effects. As before, the association between perceived masculinity and case outcomes is smaller. Notably, the presence of a lower-court dissent no longer predicts a ruling in favor of the petitioner, while the political division of the lower court panel does. We will return to pre-trial characteristics in Section 7,

⁵<http://scdb.wustl.edu/data.php>

Table XV: Case Outcomes and Perceived Masculinity

	Dependent Variable: Case Outcome (win = 1, lose = 0)					
	-0.0149***	-0.0145***	-0.0151***	-0.0139***	-0.0864**	
Masculine	(0.00565)	(0.00559)	(0.00558)	(0.00537)	(0.0340)	
Confident	0.00508	0.00535	0.00595	0.00482	0.0851	
	(0.00387)	(0.00386)	(0.00385)	(0.00382)	(0.0539)	
Attractive	0.0000377	-0.000927	-0.000399	0.000460	-0.00237	
	(0.00445)	(0.00445)	(0.00441)	(0.00431)	(0.0501)	
Intelligent	0.00244	0.00264	0.00309	0.00166	-0.0167	
	(0.00385)	(0.00384)	(0.00381)	(0.00375)	(0.0639)	
Trust	0.00356	0.00336	0.00330	0.00305	0.0644	
	(0.00344)	(0.00343)	(0.00345)	(0.00338)	(0.0618)	
Aggressive	-0.00134	-0.00139	-0.00145	-0.00170	-0.0235	
	(0.00345)	(0.00343)	(0.00343)	(0.00339)	(0.0472)	
Likely winner	-0.000977	-0.00118	-0.000821	-0.00152	-0.0401	
	(0.00411)	(0.00411)	(0.00412)	(0.00405)	(0.0755)	
Masculinity of Name	N	Y	Y	N	N	
SCOTUS Experience	N	N	Y	N	N	
Additional Lawyer Covariates	N	N	N	Y	Y	
Collapsed	N	N	N	N	Y	
Observations	18542	18542	18542	18542	856	
R-squared	0.002	0.006	0.008	0.018	0.026	
Sample: Male Petitioners						

Table XVI: Case Outcomes and Perceived Femininity

	Dependent Variable: Case Outcome (win = 1, lose = 0)	
Feminine	0.0125* (0.00678)	0.0135** (0.00668)
Confident	0.00701 (0.00730)	0.0104 (0.00647)
Attractive	-0.00587 (0.00743)	0.00690 (0.00714)
Intelligent	-0.00578 (0.00847)	-0.00508 (0.00715)
Trust	-0.00175 (0.00647)	-0.00408 (0.00838)
Aggressive	0.00243 (0.00786)	-0.00140 (0.00629)
Likely winner	0.00361 (0.00785)	-0.00182 (0.00771)
Masculinity of Name	N	Y
SCOTUS Experience	N	Y
Additional Lawyer Covariates	N	N
Collapsed	N	N
Observations	5868	5868
R-squared	0.004	0.039
Sample: Female Lawyers		0.079
		5868
		0.079
		267
		0.088

TABLE XVII
PRE-TRIAL CHARACTERISTICS AND CASE OUTCOMES

	Dependent Variable: Case Outcome (win = 1, lose = 0)					
Masculine	-0.0153***	-0.0125**	-0.0166**	-0.00546**	-0.00554**	-0.00389*
	(0.00576)	(0.00576)	(0.00756)	(0.00225)	(0.00228)	(0.00235)
Aggressive	-0.000563	0.0000108	0.00181	0.00228	0.00263	0.00217
	(0.00352)	(0.00350)	(0.00424)	(0.00173)	(0.00175)	(0.00164)
Attractive	0.00189	-0.000299	0.00189	0.00202	0.00172	0.00101
	(0.00454)	(0.00443)	(0.00570)	(0.00179)	(0.00172)	(0.00191)
Confident	0.00554	0.00390	0.00199	0.00284	0.00258	-0.00221
	(0.00396)	(0.00384)	(0.00461)	(0.00191)	(0.00196)	(0.00192)
Intelligent	0.00250	0.00260	0.00191	0.000816	0.000983	0.00250
	(0.00391)	(0.00389)	(0.00536)	(0.00175)	(0.00178)	(0.00179)
Trust	0.00391	0.00560	0.00421	0.00232	0.00242	0.000647
	(0.00352)	(0.00345)	(0.00428)	(0.00179)	(0.00178)	(0.00172)
Likely winner	-0.00194	-0.00230	-0.00435	0.000736	0.000825	0.000916
	(0.00421)	(0.00426)	(0.00518)	(0.00219)	(0.00225)	(0.00204)
Lower Court Disagreement		0.107***	0.142***		0.0323	-0.0659
		(0.0366)	(0.0518)		(0.0537)	(0.0786)
Lower Court Reversing		-0.0218	-0.00438		-0.0909*	-0.0964
Trial Court		(0.0385)	(0.0483)		(0.0543)	(0.0703)
Lower Court Liberal		0.0750**	0.0406		0.0155	0.0127
Decision		(0.0355)	(0.0471)		(0.0592)	(0.0771)
Lower Court Opinion			-0.00612			0.0203
Word Count			(0.0134)			(0.0256)
Lower Court Opinion			-0.00130			-0.647
Self-Certainty Words			(0.402)			(0.710)
Lower Court Politically			0.0977*			0.139*
Divided			(0.0559)			(0.0801)
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						

which considers the previous best existing prediction models of Supreme Court votes. These models incorporate trends at the Justice, Supreme Court, and lower-court levels.

5.3 Do Judges Differ in How Their Votes Correlate with Masculine Voices?

Votes of Democrats, but not Republicans, are negatively associated with perceived masculinity (Figure 2). In the higher intercept for Republicans, we see that Republicans favor males, as documented previously by 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 gra-

TABLE XVIII
POLITICAL PARTY AND RESPONSE TO MALE PETITIONER MASCULINITY ($p < 0.01$)

	(1)
	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

dent for Republicans indicates that the adverse outcome for masculine males can be entirely attributed to Democratic votes. All analyses in this sub-section regress:

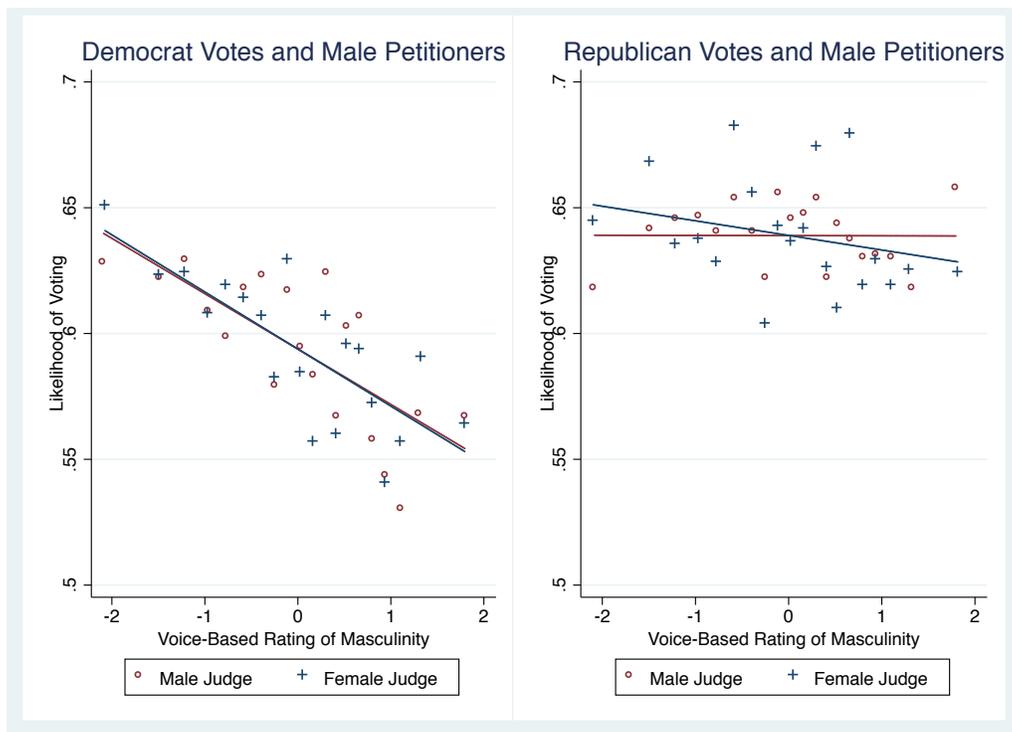
$$(1) \quad \text{judge votes for advocate}_{itwj} = \alpha + \mathbf{attribute}'_{itwj}\beta + \mathbf{x}'_{itwj}\gamma + \varepsilon_{itwj},$$

where w subscripts for rater and j for judge. Standard errors are clustered at the lawyer and judge levels, and judge fixed effects are included. Interactions between $\mathbf{attribute}'_{itwj}$ and indicators for whether the judge is a Democrat or Female are included in some models. In the remainder of Section 5, we alternatively consider interactions with case, firm, and rater characteristics.

The heterogeneity along the dimension of judges' 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 XVIII). Nor are experience and age driving factors. Although only 13 individuals serve as Supreme Court Justices during our time frame, the differences between judges based on party affiliation are highly significant ($p < 0.01$).

If we fully interact gender and party, we see a slight gender effect (Figure 11). This is largely due to Sandra Day O'Connor, whose reaction to perceived masculinity exists relative to her male Republican colleagues, but is far more muted than how her female Democratic and male Democratic colleagues react.

Figure 11: Political Party and Response to Male Petitioner Masculinity ($p < 0.01$)



If we replace party with judges' ideology scores (Martin and Quinn 2002) or include both together, we find that ideology ($p < 0.01$) is more important than party, which is rendered insignificant (Figure 12).

Figure 13 considers judge heterogeneity for within-lawyer variation and for lawyers who appeared only once in the Supreme Court. When we include lawyer fixed effects on the left side of Figure 13, it suggests that something other than judges' ideology would need to explain within-lawyer effects. Notably, the right-hand side indicates that the judicial heterogeneity for between-lawyer effects when restricted to lawyers who argued only once is quite strong, so the between-lawyer variance for non-repeat advocates may have been masked in the analysis looking at court outcomes only.

We also find that Republicans vote for feminine-sounding females more than Democrats do (Figure 14). The difference is significant at $p < 0.1$. Notably, this rules out an explanation in which Republican judges have "made up their minds" regardless of the lawyer and only Democrats pick up on case weakness correlated with more masculine male voices. For an explanation of case weakness as an omitted variable, one would need Democrats to notice this more for males and Republicans to notice this more for females, which is a somewhat more complicated hypothesis than one in which firms and clients perceive masculine voices as winning voices and select them at the risk of losing in the court.

This result is again not affected by the judge's gender, in a specification including interactions with gender and party. Fully interacting gender and party indicates that among

Figure 12: Judicial Ideology and Response to Male Petitioner Masculinity ($p < 0.01$)

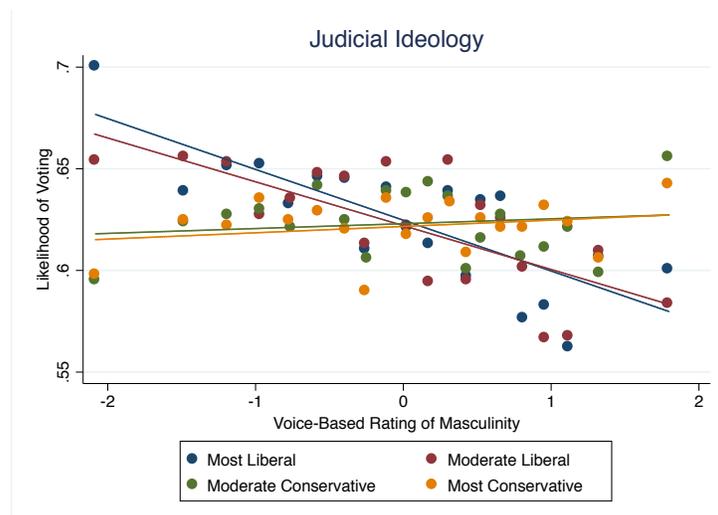


Figure 13: Political Party and Response to Male Petitioner Masculinity (Within and Between Lawyers)

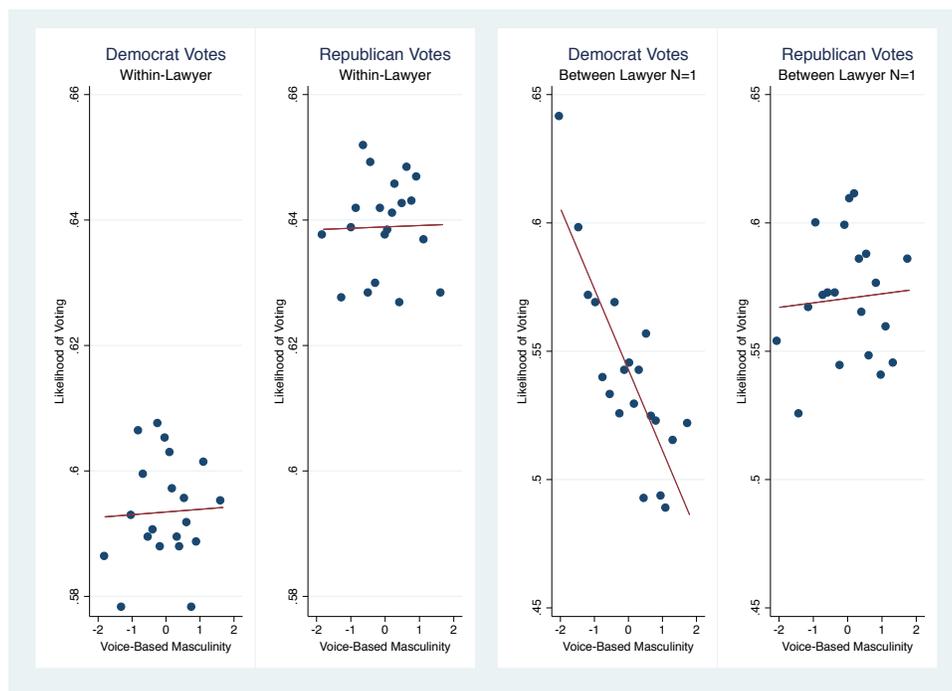
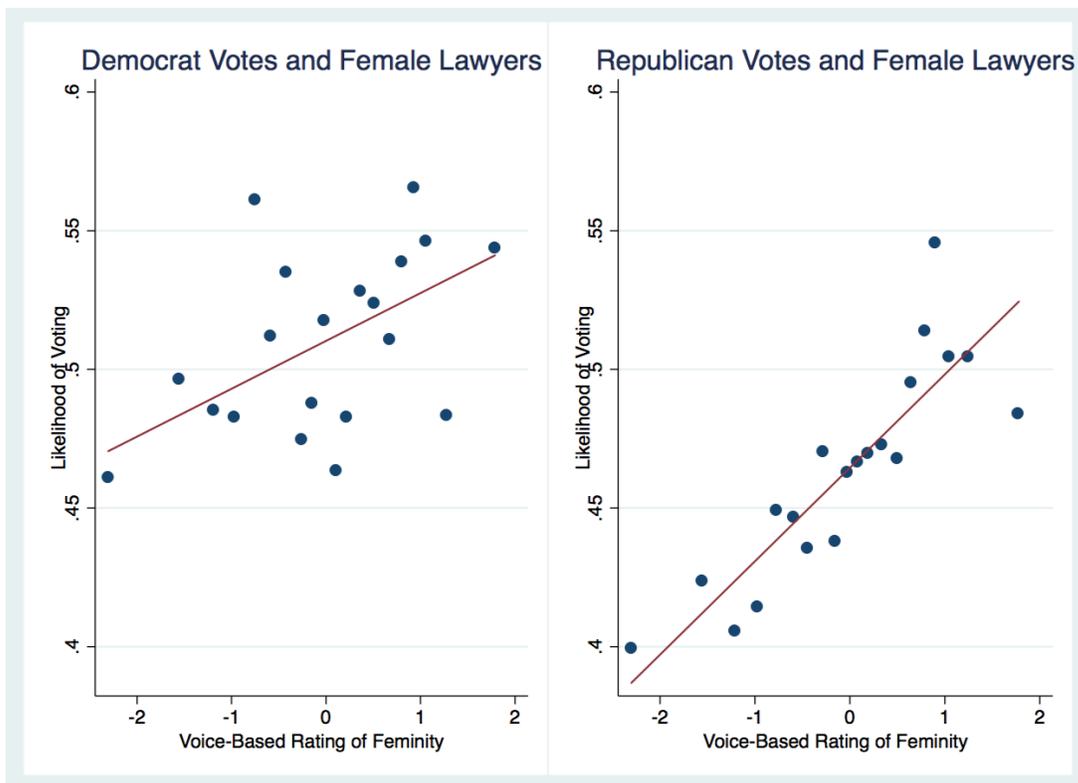


Figure 14: Political Party and Response to Femininity ($p < 0.1$)



Democrats, males respond slightly more favorably to feminine females, while among Republicans, Sandra Day O'Connor responds more favorably to feminine females (Figure 15). Among females, the Republican (Sandra Day O'Connor) significantly responds more favorably to feminine females than her female Democratic counterparts. The ideological difference is more muted for males.

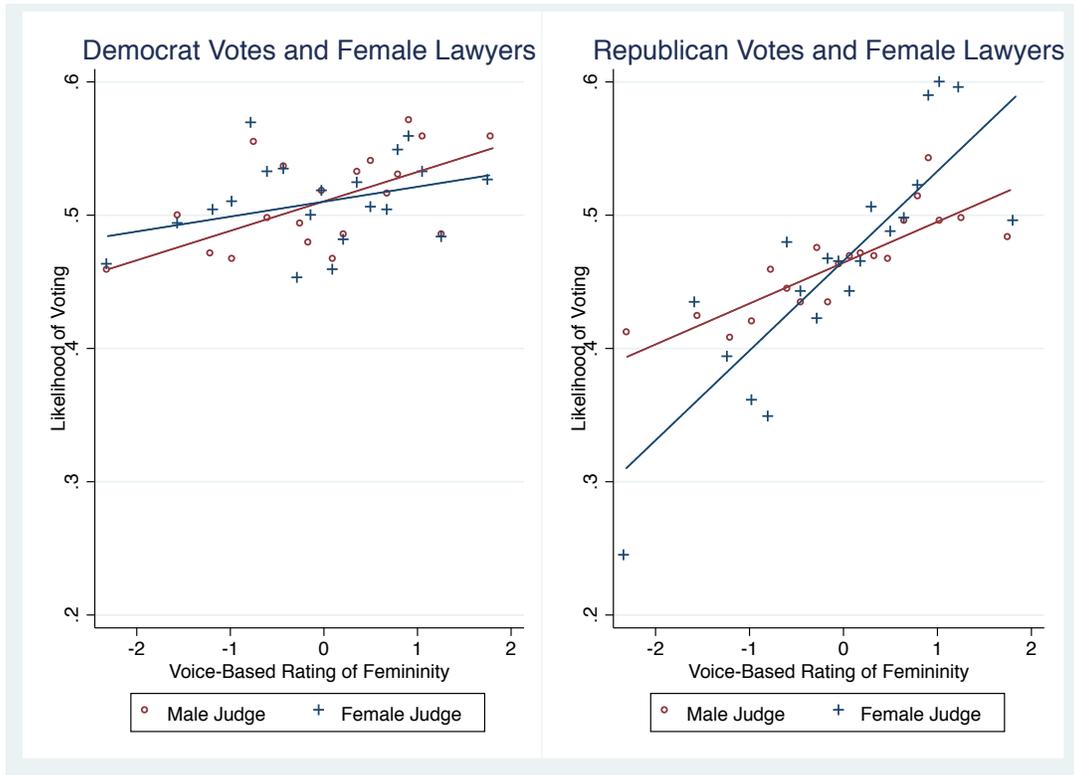
5.4 Why are Votes of Judges Correlating Differentially with Masculine Voices?

There are at least three possible reasons why judges vote differently to perceived masculinity. We consider these three reasons in turn: strategic voting, sincere voting, and homophily.

H_0 : With strategic voting, judges could vote to signal their attitudes towards gender role issues. Democrats want to signal they are progressive and support gender-neutral sounding men. Republicans want to signal their support for traditional values and gender roles: females should sound female, and men and women should have a distinct role. Under strategic voting, we should see no variation across cases.

H_1 : With sincere voting, judges value information more highly from advocates who sound less masculine, since they must be more highly skilled than their more masculine-sounding

Figure 15: Political Party and Response to Femininity ($p < 0.01$)



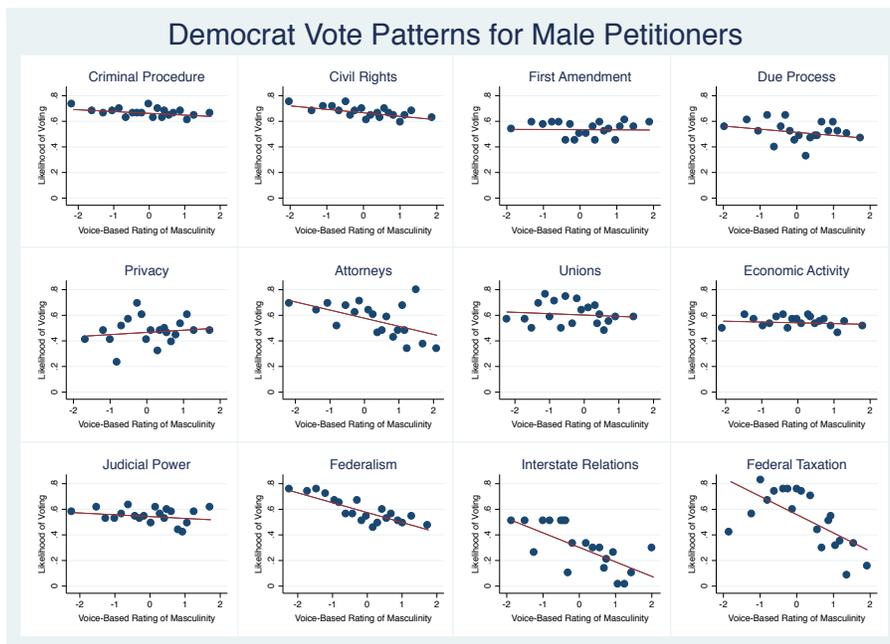
colleagues in order to have reached this level. For both men and women this means not sounding masculine. Under sincere voting, we should see this effect particularly in difficult or technical cases.

H_2 : With homophily, voices perceived as masculine may signal ideology and engender trust. Labor, consumer, and civil rights advocacy groups have traditionally sought to put forth attorneys who share their ideology. Recent studies find that Republican women in Congress look feminine (Carpinella and Johnson 2013) and Republican Senators look “powerful” while Democrats look “warm” (Rule and Ambady 2010), which are potential synonyms for more and less masculine. Under homophily, we should see this effect particularly in ideological cases.

Before proceeding, it is important to note that the scope of the data precludes identifying the exact reason. For instance, we are unable to distinguish behavior that is intentional vs. unintentional or due to signaling vs. preference. Nevertheless, we use the data that we have on cases to explore these three broad classes of reasons.

Examining the raw correlation between Democratic votes and male petitioners in different case categories (Figure 16) shows that the correlation is more salient in the case categories of: attorneys, interstate relations, and federal taxation ($p < 0.05$) (also federalism). With the exception of the last category, these are less ideological areas according to their ranking in

Figure 16: Substantial Variation by Case, Stronger in Less-Ideological Cases



percent of decisions with 5-4 outcomes.⁶

Comparing with Republican vote patterns by case type (Figure 17), we see that both Democrats and Republicans vote for less-masculine lawyers in federal taxation ($p < 0.1$), interstate relations, federalism, and attorneys cases, which would be consistent with sincere voting. However, the slope differs in ideological cases like unions and civil rights, which is consistent with homophily.

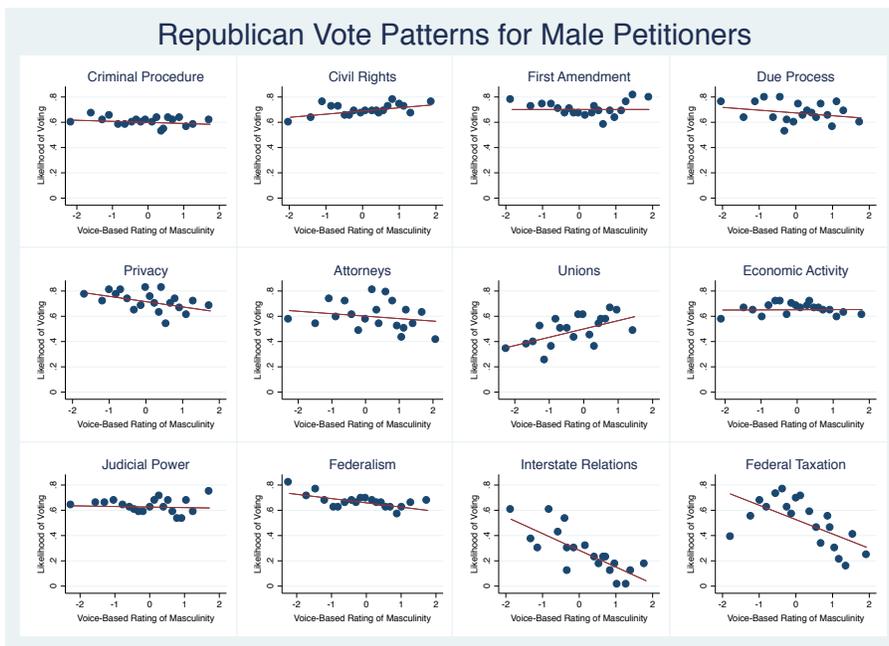
Using a dummy indicator for the less-ideological case categories, we regress:

$$(2) \quad \begin{aligned} \text{judge votes for advocate}_{itw} = & \alpha + \mathbf{masc}'_{itw}\beta_0 + \mathbf{masc}'_{itw} * \text{Dem}_w\beta_1 \\ & + \mathbf{masc}'_{itw} * \text{LessIdeo}_t\beta_2 + \mathbf{x}'_{itw}\gamma + \varepsilon_{itw} \end{aligned}$$

The results suggest that the overall effect can be broken down into ideology for 75% (H_2) of the effect, and information for the remaining 25% (H_1). The reason is that despite the visual agreement in the previous figures, there are actually very few tax and attorney cases. This breakdown needs to be viewed as an approximation and it is unlikely that only one of the three hypotheses explains the entirety of the judicial heterogeneity. Inconsistent with strategic voting is the fact that Republicans do not favor masculine-sounding males. Inconsistent with

⁶We observe that 5-4 decisions occur with the following frequency in these legal areas: unions - 34%, criminal procedure - 31%, federalism - 28%, civil rights - 27%, first amendment - 26%, judicial power - 22%, interstate relations - 22%, due process - 21%, privacy - 21%, attorneys - 21%, economic activity - 12%, and federal taxation - 10%.

Figure 17: Even Republicans Vote like Democrats in Less-Ideological Cases



sincere voting is the fact that Republicans appear not to value this information unless it comes from a female. Inconsistent with homophily is the fact that it would not explain within-lawyer effects, unless the within-lawyer change in voice corresponds to changes in ideology.

Indeed, we can see a somewhat similar pattern for female lawyers: agreement in less-ideological cases that feminine-sounding females win the votes of both Democrats and Republicans, and disagreement in more-ideological cases. Figure 18 presents Republican votes for female advocates by case category. (The smaller number of females entails the omission of one category.) Attorneys ($p < 0.01$) (also federalism, federal taxation, due process, and privacy) cases see a more marked Republican gradient.

For Democrats, we also see attorneys ($p < 0.05$) (also federal taxation) cases having positive gradients (Figure 19).

Regardless of the source of the judicial heterogeneity, lawyers should adjust their voices unless another audience matters, which we explore in the next two sub-sections.

5.5 Is Masculinity of Voice Differentially Correlated with Court Outcomes across Firms?

Thus far, we have observed that masculine-sounding males lose in the Supreme Court, which does not seem to be due to omitted variables at the lawyer or pre-trial level. Part of masculine males losing seems due to a combination of identity and ideology (Democrats but not Republicans vote in favor of less-masculine males) and partly by information (less-masculine males win both Democrat and Republican votes in the less-ideological, more technical areas

Figure 18: Substantial Variation by Case, Stronger in Less-Ideological Cases

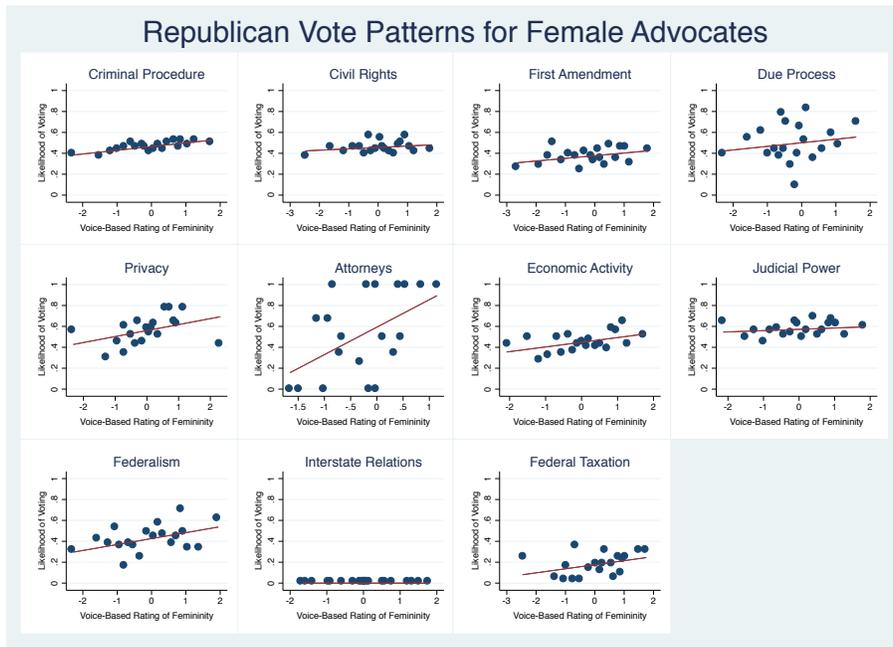


Figure 19: Even Democrats Vote like Republicans in Less-Ideological Cases ($p < 0.05$)

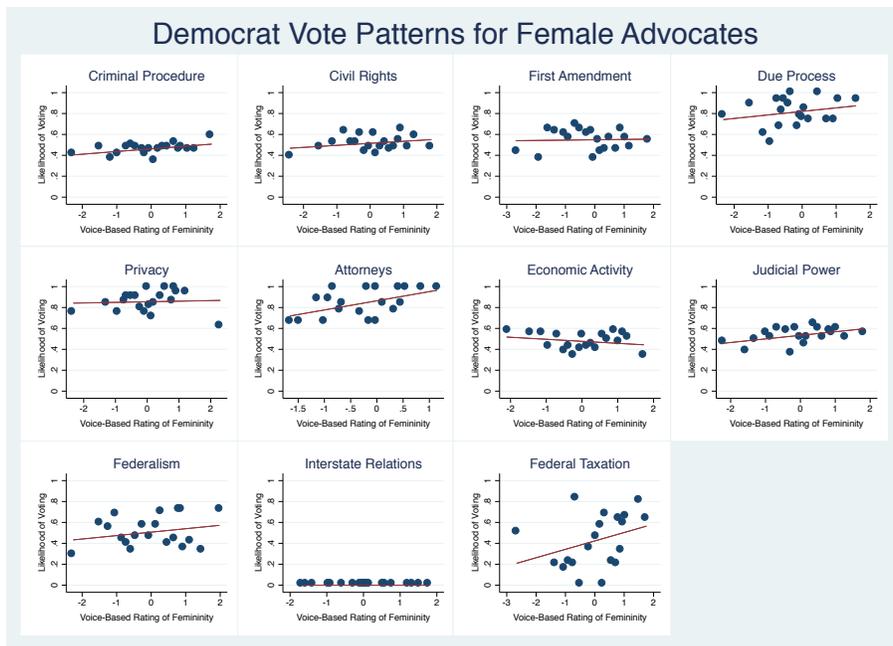


TABLE XIX
PUBLIC VS. PRIVATE FIRMS AND RESPONSE TO MASCULINITY

	Masculine Rating
Public	-0.0480* (0.0287)
Observations	1580

of law, which is consistent with firms choosing lawyers who sound more masculine at the risk of losing cases). We now address the role of firm selection more directly using data on firm characteristics and rater characteristics. In the model, industries that value $d > 0$ more highly, with a greater relative taste for M lawyers than for F lawyers, the price of masculinity increases: they will choose more masculine-sounding lawyers with the consequence of losing more often.

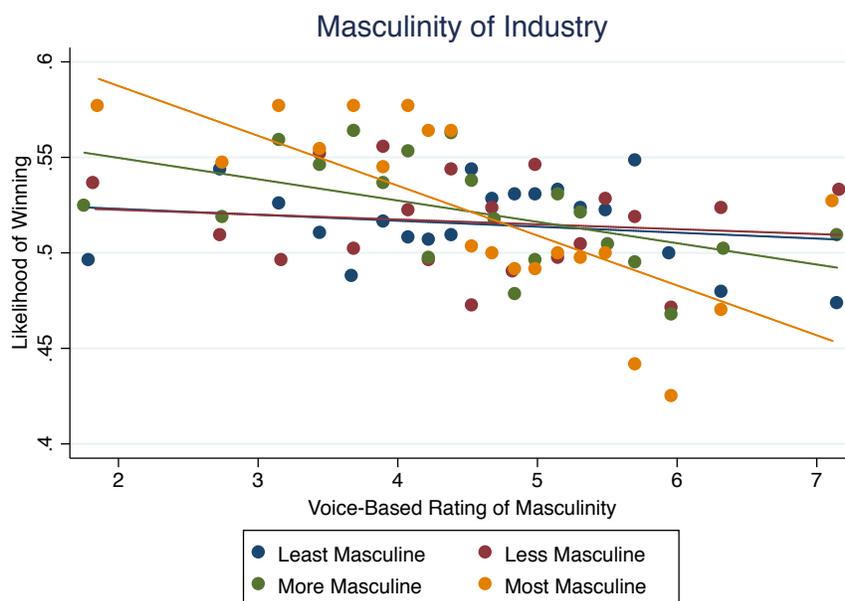
We begin by simply dividing the set of firms into two categories: not-for-profit (e.g., government) vs. private. Prior research finds that, in the private sector, CEOs with deeper voices manage larger companies (Mayew and Venkatachalam 2012), which may lead private firms to prefer masculine-sounding lawyers. Indeed, not-for-profit firms' lawyers are slightly less masculine than private firm lawyers ($p < 0.1$) (Table XIX). We also observe stronger negative correlations between perceived masculinity and court outcomes among private firms and in industries with more masculine voices ($p < 0.05$) (Figure 20), and essentially no correlation between perceived masculinity and win rates for public lawyers advocating on behalf of the government, which renders less likely an explanation based solely on Justices ideologically profiling lawyers.

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 20 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 $d > 0$. 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.

5.6 Whose Perceptions of Masculinity Predict Outcomes?

Regardless of the reason for judges to vote differently according to voice masculinity, lawyers should adjust, unless another audience 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

Figure 20: Masculinity of Industry and Response to Masculinity ($p < 0.05$)



more likely to win? Affirmative answers to the two questions on rater heterogeneity would suggest that law firm employers are the true audience for the lawyers' masculinity.

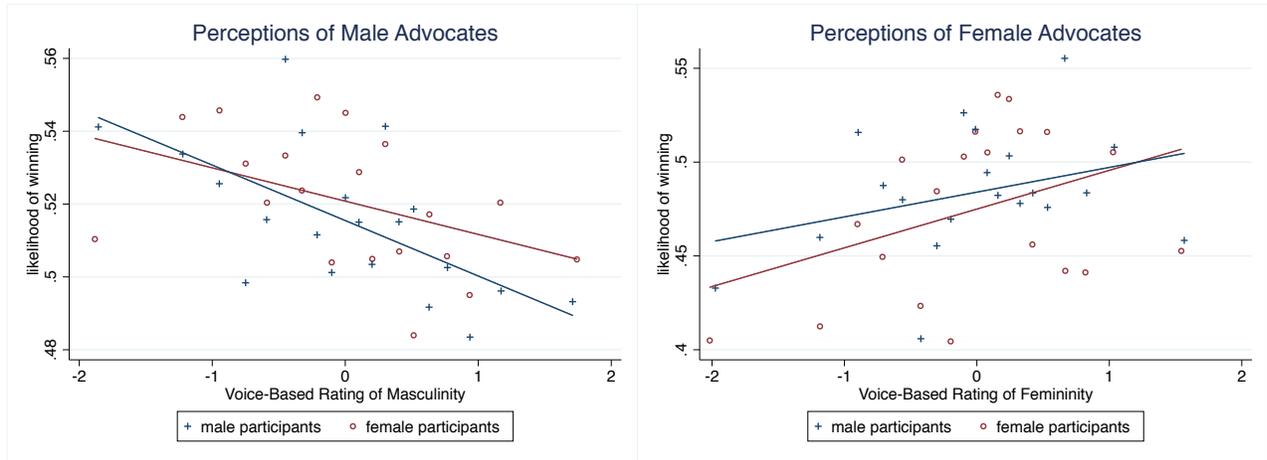
We begin by analyzing the raters' gender in order to match priors. Previous studies find that members of the same sex may use certain criteria to rate each other, which differ from those used to rate members of the opposite sex (Babel et al. 2014). Table A.5 shows that partial correlations among perceived attributes of men and women differ when rated by men or by women. We then investigate other demographic characteristics that play a bigger role than gender.

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 21).

Tables XX 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.⁷ 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

⁷ 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).

Figure 21: Voice-Based Perceptions and Court Outcomes by Advocate and Participant Gender



Notes: Binned scatterplots illustrating the association between voice-based masculinity (femininity) rating and court outcomes when lawyer voices are rated separately by same-sex and cross-sex participants. The figures reflect the models presented in Columns 2, 4, 6 and 8 of Table A.6.

likelihood of winning the case if the advocate is female. These associations are displayed in Figure 9. The results for unmatched gender specification have substantially smaller coefficient estimates and are less statistically distinguishable from 0, which can be seen in Figure 21.

TABLE XX
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

Notes: Using the raw perceptual ratings, 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 percent, 5 percent, and 1 percent levels, respectively.

Next, we observe even greater heterogeneity according to the rater's race and region (Figures 22 and 23). White ($p < 0.05$) and non-Southerners' ($p < 0.05$) perceptions of masculinity are more significantly predictive of court outcomes than other demographic groups' perceptions. 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.

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.

Interestingly, White and non-Southerners' perceptions of femininity are also more predictive of court outcomes, though the difference is less salient than for masculinity (Figure 24).

When we examine raters' perceptions of winning and masculinity, we see that poor individuals ($p < 0.05$) and non-whites ($p < 0.05$) are significantly less likely to correlate masculinity

Figure 22: White Raters' Perceptions of Masculinity Predicted Court Outcomes ($p < 0.05$)

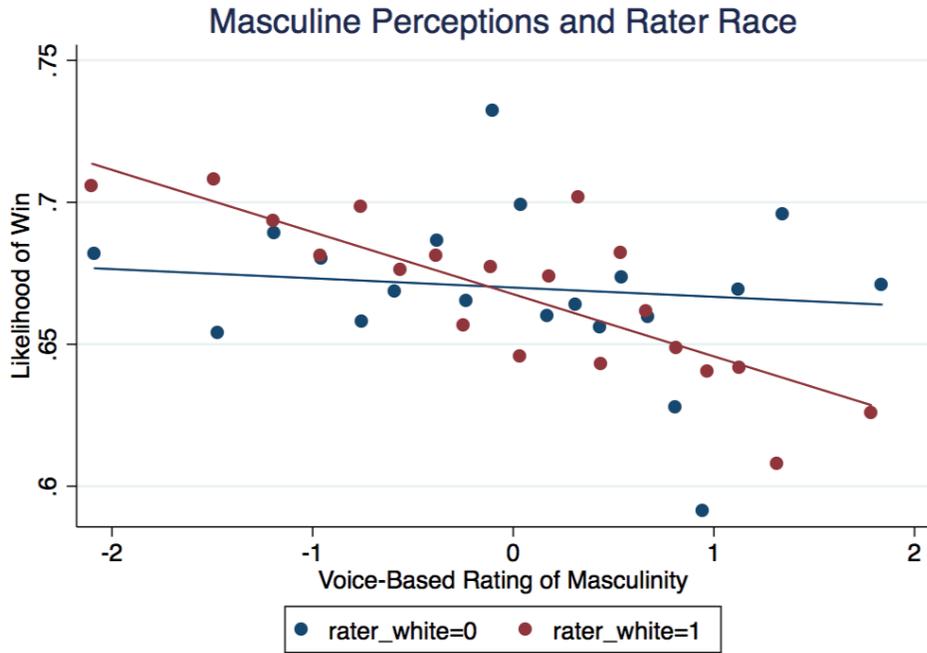
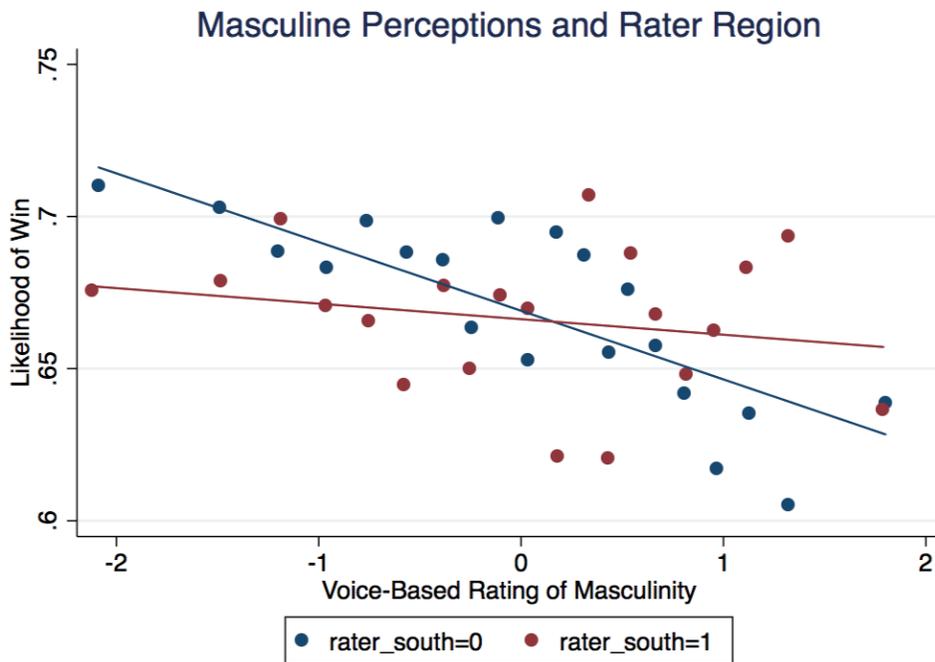


Figure 23: Non-Southerners' Perceptions of Masculinity Predicted Court Outcomes ($p < 0.05$)



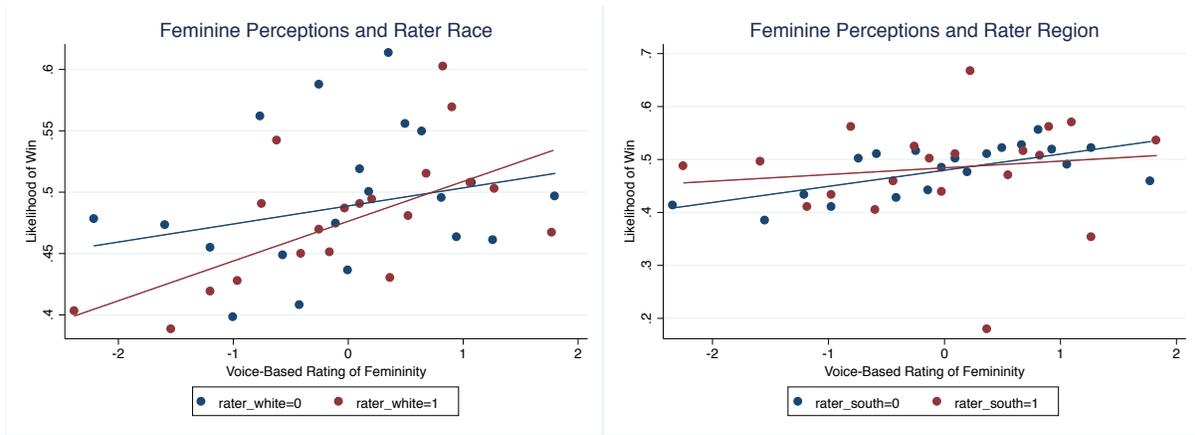


Figure 24: White non-Southerners’ Perceptions of Femininity also Predict Court Outcomes ($p > 0.1$)

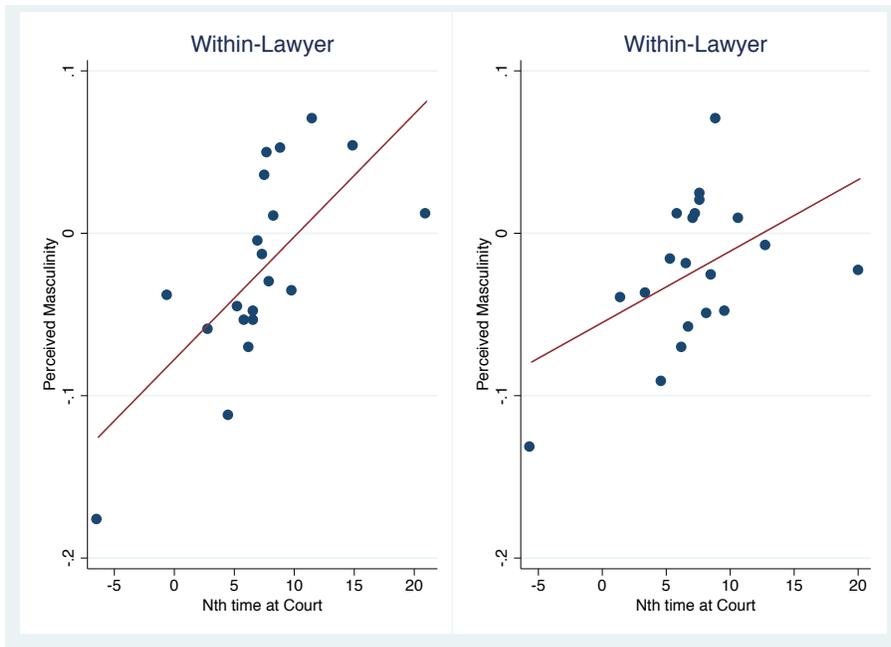
with winning. Poor individuals ($p < 0.05$), males ($p < 0.1$), and the less-educated ($p < 0.05$), are significantly less likely to correlate femininity with winning. This is notable for three reasons. First, if these individuals are more likely to comprise the Democratic Party, the findings are consistent with judicial heterogeneity. Second, 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. It also suggests that more diversity in the set of firm selectors can be financially profitable for the firm. Third, perceptions of masculinity are subjective and not everyone perceives the same voice as being masculine. Different stereotypes persist, which may drive behavioral responses. In the next section, we ask whether individuals can be de-biased.

6 De-biasing Experiment

6.1 What are Lawyers Doing Over Time?

Voice appears to be mutable, and perceived masculinity is negatively correlated with winning in the Supreme Court. We begin by exploring whether lawyers are learning 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 25 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 26 presents the cross-sectional variation without lawyer fixed effects. Note that more-experienced lawyers generally have a less-masculine voice, so the within-lawyer analysis (which biases towards selecting lawyers who repeatedly win) will be biased towards finding less-masculine voices. In fact, we find the opposite.

Figure 25: Masculine Ratings and Order of Appearance



Notes: Within-lawyer graph is residualized on lawyer fixed effects, so negative means early years of experienced advocates.

Figure 26: Cross-sectional variation

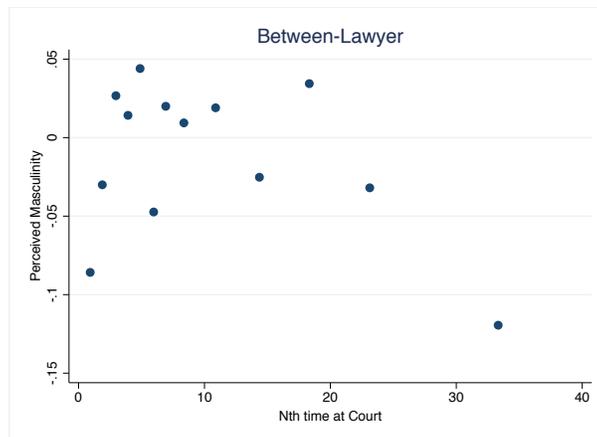


TABLE XXI
ARE LAWYERS LEARNING?

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

Table XXI rejects the hypothesis that losses are significantly associated with decreases in perceived masculinity. Column 1 repeats the negative association between masculinity rating and court outcome. Column 2 presents the correlation between a loss in the previous 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, which biases upwards the correlation in levels. Whatever drives the increase in perceived masculinity (it may be “covering” to make one sound more masculine in order to be hired again), lawyers do not seem to be taking actual win rates into account.

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 ($I_j = 0$).

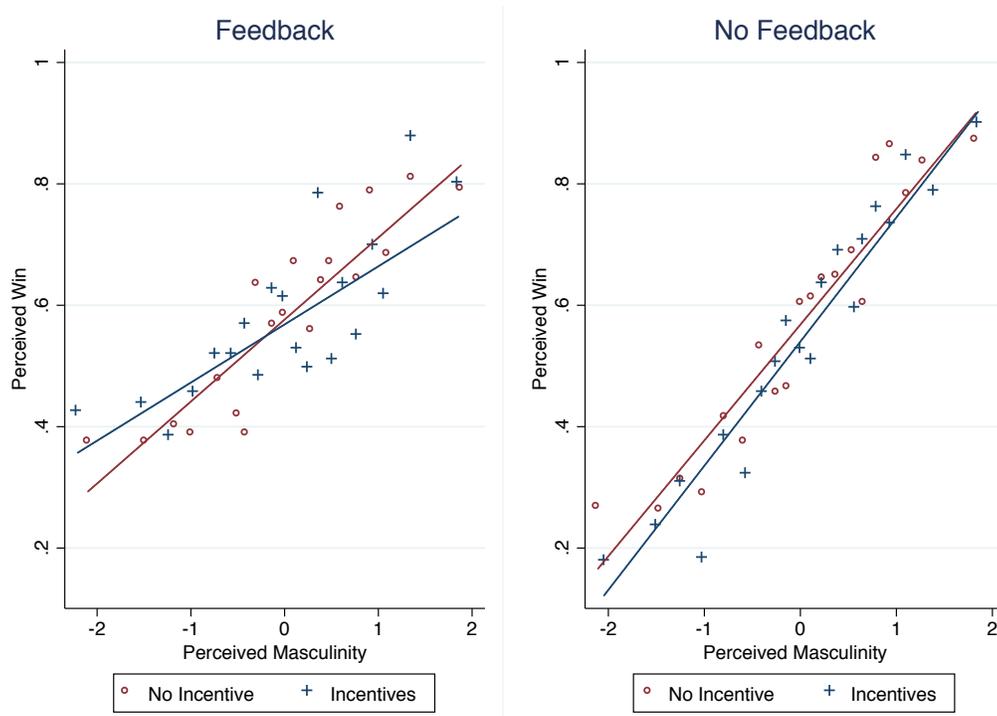
The estimation equation takes the following form:

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

where $V_i \in \{H, L\}$ is advocate i ’s voice type and $T_J \in \{[F_j], [I_j]\}$ is a treatment vector (2x2 treatment). The incentives treatment entailed an additional \$0.5 for raters who get two-thirds or more of the outcomes correct.

Figure 4 shows that information reduces 40% of the correlation between perceived masculinity and perceived win ($p < 0.05$). In fact, all ratings’ positive correlations with perceived

Figure 27: Incentives ($p < 0.05$) with Feedback



wins are affected by information but not incentives. Recall that none of the other ratings predicted actual wins, but raters generally associated all the attributes with a greater likelihood of winning.

When we fully interact the treatments in Figure 27, we observe that incentives further reduce the correlation. Recall that incentives to choose correctly erode the effect of taste on choices ($\pi_F - \pi_M > \frac{d}{\alpha}$). Providing incentives should bring $d \rightarrow 0$. The existence of the effect ($p < 0.05$, 30% of association with feedback) also means $d > 0$ (else $\frac{d}{\alpha}$ would be 0 regardless of the size of α).

7 Explanation of Variance

7.1 How Does Predictive Power Compare with Previous Best Multivariate Models?

We consider predictions using random forest tree methods (Katz et al. 2014; Guimerà and 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-trial characteristics (which collectively received roughly 25% of the importance weight) as well as court and judge historical trends specific to issue, parties,

and lower courts (which collectively received roughly 75% of the importance weight). Feature weights are presented in Table XXII, which comes from Katz et al. (2014).

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. Does perceived masculinity have an explanatory effect above and beyond this best predictor? If so, how much R-square improvement?

Table XXIII shows that the predicted vote correlates strongly with the actual judge vote ($p < 0.01$) in Column 1. Column 2 examines the separate role of perceived masculinity ($p < 0.05$). 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-trial case characteristics, as seen in Section 5, and also orthogonal to court and judge 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.

Why does perceived masculinity do so well relative to the best prediction model? Notably, Katz et al. (2014) state that their model performs 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). In fact, in close cases affirming the lower court, the model predicts the outcome with only 25% accuracy.⁸ We showed earlier in the paper that perceived masculinity is found to be relevant with the swing voter, which is where the random forest model may do poorly.

7.2 How Much Do Perceptions Explain beyond Objective Acoustic Data?

We now analyze acoustic features of the clips using a set of phonetic traits previously 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. Some analysis documenting the measurements of these traits in the Supreme Court oral arguments is provided elsewhere (Yu et al. 2015). Note that some of these capture pitch range, which is hypothesized to be correlated with ideology (Reeve 2015).

Table XXIV 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

⁸Figure 6 in Katz et al.

TABLE XXII
FINAL FEATURE WEIGHTS (KATZ ET AL. 2014)

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

TABLE XXIII
BEST PREDICTION AND PERCEIVED MASCULINITY

	Judge Votes for Lawyer					
Predicted Vote	0.257***		0.258***	0.250***		0.248***
from Random Forest	(0.0486)		(0.0487)	(0.0485)		(0.0489)
Masculine		-0.0223**	-0.0207**		-0.0852**	-0.0780**
		(0.0101)	(0.0101)		(0.0359)	(0.0361)
Cluster		Lawyer and Judge				
Collapsed	No	No	No	Yes	Yes	Yes
Observations	26447	26391	26391	1229	1229	1229
R-squared	0.061	0.002	0.063	0.058	0.008	0.064
Sample: Male Petitioners, Democrat Judges						

controls. The association with perceived masculinity is basically invariant for two reasons. First, perceived masculinity is correlated with many acoustic variables (Puts et al. 2011). 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 and Schuckers 2007).

8 Conclusion

Mutable characteristics are relevant to current policy debates in the United States and the EU. While economic theorists have modeled such characteristics, mutable characteristics have received less attention in the empirical literature (Bertrand and Duflo 2016).

We find that snap judgments of advocate voices predict SCOTUS outcomes based on perceived masculinity. Males who are perceived as less masculine are more likely to win. The petitioner (the first lawyer to argue) is the main driver. Among repeat petitioners, 30% of the variance can be explained by within-male-lawyer differences, and 70% by between-male-lawyer variation. As for why voices predict SCOTUS outcomes, we find that Democratic judges prefer less-masculine-sounding males and Republican judges prefer more-feminine-sounding females, and that there is a stronger effect in private firms and in industries characterized by more-masculine voices. This is consistent with linguistic profiling by judges and firms, but further research is needed.

We conduct a de-biasing experiment and find that information reduces 40% of the correlation between perceived masculinity and perceived win, and that incentives reduce another 20% when there is information. These findings separately distinguish prejudice from statisti-

TABLE XXIV
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

cal discrimination. Finally, perceived masculinity explains additional variance beyond and is orthogonal to the best the prediction models.

Judges responding differently to masculine-sounding voices renders unlikely an explanation where the entire association is due to perceived masculinity proxying for case weakness. Among male petitioners, being above or below median in masculine rating is equivalent to 7 percentage points in the likelihood of winning. The magnitude of the effects and the stakes of Supreme Court outcomes (Crawford and Malik 2016) also renders unlikely a compensating salary differential whereby less-masculine-sounding males are paid for their higher win rates. Lawyers become more masculine with experience, while age and experience are cross-sectionally associated with less-masculine voices, and lawyers are not making voices less masculine following a loss, also rendering unlikely an explanation where lawyers and firms learn over time.

Our results also contribute to a growing literature on the relevance of extraneous factors in courtrooms. Although it is widely assumed that judicial behavior is governed by legal doctrine (Kornhauser 1999), and that judges strictly hew to legal doctrine and court precedent in their rulings, judges' decisions can be affected by their policy preferences (Cameron 1993) and self-interest (Posner 1973). The associations we find are comparable to the effects of other external factors that have been shown to influence judicial behavior. Most analyses of judicial behavior document disparities across judges; judges' demographic background characteristics, such as gender, race, and in particular, the party of the appointing president predict judges' decisions (Peresie 2005; Sunstein et al. 2006). However, a handful of studies examine within-judge variation, such as the effects of emotional shocks on disposition length stemming from

unexpected football game losses (Eren and Mocan 2016). Danziger et al. (2011) detect time-of-day patterns in Israeli judges' parole decisions. Concretely, they show that parole approval rates drop with the time from the judges' last meal. In other work, asylum judges are 2 percentage points more likely to deny asylum to refugees if their previous decision granted asylum. Likewise, asylum judges are roughly 2 percentage points more likely to grant asylum on the day after a home-city Sunday football game win instead of a loss. U.S. District judges are 0.3 percentage points less likely to assign any prison length in criminal sentencing cases after a home-city football game win instead of a loss. And they assign 8% longer prison sentences when the first letter of the defendant's name matches the first letter of their own name. U.S. Presidential elections affect judicial decisions. And the explanatory power of extraneous factors persist after employing the best prediction models of asylum and sentencing decisions. See e.g., Berdejo and Chen 2016; Barry et al. 2016; Chen and Eigel 2016; Chen and Prescott 2016; Chen et al. 2016d; 2014; 2016b; 2016. In an analysis of the Supreme Court, Chen et al. (2016a) examine audio convergence during oral arguments and find that mimicry (phonetic accommodation) explains additional variation beyond the best predictive models.

There are a variety of directions for future research. One can employ all data back to 1955, more lawyer covariates (faces, campaign donations, which Justice they clerked for, if any), more acoustic data, data on Supreme Court clerks (and their ideology) at the time of the decision, and the subsequent behavior of Justices (number of questions, interruptions). In a separate paper, Chen et al. (2016c) collects 70% of oral advocates faces from 1998-2012 and analyzes them using a face database to predict perceptions of face. The paper finds that these features predict decisions. It also uses all audio features of oral advocates from 1955-present and finds that these features also predict decisions beyond the current best prediction models. Dietrich et al. (2016) also examine audio features from a dataset beginning in the 1980s and finds that male lawyers with higher pitch are more likely to win, which is consistent with less masculine-sounding males winning (but it does not elicit human ratings nor does it restrict to the identical introductory sentences of lawyers). Besides extending the data backwards in time or extending it to collect more variables, other directions for future research include experiments that manipulate the presence of text or appearance of faces to separate the predictive power of audio or facial features from the content of the argument. Future experiments could manipulate the voices to sound more masculine, or present both petitioner and respondent clips from a case for raters to evaluate. Ideological mechanisms can be explored with data measuring the audio—the perceived masculinity/femininity—of male and female Democrats and Republicans. The de-biasing experiment can measure willingness to bet that a voice will win in order to elicit a revealed preference measure of confidence in their beliefs.

Affirmative action typically targets immutable characteristics, but some have asked whether society should help all women past the glass ceiling (rather than only women who are mas-

culine) past the glass ceiling: it has been argued, “[o]nce that barrier is broken, there may be a little less pressure to keep appointing people from that previous excluded category” (Zweigenhaft 1987; Zweigenhaft and Domhoff 1998). We refrain from policy inference from these findings for two reasons. First, the welfare consequences of any policy change are unknown, since they are determined by decisions about what factors to weigh in the social welfare function. For example, whether ‘fit’ should be a hiring criterion depends on one’s attitude towards fit. Second, for any policy proposal, one also needs to consider the general equilibrium consequences, which would require the random assignment of laws (Abramowicz et al. 2011; Belloni et al. 2012; Chen and Yeh 2014a,b; Chen et al. 2014a). However, our results do highlight the relevance of assimilation and integration, which relates to EU refugee policy and *laïcité* which is at least partly based on mutable characteristics. Our results also highlight the mechanism that social stereotypes and mental stereotypes (snap judgments) associated with speakers’ words can be relevant for the behavioral response to the speaker, as per the theory and philosophy of language (Putnam 1973) and recent formal models of economic decision making (Bordalo et al. 2014).

References

- Abramowicz, Michael, Ian Ayres, and Yair Listokin**, “Randomizing Law,” *University of Pennsylvania Law Review*, March 2011, 159 (4), 929–1005.
- Adler, Andy and Michael E Schuckers**, “Comparing human and automatic face recognition performance,” *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 2007, 37 (5), 1248–1255.
- Alesina, Alberto and Eliana La Ferrara**, “Preferences for redistribution in the land of opportunities,” *Journal of public Economics*, 2005, 89 (5), 897–931.
- Ambady, Nalini and Robert Rosenthal**, “Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis,” *Psychological Bulletin*, 1992, 111, 256–274.
- Apple, William, Lynn A Streeter, and Robert M Krauss**, “Effects of pitch and speech rate on personal attributions,” *Journal of Personality and Social Psychology*, 1979, 37 (5), 715.
- Arrow, Kenneth J**, *The Theory of Discrimination* 1973.
- Atkinson, Max**, *Our Masters’ Voices: The language and body language of politics*, London; New York: Methuen & Co. Ltd., 1984.
- Austen-Smith, David and Roland G Fryer Jr**, “An Economic Analysis of ‘Acting White,’” *Quarterly Journal of Economics*, 2005, pp. 551–583.
- Babel, Molly, Grant McGuire, and Joseph King**, “Towards a more nuanced view of vocal attractiveness,” *PLOS ONE*, 2014, 9 (2), e88616. doi:10.1371/journal.pone.0088616.
- , ———, and **Santa Cruz**, “Perceived vocal attractiveness across dialects is similar but not uniform,” in “Proceedings of Interspeech” 2013, pp. 426–430.
- Baker, Scott and Claudio Mezzetti**, “A Theory of Rational Jurisprudence,” *Journal of Political Economy*, 2012, 120 (3), 513–551.
- Balkin, Jack M**, *Living originalism*, Harvard University Press, 2011.
- Ballew, Charles C and Alexander Todorov**, “Predicting political elections from rapid and unreflective face judgments,” *Proceedings of the National Academy of Sciences*, 2007, 104 (46), 17948–17953.
- Barry, Nora, Laura Buchanan, Evelina Bakhturina, and Daniel L. Chen**, “Events Unrelated to Crime Predict Criminal Sentence Length,” Technical Report 2016.
- Becker, Gary S**, *Economics of Discrimination*, University of Chicago Press, 1957.
- Beim, Deborah, Alexander V. Hirsch, and Jonathan P. Kastellec**, “Signaling and Counter-Signaling in the Judicial Hierarchy: An Empirical Analysis of En Banc Review,” *American Journal of Political Science*, 2016, 60 (2), 490–508.
- Belloni, Alex, Daniel L. Chen, Victor Chernozhukov, and Chris Hansen**, “Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain,” *Econometrica*, November 2012, 80 (6), 2369–2429.
- Benjamin, Daniel J, Jesse M Shapiro et al.**, “Thin-Slice Forecasts of Gubernatorial Elections,” *The Review of Economics and Statistics*, 2009, 91 (3), 523–536.
- Berdejo, Carlos and Daniel L. Chen**, “Electoral Cycles Among U.S. Courts of Appeals Judges,” Technical Report 2016.
- Berggren, Niclas, Henrik Jordahl, and Panu Poutvaara**, “The looks of a winner: Beauty and electoral success,” *Journal of Public Economics*, 2010, 94 (1), 8–15.
- Bertrand, Marianne and Esther Duflo**, “Field Experiments on Discrimination,” Technical Report, National Bureau of Economic Research 2016.
- and **Sendhil Mullainathan**, “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination,” *The American Economic Review*, 2004, 94, 991–1013.
- , **Esther Duflo, and Sendhil Mullainathan**, “How Much Should We Trust Differences-in-Differences Estimates?,” *Quarterly Journal of Economics*, 2004, 119 (1), 249–275.
- Biddle, Jeff E and Daniel S Hamermesh**, “Beauty, productivity, and discrimination: Lawyers’ looks and lucre,” *Journal of Labor Economics*, 1998, 16 (1), 172–201.
- Biskupic, Joan, Janet Roberts, and John Shiffman**, “At America’s court of last resort, a handful of lawyers now dominates the docket,” 2014.
- Black, Ryan C, Sarah Ann Roberts, Timothy R Johnson, and Jerry Goldman**, “Emotions, oral arguments, and supreme court decision making,” *Journal of Politics*, 2011, 73 (2), 572–581.
- Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer**, “Stereotypes,” Technical Report, National Bureau of Economic Research 2014.
- Brewer, Marilyn B**, “A dual process model of impression formation,” *Advances in social cognition*, 1988, 1,

- Buller, DB, JK Burgoon, and WG Woodall**, *Nonverbal communications: The unspoken dialogue*, New York: McGraw-Hill, 1996.
- Butler, Judith**, *Gender trouble: Feminism and the subversion of identity*, New York, NY: Routledge, 1990.
- Cameron, Charles M.**, “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 Rochester, NY 1993.
- Campbell-Kibler, Kathryn**, “Sociolinguistics and perception,” *Language and Linguistics Compass*, 2010, 4 (6), 377–389.
- Carpinella, Colleen M and Kerri L Johnson**, “Appearance-based politics: Sex-typed facial cues communicate political party affiliation,” *Journal of Experimental Social Psychology*, 2013, 49 (1), 156–160.
- Case, Mary Anne C**, “Disaggregating gender from sex and sexual orientation: The effeminate man in the law and feminist jurisprudence,” *Yale Law Journal*, 1995, 105 (1), 1.
- Charles, Kerwin Kofi and Jonathan Guryan**, “Prejudice and Wages: An Empirical Assessment of Becker’s The Economics of Discrimination,” *Journal of Political Economy*, 2008, 116, 773–809.
- Chen, Daniel L.**, “Gender Violence and the Price of Virginity: Theory and Evidence of Incomplete Marriage Contracts,” Working Paper, University of Chicago, Mimeo 2004.
- , “Club goods and group identity: evidence from islamic resurgence during the indonesian financial crisis,” *Journal of Political Economy*, 2010, 118 (2), 300–354.
- , “The Deterrent Effect of the Death Penalty? Evidence from British Commutations During World War I,” Working Paper, ETH Zurich May 2013.
- , “Priming Ideology: Why Presidential Elections Affect U.S. Courts of Appeals Judges,” Technical Report 2016.
- and **Holger Spamann**, “This Morning’s Breakfast, Last Night’s Game: Detecting Extraneous Factors in Judging,” Working Paper, ETH Zurich 2014.
- and **J. J. Prescott**, “Implicit Egoism in Courtrooms: First Initial Name Effects with Randomly Assigned Defendants,” Technical Report 2016.
- and **Jasmin Sethi**, “Insiders, Outsiders, and Involuntary Unemployment: Sexual Harassment Exacerbates Gender Inequality,” Working Paper, University of Chicago October 2011.
- and **Jess Eagel**, “Can Machine Learning Help Predict the Outcome of Asylum Adjudications?,” Technical Report 2016.
- and **Martin Schonger**, “Social Preferences or Sacred Values? Theory and Evidence of Deontological Motivations,” Working Paper, ETH Zurich, Mimeo August 2013.
- and ———, “A Theory of Experiments: Invariance of Equilibrium to the Strategy Method of Elicitation and Implications for Social Preferences,” Working Paper, ETH Zurich January 2015.
- and **Susan Yeh**, “Government Expropriation Increases Growth and Inequality: Evidence from Eminent Domain,” Working Paper, ETH Zurich and George Mason University December 2014.
- and ———, “How Do Rights Revolutions Occur? Free Speech and the First Amendment,” Working Paper, ETH Zurich December 2014.
- , **Damian Kozbur, and Alan Yu**, “Mimicry: Phonetic Accommodation Predicts U.S. Supreme Court Votes,” Working Paper, ETH Zurich 2016.
- , **Matt Dunn, Rafael Garcia Cano Da Costa, Ben Jakubowki, and Levent Sagun**, “Early Predictability of Asylum Court Decisions,” Technical Report 2016.
- , **Moti Michaeli, and Daniel Spiro**, “Ideological Perfectionists under Pressure: Theory and Evidence from the U.S. Courts of Appeals,” Working Paper, ETH Zurich January 2015.
- , **Motwani Kumar, Vishal Motwani, and Philip Yeres**, “Is Justice Really Blind? And Is It Also Deaf?,” Technical Report 2016.
- , **Tobias J. Moskowitz, and Kelly Shue**, “Decision-Making Under the Gambler’s Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires,” Technical Report 2016.
- , **Vardges Levonyan, and Susan Yeh**, “Policies Affect Preferences: Evidence from Random Variation in Abortion Jurisprudence,” Working Paper, ETH Zurich May 2014.
- , ———, **S. Eric Reinhart, and Glen B. Taksler**, “Mandatory Disclosure: Theory and Evidence from Industry-Physician Relationships,” Working Paper 2014. Mimeo.
- Chen, Daniel, Yosh Halberstam, and Alan C. L. Yu**, “Perceived Masculinity Predicts United States Supreme Court Outcomes,” Technical Report 2016.
- Crawford, Jonathan and Naureen Malik**, “Billions of Dollars at Stake in Supreme Court Power Market

- Fight,” 2016.
- Danescu-Niculescu-Mizil, Cristian, Lillian Lee, Bo Pang, and Jon Kleinberg**, “Echoes of Power: Language Effects and Power Differences in Social Interaction,” in “Proceedings of the 21st International Conference on World Wide Web” WWW ’12 ACM New York, NY, USA 2012, pp. 699–708.
- Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso**, “Extraneous factors in judicial decisions,” *Proceedings of the National Academy of Sciences of the United States of America*, 2011, *108* (17), 6889–6892.
- Deaton, Angus**, “Instruments, Randomization, and Learning about Development,” *Journal of Economic Literature*, September 2010, *48* (2), 424–455.
- Dietrich, Bryce J., Ryan D. Enos, and Maya Sen**, “Emotional Arousal Predicts Voting on the Supreme Court,” Technical Report 2016.
- Eckert, Penelope**, “Variation and the indexical field,” *Journal of Sociolinguistics*, 2008, *12* (4), 453–476.
- and **Sally McConnell-Ginet**, *Language and gender*, Cambridge, UK: Cambridge University Press, 2003.
- Edwards, Harry T. and Michael A. Livermore**, “Pitfalls of Empirical Studies That Attempt to Understand the Factors Affecting Appellate Decisionmaking,” *Duke Law Journal*, 2008, *58* (8), 1895–1989.
- Epstein, Lee, William M Landes, and Richard A Posner**, “Inferring the winning party in the Supreme Court from the pattern of questioning at oral argument,” *The Journal of Legal Studies*, 2010, *39* (2), 433–467.
- Eren, Ozkan and Naci Mocan**, “Emotional Judges and Unlucky Juveniles,” Working Paper 2016.
- Feinberg, David R, Benedict C Jones, Lisa M DeBruine, Fhionna R Moore, Miriam J Law Smith, R Elisabeth Cornwell, Bernard P Tiddeman, Lynda G Boothroyd, and David I Perrett**, “The voice and face of woman: One ornament that signals quality?,” *Evolution and Human Behavior*, 2005, *26* (5), 398–408.
- Fisher, Jeffrey L**, “A clinic’s place in the Supreme Court bar,” *Stanford Law Review*, 2013, *65* (137).
- Fitch, W Tecumseh**, “Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques,” *The Journal of the Acoustical Society of America*, 1997, *102* (2), 1213–1222.
- Fryer, Roland G. and Paul Torelli**, “An Empirical Analysis of Acting White,” *Journal of Public Economics*, 2010, *94*, 380–396.
- Gennaioli, Nicola and Andrei Shleifer**, “The Evolution of Common Law,” *The Journal of Political Economy*, 2007, *115* (1), 43–68.
- Giles, Howard and Peter Powesland**, *Accommodation theory*, Springer, 1997.
- Goffman, Erving**, *Stigma: Notes on the management of spoiled identity*, NJ: Prentice-Hall, 1963.
- Grogger, Jeffrey**, “Speech Patterns and Racial Wage Inequality,” *Journal of Human Resources*, 2011, *46* (1), 1–25.
- Guimerà, Roger and Marta Sales-Pardo**, “Justice blocks and predictability of us supreme court votes,” *PloS one*, 2011, *6* (11), e27188.
- Hamermesh, Daniel S and Jeff E Biddle**, “Beauty and the Labor Market,” *The American Economic Review*, 1994, *84* (5), 1174–1194.
- Heckman, James J and Peter Siegelman**, *The Urban Institute audit studies: Their methods and findings* Clear and Convincing Evidence: Measurement of Discrimination in America, Urban Institute Press, 1993.
- Hersch, Joni**, “Compensating Differentials for Sexual Harassment,” *The American Economic Review Papers and Proceedings*, 2011, *101* (3), 630–634.
- Hodges-Simeon, Carolyn R., Steven J. C. Gaulin, and David A. Puts**, “Different Vocal Parameters Predict Perceptions of Dominance and Attractiveness,” *Human Nature*, 2010, *21*, 406–427.
- Jackson, Robert H**, “Advocacy Before the United States Supreme Court,” *Cornell Law Review*, 1951, *37*, 1.
- Johnson, Timothy R., James F. Spriggs, and Paul J. Wahlbeck**, “The influence of oral arguments on the US Supreme Court,” *American Political Science Review*, 2006, *100* (1), 99–113.
- Jr, Stanford W Gregory and Timothy J Gallagher**, “Spectral analysis of candidates’ nonverbal vocal communication: Predicting US presidential election outcomes,” *Social Psychology Quarterly*, 2002, pp. 298–308.
- Kahneman, Daniel**, *Thinking, Fast and Slow*, Farrar, Straus and Giroux, 2011.
- Katz, Daniel Martin, Michael J Bommarito, Tyler Soellinger, and James Ming Chen**, “Law on the Market? Evaluating the Securities Market Impact of Supreme Court Decisions,” *arXiv preprint arXiv:1508.05751*, 2015.
- , **Michael James Bommarito, and Josh Blackman**, “Predicting the behavior of the Supreme Court of the United States: A general approach,” *Available at SSRN 2463244*, 2014.
- Kessler, Suzanne and Wendy McKenna**, *Gender: An ethnomethodological approach*, John Wiley and Sons., 1978.

- Kiesling, Scott**, “Men, Masculinities, and Language,” *Language and Linguistic Compass*, 2007, 1 (7), 653–673.
- Kinzler, Katherine D, Emmanuel Dupoux, and Elizabeth S Spelke**, “The native language of social cognition,” *Proceedings of the National Academy of Sciences*, 2007, 104 (30), 12577–12580.
- Klofstad, Casey A., Rindy C. Anderson, and Susan Peters**, “Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women,” *Proceedings of the Royal Society of London B: Biological Sciences*, 2012, 279 (1738), 2698–2704.
- Knowles, John, Nicola Persico, and Petra Todd**, “Racial Bias in Motor Vehicle Searches: Theory and Evidence,” *Journal of Political Economy*, 2001, 109, 203–229.
- Kornhauser, Lewis A.**, “Judicial Organization and Administration,” in Chris William Sanchirico, ed., *Encyclopedia of Law and Economics*, Vol. 5 1999, chapter 11, pp. 27–44.
- Kramer, Michael**, “Saying Bye-Bye to the Wimp Factor,” 1987.
- Künzel, Hermann J.**, “How well does average fundamental frequency correlate with speaker height and weight?,” *Phonetica*, 1989, 46 (1-3), 117–125.
- Lang, Kevin**, “A Language Theory of Discrimination,” *Quarterly Journal of Economics*, May 1986, 101 (2), 363–382.
- Lazarus, Richard James**, “Advocacy matters before and within the Supreme Court: Transforming the Court by transforming the bar,” *Georgetown Law Journal*, 2008, 96.
- Lazear, Edward P.**, “Personnel economics: Past lessons and future directions,” Technical Report, National Bureau of Economic Research 1999.
- List, John A.**, “The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field,” *Quarterly Journal of Economics*, 2004, 119, 49–89.
- Locke, John L.**, *Duels and Duets: Why men and women talk so differently*, New York: Cambridge University Press, 2011.
- Martin, Andrew D. and Kevin M. Quinn**, “Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the US Supreme Court,” *Political Analysis*, 2002, 10, 134–153.
- Martin, Andrew D, Kevin M Quinn, Theodore W Ruger, and Pauline T Kim**, “Competing approaches to predicting supreme court decision making,” *Perspectives on Politics*, 2004, 2 (04), 761–767.
- Mauro, Paolo**, “Corruption and growth,” *Quarterly Journal of Economics*, 1995, pp. 681–712.
- Mayew, William J. and Mohan Venkatachalam**, “The Power of Voice: Managerial Affective States and Future Firm Performance,” *Journal of Finance*, 2012, 67 (1), 1–43.
- McAleer, Phil, Alexander Todorov, and Pascal Belin**, “How do you say hello? Personality impressions from brief novel voices,” *PLoS ONE*, 2014, 9 (3), e90779.
- McAtee, Andrea and Kevin T McGuire**, “Lawyers, Justices, and Issue Salience: When and How Do Legal Arguments Affect the US Supreme Court?,” *Law & Society Review*, 2007, 41 (2), 259–278.
- Mobius, Markus M. and Tanya S. Rosenblat**, “Why Beauty Matters,” *American Economic Review*, 2006, 96 (1), 222–235.
- Morawetz, Nancy**, “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*, 2011, 86, 131.
- Mueller-Smith, Michael**, “Discrimination With Concealable Characteristics: Evidence and Application to Sexual Orientation in the United States,” Technical Report, Columbia 2013.
- Nallon, Steve**, “The Thatcher Years - Voice Analysis of Margaret Thatcher for US TV,” 2014.
- Nass, Clifford and Kwan Min Lee**, “Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction, and consistency-attraction,” *Journal of Experimental Psychology: Applied*, 2001, 7, 171–181.
- Neumark, David**, “Experimental Research on Labor Market Discrimination,” Technical Report, National Bureau of Economic Research 2016.
- O’Barr, William and Bowman Atkins**, ““Women’s Language” or “powerless language”?,” in Sally McConnell Ginet, Ruth Borker, and Nelly Furman, eds., *Women and languages in Literature and Society*, New York: Praeger, 1980, pp. 93–110.
- O’Barr, William M.**, *Linguistic evidence: Language, power, and strategy in the courtroom* Studies on Law and Social Control, New York: Academic Press, 1982.
- Orvice, Vikki**, “Why 90s Women are Lowing Their Tone,” 1993.
- Pemberton, Cecilia, Paul McCormack, and Alison Russell**, “Have women’s voices lowered across time? A cross sectional study of Australian women’s voices,” *Journal of Voice*, 1998, 12 (2), 208–213.
- Peresie, Jennifer L.**, “Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts,” *The Yale Law Journal*, 2005, 114 (7), 1759–1790.

- Phelps, Edmund S.**, “The Statistical Theory of Racism and Sexism,” *American Economic Review*, 1972, 62, 659–661.
- Posner, Richard A.**, “An Economic Approach to Legal Procedure and Judicial Administration,” *Journal of Legal Studies*, 1973, 2 (2), 399–458.
- Purnell, Thomas, William Idsardi, and John Baugh**, “Perceptual and Phonetic Experiments on American English Dialect Identification,” *Journal of Language and Social Psychology*, 1999, 18 (10), 10–30.
- Putnam, Hilary**, “Meaning and reference,” *Journal of Philosophy*, 1973, 70 (19), 699–711.
- Puts, David A, Julia L Barndt, Lisa LM Welling, Khytam Dawood, and Robert P Burriss**, “Intra-sexual competition among women: Vocal femininity affects perceptions of attractiveness and flirtatiousness,” *Personality and Individual Differences*, 2011, 50 (1), 111–115.
- Rao, Gautam**, “Familiarity Does Not Breed Contempt: Generosity, Discrimination and Diversity in Delhi Schools,” Technical Report, UC Berkeley Mimeo 2014.
- Reeve, Elspeth**, “Why Do So Many People Hate the Sound of Hillary Clinton’s Voice?,” 2015. New Republic, May 1, 2015.
- Roberts, John G.**, “Oral Advocacy and the Re-emergence of a Supreme Court Bar,” *Journal of Supreme Court History*, 2005, 30 (1), 68–81.
- Ruger, Theodore W, Pauline T Kim, Andrew D Martin, and Kevin M Quinn**, “The Supreme Court forecasting project: Legal and political science approaches to predicting Supreme Court decisionmaking,” *Columbia Law Review*, 2004, pp. 1150–1210.
- Rule, Nicholas O and Nalini Ambady**, “Democrats and Republicans can be differentiated from their faces,” *PLoS One*, 2010, 5 (1), e8733.
- Scherer, Klaus R.**, “Voice and speech correlates of perceived social influence in simulated juries,” in Klaus R. Scherer, Howard Giles, and R. St. Clair, eds., *The social psychology of language*, London: Blackwell, 1979, pp. 88–120.
- Schroeder, Juliana and Nicholas Epley**, “The Sound of Intellect Speech Reveals a Thoughtful Mind, Increasing a Job Candidate’s Appeal,” *Psychological Science*, 2015, 26 (6), 877–891.
- Schubert, James N., Steven A. Peterson, Glendon Schubert, and Stephen L. Wasby**, “Dialect, Sex and Risk Effects on Judges Questioning of Counsel in Supreme Court oral Argument,” in Frank K. Salter, ed., *Risky Transactions: Trust, Kinship, and Ethnicity*, Berghan Books, 2002, chapter 10, p. 304.
- Schubert, James N, Steven A Peterson, Glendon Schubert, and Stephen Wasby**, “Observing Supreme Court oral argument: A biosocial approach,” *Politics and the Life Sciences*, 1992, pp. 35–51.
- Starecheski, Laura**, “Can Changing How You Sound Help You Find Your Voice?,” 2014. National Public Radio, October 14, 2014.
- Sunstein, Cass R., David Schkade, Lisa M. Ellman, and Andres Sawicki**, *Are Judges Political? An Empirical Analysis of the Federal Judiciary*, Brookings Institution Press, 2006.
- Szmer, John J, Tammy A Sarver, and Erin B Kaheny**, “Have We Come a Long Way, Baby? The Influence of Attorney Gender on Supreme Court Decision Making,” *Politics & Gender*, 2010, 6 (01), 1–36.
- Tigue, Cara C., Diana J. Borak, Jillian J.M. O’Connor, Charles Schandl, and David R. Feinberg**, “Voice pitch influences voting behavior,” *Evolution and Human Behavior*, 2012, 33, 210–216.
- Todorov, Alexander, Anesu N Mandisodza, Amir Goren, and Crystal C Hall**, “Inferences of competence from faces predict election outcomes,” *Science*, 2005, 308 (5728), 1623–1626.
- Toft, Catherine A and Timothy F Wright**, *Parrots of the Wild: A Natural History of the World’s Most Captivating Birds*, University of California Press, 2015.
- Tversky, Amos and Daniel Kahneman**, “Advances in prospect theory: Cumulative representation of uncertainty,” *Journal of Risk and Uncertainty*, 1992, 5 (4), 297–323.
- Uhlmann, Eric Luis and Geoffrey L Cohen**, “Constructed criteria redefining merit to justify discrimination,” *Psychological Science*, 2005, 16 (6), 474–480.
- White, David, P Jonathon Phillips, Carina A Hahn, Matthew Hill, and Alice J O’Toole**, “Perceptual expertise in forensic facial image comparison,” in “Proc. R. Soc. B,” Vol. 282 The Royal Society 2015, p. 20151292.
- Yoshino, Kenji**, “The epistemic contract of bisexual erasure,” *Stanford Law Review*, 2000, pp. 353–461.
- , *The Hidden Assault on Our Civil Rights*, Random House Incorporated, 2006.
- Yu, Alan CL, Carissa Abrego-Collier, Jacob Phillips, Betsy Pillion, and Daniel Chen**, “Investigating Variation In English Vowel-To-Vowel Coarticulation In A Longitudinal Phonetic Corpus,” Technical Report 2015.
- Zweigenhaft, Richard L**, “Minorities and women of the corporation: Will they attain seats of power,” *Power elites and organizations*, 1987, pp. 37–62.

_____ and **G William Domhoff**, *Diversity in the power elite: Have women and minorities reached the top?*, Yale University Press, 1998.

TABLE A.1
 ADVOCATE EXPERIENCE AT THE SCOTUS AND OUTCOMES (NORMALIZED RATINGS, MICRO LEVEL)
 Dependent Variable: Case Outcome (win = 1, lose = 0)

	Petitioner (1)	Respondent (2)	Both (3)	Petitioner (4)	Respondent (5)	Both (6)
	argued only once			argued more than once		
Masculine	-0.00332 (0.0142)	0.00994 (0.0139)	-0.000833 (0.0103)	-0.0238** (0.0116)	0.00900 (0.0132)	-0.0145 (0.00938)
Confident	0.00497 (0.00967)	-0.00410 (0.00978)	0.00511 (0.00717)	0.00975 (0.00693)	-0.0146 (0.00910)	0.00830 (0.00600)
Attractive	-0.00159 (0.0105)	-0.00731 (0.00943)	-0.00345 (0.00734)	-0.00417 (0.00762)	0.00323 (0.00941)	-0.00145 (0.00645)
Intelligent	0.00340 (0.00854)	0.00234 (0.00829)	0.00364 (0.00593)	0.00649 (0.00669)	0.00320 (0.00747)	0.00707 (0.00526)
Trustworthy	-0.00278 (0.00729)	-0.00959 (0.00699)	-0.00638 (0.00503)	0.00418 (0.00557)	-0.00950 (0.00723)	-0.000708 (0.00465)
Aggressive	-0.00194 (0.00915)	0.00242 (0.00813)	0.000857 (0.00644)	-0.00560 (0.00632)	0.00183 (0.00825)	0.0000956 (0.00550)
Likely winner	-0.0146* (0.00845)	-0.0278*** (0.00861)	-0.0182*** (0.00612)	0.00414 (0.00702)	-0.00373 (0.00837)	0.00479 (0.00576)
Observations	7620	7579	15199	10045	8422	18467

Notes: Using the normalized perceptual ratings, Column 1 tests whether perceived masculinity is associated with case outcome among petitioners who argued only once in the Supreme Court from 1998-2012. Column 2 tests whether perceived masculinity is associated with case outcome among respondents who argued only once in the Supreme Court from 1998-2012. Column 3 tests the same for both sets of advocates. Columns 4-6 do the same as Columns 1-3 but among advocates who argued more than once. Standard errors are clustered by case. *, **, and *** indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

TABLE A.2
VOICE PERCEPTIONS AND GENDER (NORMALIZED 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	-0.0173** (0.00882)	-0.00975 (0.00684)	0.0191 (0.0163)	0.0261** (0.0122)	-0.0197** (0.00914)	-0.0122* (0.00695)	0.0238 (0.0159)	0.0172 (0.0104)
Confident					0.00912 (0.00591)	0.00738 (0.00465)	-0.0173 (0.0166)	0.00792 (0.0125)
Attractive					-0.00103 (0.00641)	-0.000228 (0.00481)	0.00127 (0.0161)	-0.0125 (0.0105)
Intelligent					0.00736 (0.00534)	0.00747* (0.00401)	-0.00410 (0.0211)	-0.0126 (0.0152)
Trustworthy					0.00166 (0.00450)	-0.00305 (0.00343)	-0.00952 (0.0130)	0.00663 (0.00831)
Aggressive					-0.00496 (0.00534)	-0.000718 (0.00415)	0.0120 (0.0175)	0.00166 (0.0126)
Likely winner					-0.00270 (0.00545)	-0.00451 (0.00419)	0.00786 (0.0148)	0.0104 (0.0110)
Observations	18514	35276	2543	5868	17665	33666	2486	5759

Notes: Using the normalized perceptual ratings, 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 for femininity 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 percent, 5 percent, and 1 percent levels, respectively.

TABLE A.3
GENDER, LAWYER, AND MTURKER RATER CONTROLS (NORMALIZED 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	-0.0198** (0.00913)	-0.0123* (0.00695)	0.0240 (0.0161)	0.0169 (0.0105)	-0.00766** (0.00372)	-0.00879*** (0.00312)	-0.00508 (0.00427)	-0.00299 (0.00529)
Confident	0.00919 (0.00591)	0.00747 (0.00463)	-0.0179 (0.0170)	0.00816 (0.0126)	0.00396 (0.00301)	0.00276 (0.00270)	0.00482 (0.00896)	-0.000308 (0.00630)
Attractive	-0.00134 (0.00641)	-0.000239 (0.00481)	0.000382 (0.0161)	-0.0123 (0.0105)	-0.000188 (0.00275)	0.00103 (0.00242)	0.000512 (0.00421)	-0.00288 (0.00474)
Intelligent	0.00744 (0.00531)	0.00743* (0.00400)	-0.00446 (0.0210)	-0.0129 (0.0153)	0.00291 (0.00240)	0.00336 (0.00215)	0.00535 (0.00714)	0.00661 (0.00661)
Trustworthy	0.00142 (0.00448)	-0.00314 (0.00342)	-0.00995 (0.0130)	0.00671 (0.00833)	0.00166 (0.00237)	0.000696 (0.00201)	0.00472 (0.00370)	0.00697 (0.00441)
Aggressive	-0.00493 (0.00533)	-0.000701 (0.00414)	0.0119 (0.0180)	0.00113 (0.0127)	0.00280 (0.00265)	-0.000389 (0.00227)	-0.00302 (0.00539)	-0.000171 (0.00661)
Likely winner	-0.00237 (0.00542)	-0.00441 (0.00419)	0.00871 (0.0151)	0.0103 (0.0110)	0.00294 (0.00293)	0.00414 (0.00276)	-0.00905* (0.00528)	0.00117 (0.00500)
MTurker controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lawyer FE	No	No	No	No	Yes	Yes	Yes	Yes
Observations	17665	33666	2486	5759	17665	33666	2486	5759

Notes: Using the normalized perceptual ratings, 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 for femininity 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 percent, 5 percent, and 1 percent levels, respectively.

TABLE A.4
VOICE PERCEPTIONS AND GENDER (NORMALIZED 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	
	Petitioner	Both	Petitioner	Both	Petitioner	Both	Petitioner	Both
Masculine/ Feminine	-0.0580*	-0.0314	0.122	0.152**	-0.0762*	-0.0357	0.381*	0.166
Confident	(0.0305)	(0.0228)	(0.106)	(0.0709)	0.177**	0.106	-0.454*	-0.0409
Attractive					(0.0887)	(0.0666)	(0.262)	(0.172)
Intelligent					-0.0385	-0.00770	0.0757	-0.107
Trustworthy					(0.0637)	(0.0457)	(0.205)	(0.140)
Aggressive					0.0689	0.108*	-0.0336	-0.00119
Likely winner					(0.0814)	(0.0583)	(0.146)	(0.107)
Observations	856	1634	116	267	0.0117	-0.0582	-0.241	0.102
					(0.0769)	(0.0567)	(0.256)	(0.153)
					-0.0995	-0.0209	0.206	-0.0168
					(0.0705)	(0.0546)	(0.187)	(0.133)
					-0.0697	-0.105	0.111	0.0214
					(0.0906)	(0.0708)	(0.269)	(0.185)

Notes: Using the average normalized rating per advocate, 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 for femininity 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 percent, 5 percent, and 1 percent levels, respectively.

TABLE A.5
CORRELATIONS OF TRAIT JUDGMENTS WITH MASCULINE (FEMININE) RATINGS (NORMALIZED RATINGS, MICRO LEVEL)

Variable	Male raters		Female raters	
	Male advocate	Female advocate	Male advocate	Female advocate
Confident	0.4234	0.446	0.4598	0.4877
Attractive	0.3068	0.2244	0.3686	0.2408
Intelligent	0.194	0.0478	0.2714	0.0793
Trustworthy	0.2058	0.4202	0.1948	0.3997
Aggressive	0.3223	0.3226	0.3603	0.3023
Likely winner	0.3909	0.5031	0.4351	0.5346

Notes: This table presents the correlation between other perceptual ratings and masculine (feminine) rating for each combination of male or female advocate rated by male or female raters.

TABLE A.6
VOICE PERCEPTIONS AND GENDER MATCH (NORMALIZED RATINGS, MICRO LEVEL)

Dependent Variable: Case Outcome (win = 1, lose = 0)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Male raters				Female raters			
	Male advocate		Female advocate		Male advocate		Female advocate	
	Petitioner	Both	Petitioner	Both	Petitioner	Both	Petitioner	Both
Masculine/ Feminine	-0.0228** (0.00957)	-0.0150** (0.00740)	0.0125 (0.0205)	0.0130 (0.0132)	-0.0158 (0.0107)	-0.00888 (0.00802)	0.0319 (0.0203)	0.0205 (0.0127)
Confident	0.00884 (0.00767)	0.00714 (0.00586)	-0.0212 (0.0234)	0.00395 (0.0151)	0.00927 (0.00795)	0.00757 (0.00605)	-0.0133 (0.0187)	0.0114 (0.0143)
Attractive	0.00627 (0.00712)	0.00312 (0.00545)	-0.00757 (0.0241)	-0.0180 (0.0150)	-0.00803 (0.00819)	-0.00361 (0.00625)	0.00940 (0.0171)	-0.00719 (0.0122)
Intelligent	0.0129* (0.00681)	0.0111** (0.00499)	-0.0108 (0.0218)	-0.0173 (0.0166)	0.00188 (0.00681)	0.00377 (0.00511)	0.00186 (0.0258)	-0.00849 (0.0173)
Trustworthy	-0.00120 (0.00622)	-0.00386 (0.00477)	0.000968 (0.0200)	0.0101 (0.0113)	0.00449 (0.00605)	-0.00199 (0.00462)	-0.0193 (0.0180)	0.00368 (0.0120)
Aggressive	-0.00656 (0.00686)	-0.00376 (0.00527)	0.00307 (0.0237)	-0.00545 (0.0151)	-0.00354 (0.00654)	0.00250 (0.00532)	0.0199 (0.0186)	0.00869 (0.0142)
Likely winner	-0.0000673 (0.00730)	0.000217 (0.00573)	0.0191 (0.0235)	0.0160 (0.0144)	-0.00508 (0.00705)	-0.00957* (0.00564)	-0.000289 (0.0176)	0.00493 (0.0144)
Observations	8736	16606	1141	2605	8929	17060	1345	3154

Notes: Using the normalized perceptual ratings, 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 femininity among 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 percent, 5 percent, and 1 percent levels, respectively.

TABLE VII
ADDITIONAL PERCEPTIONS (NORMALIZED RATINGS, CASE LEVEL)
Dependent Variable: Case Outcome (win = 1, lose = 0)

	(1)
Masculine	-0.555*** (0.179)
Confident	0.379 (0.300)
Attitude	-0.020 (0.283)
Competent	0.588 (0.405)
Empathy	0.212 (0.286)
Respect	-0.389 (0.320)
Likely winner	-0.367 (0.259)
Observations	60
R-squared	0.179

TABLE VIII
 ADDITIONAL PERCEPTIONS (NORMALIZED RATINGS, MICRO LEVEL)
 Dependent Variable: Case Outcome (win = 1, lose = 0)

	(1)
Masculine	-0.080** (0.031)
Confident	0.020 (0.027)
Attitude	0.005 (0.019)
Competent	0.017 (0.019)
Empathy	0.006 (0.014)
Respect	-0.003 (0.016)
Likely winner	-0.008 (0.021)
Observations	1260
R-squared	0.021

Figure 1: First Screenshot of Experiment

Please do not complete more than one HIT in this batch. You will not be paid for multiple submissions in this batch.

Survey on Speech Perception

Phonology Laboratory, Department of Linguistics, University of Chicago

Overview

- This survey aims to better understand speech perception. The sections of the survey, all to be completed within 1 hour in one single sitting, are in the format of short questionnaires and tasks on:
 - Your demographic information
 - The way you speak and understand language
- All information collected will be kept anonymous and confidential.

Headphones

- To participate in this task you will need headphones.
- Please put on your headphones now if you have not done so. Check browser compatibility and adjust the speaker volume using the following sound.



Web browser

- **For this survey, we strongly recommend Chrome**, as there are reported problems with other browsers (e.g., cannot submit your answers, cannot play sound files in the middle of the HIT).
- A broadband internet connection is strongly recommended, as a good amount of audio files are loaded for this HIT. Before you begin this HIT, please ensure that your internet connection will be stable throughout the HIT, close all other applications with a high data-streaming demand, and close all other tabs of this browser. If your current actual download speed is at least 1 Mbps (to test your current connection speed in a new tab: <http://www.speedtest.net/>), your connection should be fast enough. To ensure that the HIT runs smoothly, it is also recommended that other applications and background tasks are closed.
- If you are experiencing any problem, please let us know the browser and system you are using and you will be able to participate in a future study.

Figure 2: Second Screenshot of Experiment

Instructions

- In the following task, you will listen to a series of audio recordings from different talkers. Please follow the instructions and provide your impression of the talkers.
- There are altogether 66 questions. Please expect to spend roughly 25 minutes for all of them (approx. 20 seconds per question).
- For each question, the associated audio clip will be automatically played once when the question appears, and you may play it as many times as you like.
- Please respond as quickly as possible. Please report your first impression and do not dwell on a particular file for too long.
- Click on "Next" to begin the task.

Next

Figure 3: Third Screenshot of Experiment

Recording 1 of 66



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

Very Attractive	<input type="radio"/>	Very Unattractive						
Very Masculine	<input type="radio"/>	Not At All Masculine						
Not Intelligent	<input type="radio"/>	Intelligent						
Very Unaggressive	<input type="radio"/>	Very Aggressive						
Not Trustworthy	<input type="radio"/>	Trustworthy						
Very Confident	<input type="radio"/>	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 4: Fourth Screenshot of Experiment

Profile

Please answer the following questions:

Year of birth:

Gender:

State of residence:

Highest educational attainment:

Race:

Total annual household income:

[Next](#)