University of Southampton

Economics Department
University of Southampton
Southampton SO17 1BJ, UK

Discussion Papers in
Economics and Econometrics

Lady Justice: The impact of female judges on jury trials verdicts in North Carolina.

Foresta Alessandra
University of Southampton

No. 2305

This paper is available on our website
http://www.southampton.ac.uk/socsci/economics/research/papers

ISSN 0966-4246

Lady Justice: The impact of female judges on jury trials verdicts in North Carolina.

Foresta Alessandra*†[1]

[1]University of Southampton

October 31, 2023

Abstract

This study evaluates the impact of judges' gender on jury trial outcomes in the U.S. state of North Carolina. The identification strategy is based on judges' rotation across different districts. The results indicate that, in trials presided over by female judges, juries are more likely to render guilty verdicts. I conduct a series of robustness and heterogeneity checks. Finally, I discuss the potential mechanisms underlying these findings and explore the influence of the jury selection process, the role of judges' demeanour, and women's attitudes towards the courts and sentencing.

JEL Codes: K10, K40, J16
Keywords: Gender, Judge, Trials behaviours

*a.foresta@soton.ac.uk

†I thank the North Carolina Sunshine Project, Ron Wright, Francis Flanagan, Shea R. Denning for the help with the dataset and the information over the North Carolina Judicial system. I would also like to thank Andrew Pickering, Cheti Nicoletti, Emma Tomminey, Neil Cummins, Brendon Mcconnell, along with the participants of the AME cluster seminar in York, SES2022, EPCS2022, seminar in Southampton, and the 2nd Workshop in Gender and Economics at the University of Luxembourg for their constructive suggestions that greatly enriched this work. This article has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 897497.

1 Introduction

In many countries, the concept of *Justice* is represented by *Lady Justice*, a blindfold woman holding a scale and a sword. The sword symbolizes authority, while the scale measures the balance between support and opposition in a given case. The blindfold represents impartiality, a fundamental principle that dictates justice should be applied without regard to power, wealth, or status. Impartiality stands as a cornerstone in numerous justice systems, with judges often being described as its "embodiment". However, in reality, judges do not always exhibit the level of neutrality and impartiality they are expected to uphold. For example, works like Spitzer and Talley (2013) and Johnson (2014) indicate that certain judges' characteristics, such as political affiliation or gender, the duration of sentences in bench trials within the US.[1] In jury trials, Lenehan and O'Neill (1981) and Halverson et al. (1997) find that judges non-verbal behaviours can sway juries toward one decision or another.

In this work, I explore the influence of judges' gender on the verdicts rendered in jury trials. To the best of my knowledge, this article is the first to examine the impact of judges' gender on the outcomes of jury trials. The empirical analysis uses data on felony trials conducted in North Carolina from 2010 to 2012. The identification strategy exploits the (mandatory) judges' rotation across different districts. The findings suggest that female judges enhance the likelihood of guilty verdicts in jury trials. Furthermore, I implement a battery of robustness checks that encompass diverse model specifications, district dimensions, and workloads. A sensitivity check is also conducted, involving the exclusion of individual judges, and the outcomes consistently align with the primary results. I also run a series of heterogeneity checks based on judges' traits. Finally, I investigate the possible mechanisms behind these findings. While the jury selection process appears to lack a statistically significant impact, discernible effects arise from disparities in the sternness exhibited by female and male judges, as well as differing attitudes held by women and men towards criminal offences and sentencing.

This article draws heavily from existing literature in various ways. First, it is connected with the vast body of research exploring gender differences in the decision-making process. For example, Gilligan (1982) suggests that men and women employ distinct criteria when making decisions. According to this author, women often base decisions on an "ethic of care", emphasizing factors such as "correctness", responsibility and contextual considerations. In contrast, men tend to utilize an

[1] *Bench trials* involve judges taking on the role of fact-finder in addition to delivering sentences. The trials examined in this paper are *jury trials*, where the jury determines the facts, and the judge oversees the trial, issuing a sentence after the jury reaches a verdict. For further details, refer to Section 2.

"ethics of rights", relying on abstract rules and principles such as individual liberty and hierarchical structures. These gender-based differences in decision-making criteria could potentially result in variations in the decision-making and behaviours of judges. Fox and Van Sickel (2000) indicate that female judges are more likely to align with the prosecution, while their male counterparts are more inclined to align with the defence. Their findings further suggest that female judges tend to adopt an inclusive and procedural sentencing style. Conversely, a consensual and authoritarian sentencing style is typically associated with male judges. These distinctions provide valuable insights into the potential impact of judges' gender on courtroom dynamics and outcomes.

Additionally, this article is linked to the literature that explores the relationship between judges' characteristics and trial outcomes. While a significant portion of prior research has concentrated on bench trials and sentencing, jury trials have garnered relatively less attention. According the existing body of work, the impact of judges' gender on sentencing appears to be less straightforward compared to the influence of other judicial attributes, such as political affiliation (as exemplified by studies like Schanzenbach and Tiller (2008), Fischman and Schanzenbach (2011), and Spitzer and Talley (2013)). Johnson (2014) suggests that female judges lean toward leniency in their sentencing, while contrasting viewpoints are presented by Steffensmeier and Hebert (1999) and Spohn (1991), indicating that women judges may actually impose more severe sentences. Finally, Songer et al. (1994) propose that distinctions in sentencing behaviours between male and female judges are predominantly observable in cases related to employment discrimination.

Furthermore, this study also builds upon prior research examining the influence of judges on juries. While judges are ideally perceived as impartial figures devoid of any sway over juries, existing literature indicates that juries are often privy to the inclinations of judges regarding a case with an significant role played by judges non-verbal behaviours (e.g., Hart (1992), Hart (1995)). Ekman and Friesen (1969) investigate the impact of judges' non-verbal behaviours on juries' verdicts and they find that even the most composed and controlled judge can inadvertently convey biased information through non-verbal cues. Moreover, there are multiple cases in which courts rule that defendants' rights have been violated by judges' non-verbal behaviours (e.g., Blanck et al. (1985)). This article tries to (partially) fill the gap and investigates the impact of judges' gender in the relationship between judges and juries.

The remainder of the paper is organised as follows: in Section 2 I focus on the legal system in North Carolina, in Section 3 on the dataset and in Section 4 on the empirical strategy. In Section 5, I present the main results, in Section 5.2 and Section 5.3 the robustness and heterogeneity checks. Finally, in Section 6 I speculate over the possible explanations behind the results and, in Section

7, I list my conclusions.

2 North Carolina Legal system and Judge

Jury trials in the US involve various key "participants": the defendant, the defence attorney, the prosecutor, a (seated) jury and a judge. The defendant is the individual who stands accused of the offence(s), and the defence attorney is the lawyer responsible for advising and advocating on behalf of the defendant. Prosecutors, on the other hand, act on behalf of the state and present the case against the defendant. They also have substantial influence in determining whether and when trials proceed, as suggested by Bandyopadhyay and McCannon (2014). In the context of North Carolina, prosecutors are called *Assistant District Attorneys*. At the helm of the prosecution office is the "District Attorney", an elected official who holds a four-year term.[2]

The most distinctive feature of Anglo-American legal system is the jury, tasked primarily with evaluating evidence and delivering a verdict. Specifically, the jury overseeing a trial is referred to as the "seated jury", and in North Carolina, it is composed by 12 jurors along with alternates.[3] *Jury pool* is the group of potential jurors, individuals who may be selected to serve.[4] The process of jury selection, elaborated further in Section 6.1, serves to transform the jury pool into the seated jury by excluding jurors who might exhibit potential biases.

The person with the greatest power in the courtroom is the judge. In the context of jury trials, judges hold the responsibility of overseeing the proceedings and ultimately delivering a sentence subsequent to the jury's verdict.[5] The judges' role is to ensure that laws and procedures are always respected during the trials. For example, they review whether there are any illegality issues in the submitted evidences and they are responsible to provide jury instructions.[6] The cornerstone of judges' conduct rests upon their impartiality, an imperative that necessitates divorcing personal opinions, expectations, backgrounds, and attributes from the proceedings. The trials in the analysed sample are presided by a specific group of state judges, namely North Carolina's Superior Court Judges. These judges are elected officials who serve eight-year terms. A minority of judges in the sample, approximately 14%, are appointed by the Governor and fulfil five-year terms.[7] While

[2] North Carolina is divided in 48 prosecutorial districts in the analysed period.

[3] Alternates are jurors designated to replace seated jurors if the latter become unable to fulfil their role. The number of alternates per trial is determined by the presiding judge.

[4] According to North Carolina legislation, a qualified potential juror must be; U.S. citizen, a resident of the county where the summons was issued, at least 18 years old, able to understand English and physically/mentally competent. Potential jurors must not: have served as a juror during the previous two years; have served a full term as a grand juror in the last six years and have been convicted of a felony (unless citizenship rights have been restored). Source: North Carolina Jury Service.

[5] For a more comprehensive exploration of distinct trial types, refer to note 1.

[6] Jury instructions encompass the guidelines provided by the judge to the jury after the conclusion of evidence presentation. These instructions elucidate the pertinent laws applicable to the case.

[7] This aspect is accounted for in the analysis by including a control for elected judges (see Section 5).

all judges must hold legal qualifications, they are prohibited from practising law privately during their tenure on the bench. There are no term limits, though judges must be below 72 years of age.[8]

3 Data

The dataset is collected by the research team of North Carolina Jury Sunshine Project.[9] It consists in felony trials in North Carolina between 2010 and 2012.[10] The dataset includes judges' information such as name and surname, ethnicity,[11] gender, political affiliation, elections details[12] and experience.[13] The dataset also incorporates trial-specific data, such as the list of charges, verdict outcomes, sentences, defendant characteristics (age, ethnicity, and gender), and traits of both potential and seated jurors (political affiliation, gender, and ethnicity). For the purposes of this article, the sample is confined to non-capital felony trials.[14] Further limitations involve focusing solely on cases with one defendant and one charge, where the verdict is known.[15] The main descriptive statistics are presented in Table 1.

[TABLE 1 APPROXIMATELY HERE]

As depicted in Table 1, within the sample, 12 % of the defendants is female, 63 % is identified as black and 34 % is identified as white. The majority of defendants, 64 %, is aged over 30. Following Anwar et al. (2012, 2014) and Flanagan (2018), I construct a set of dummies to represent various charges. Notably, the most prevalent offences within the sample are property offences (30%) and drug offences (19%). On average, the composition of juries pools is roughly balanced in terms of gender, with a relatively higher percentage of individuals identifying as white (61 %) and a smaller proportion identifying as black (17 %). Moreover, the political affiliation within jury pools indicates that Republican and Democratic jurors make up approximately 26% and 32% of the composition, respectively. Overall, 73% of the defendants included in the sample have been

[8] Source: North Carolina Judiciary branch site.

[9] Website: Sunshine Project Website. For an in-depth description of the data collection methodology, consult Flanagan (2018).

[10] There are 7 trials in the data that are set in 2008/2009 or 2013/2014 due to delays in the court system and lags between the jury selections and the conclusion of the trials. In accordance with Flanagan (2018), I include them in the analysis.

[11] Given the limited number of observations in some categories, I re-frame ethnicity as white, black, unknown and other, which includes the categories Asian, Native American or indigenous, Hispanic and other.

[12] In instances where election data is absent, I have leveraged sources such as Ballotpedia to fill in the gaps.

[13] There are some unknown judges in the dataset. By cross-referencing trial dates and court calendars, I was able to identify some of the missing judges. However, judges who remained unidentified were excluded from the sample (32 observations)

[14] Exclusions encompass murder charges classified as A1 felonies in North Carolina (33 observations).

[15] Given the structure of judicial courts in North Carolina, I remove trials with less than 12 jurors in the pool and in the seated jury, trials with more than 20 seated jurors and with more than 60 jurors in the jury pool (37 obs.). Given the low number of libertarians, I also removed trials with a libertarian in the jury pool (20 obs.). I also remove trials with unknown judges' gender (17 obs.), mistrials (59 obs.) and trials with defendant of unknown gender and/or ethnicity (23 obs.).

convicted. To mitigate potential influences stemming from judge gender on jury selection and any resultant pre-trial interactions that could potentially affect jurors' attitudes (as suggested by Anwar et al. (2012)), I incorporate jury pool characteristics as controls in the analysis.

Trials presided by a female judges constitute 8 % of the sample, while those presided over by non-white judges account for 15%. Democrats judges oversee 62 % of trials, whereas Republicans judges are in charge of only 27 %. Moreover, Johnson (2014) and Steffensmeier and Hebert (1999) suggest that tenure on the bench might have an impact on sentences duration. The variable *Experience (per 100)* represents the number of years elapsed since receiving the "Juris Doctor" degree, a graduate-entry professional law degree, serving as a proxy for experience.[16] On average, judges possess 30 years of experience and manage a workload of approximately 3.48 trials. Given the possible relevance of these characteristics on the main outcomes, I investigate if these variables are, on average, statistically different between judges' gender. In Table 2, I provide the results for the t-test conducted to compare various characteristics of judges between female and male judges.

[TABLE 2 APPROXIMATELY HERE]

The preliminary insights gleaned from Table 2 reveal, as expected, certain statistically significant distinctions between female and male judges. In aggregate, women judges appear to possess comparatively less experience, exhibit a lower proportion of white ethnicity, manage fewer cases, and have a greater likelihood of facing an upcoming election in the following year. Considering the meaningful nature of these observed disparities between female and male judges and their potential repercussions on the primary outcomes, I introduce these variables as control factors within the main analysis (refer to Equation 2). This inclusion helps to mitigate the potential influence of these factors on the relationships under scrutiny and enhances the accuracy of the conclusions drawn from the study.

4 Research Design

To identify the impact of judges' gender on the outcomes of jury trials, I employ a research design based on judges' rotation and fixed effects. Judges' rotation[17] is mandate by North Carolina Constitution (Article 4, Section 11).[18] In more detail, North Carolina is divided in 100 counties,

[16] To be fully authorized to practice law in North Carolina, individuals must also pass a bar examination. However, this specific information is not available in the database.

[17] The rotation was suspended in 1990, 2002 and 2009 due to budgetary constraints. However, these years are not included in the sample. Source: Article.

[18] The Article specifically states: "[t]he principle of rotating Superior Court judges among various districts of a division is a salutary one and shall be observed.".

50 districts[19] and 8 divisions, as shown in Fig 1.

[FIGURE 1 APPROXIMATELY HERE]

Superior Court Judges are assigned to one of the 50 districts, and every six months they rotate across different districts within the same division (N.C. Gen. Stat. sec. 7). For instance, consider judge A, initially assigned to Robson County, corresponding to district 16B in the fourth division. After six months, judge A leaves district 16B and moves to district 16A, which is also part of the fourth division and encompasses two counties: Scotland and Hoke. Judges' schedules are usually determined well in advance and are available online for consultation.[20] This mechanism was designed to eliminate possible conflicts of interest and corruption incentives, ensuring that judges do not have the discretion to select the trials they preside over.

Following Flanagan (2018), the identification strategy incorporates several fixed effects. Firstly, the research design includes time and district fixed effects to account for factors such as judges' availability, crime trends, and economic conditions. In more detail, the time fixed effects are based on the division of the judges' calendar proposed by the North Carolina Administrative Office of the Courts.[21] This office divides the year in two sessions: a *Spring session* from January to June and a *Fall session* from July to December. Time fixed effects are computed following this division.[22] Secondly, I introduce prosecutor fixed effects. Existing literature underscores the substantial influence of prosecutors on trial scheduling in the US. According to Bandyopadhyay and McCannon (2014), prosecutors possess considerable power in assigning trials, which could potentially compromise the research design. For instance, judges' personal preferences and tendencies might align with their gender, and prosecutors might strategically "manipulate" the calendar to allocate specific trials to particular judges. This deliberate allocation of trials could undermine the causal interpretation of my findings. To mitigate this concern, I incorporate prosecutor fixed effects into the primary regression models.[23]

Lastly, in line with prior literature (e.g., Anwar et al. (2012) and Foresta (2022)), I validate my identification strategy by conducting a regression of judges' attributes on observable trial characteristics, as illustrated in Equation (1).

$$FemaleJudge_i = \alpha + \beta_1 Trial_i + \beta_2 Def_i + \beta_3 Jury_i + DistFE_i + ProsFE_i + TimeFE_i + \epsilon_i \quad (1)$$

[19] A reform that took place in 2016 reduced the number of districts from 50 to 48. All the trials present in the sample took place before 2016.
[20] Source: North Carolina Courts Website.
[21] Source: Website.
[22] I also replicate the main analysis using year fixed effects, with similar findings. Results available upon request.
[23] I also replicate the analysis without prosecutor fixed effects, yielding similar findings to Tables 3 and Table 4. Available upon request.

where $FemaleJudge_i$ is a binary variable denoting judges' gender, $Trial_i$ encompasses trial characteristics, Def_i represents defendant characteristics, and $Jury_i$ encapsulates jury characteristics. $DistFE_i$, $ProsFE_i$, and $TimeFE_i$ denote district, prosecutor, and time fixed effects, respectively. If judges' gender is indeed randomly assigned to each trial, there should be minimal or no correlation among the variables in Equation (1). In Table 3, I present the results with and without fixed effects in Columns (1) and (2), respectively. Should judges' rotation proves adequate in ensuring the random assignment of judges' gender, the coefficients in Column (1) would likely lack statistical significance. Conversely, if rotation falls short and fixed effects are necessary, certain coefficients in Column (1) could attain statistical significance, while such significance would likely dissipate in Column (2).

[TABLE 3 APPROXIMATELY HERE]

The findings in Table 3 are highly encouraging. In Column (1), there are three coefficients that exhibit slight statistical significance, and this significance dissipates upon inclusion of fixed effects (Column (2)). While the findings in Table 3 may not entirely eliminate the possibility that judge selection is tied to unobservable characteristics, they strongly indicate that this concern is unlikely to be of significant magnitude.

5 Results

5.1 Main results

This section delves into the assessment of the causal influence of female judges on trial outcomes using a linear probability model. As mentioned in Sections 3 and 4, the regressions encompass district fixed effects, prosecutor fixed effects, time effects, and a set of control variables, as outlined in Equation 2.

$$Guilty_i = \alpha + \beta_1 FemaleJudge_i + \beta_2 Controls_i + DistrictsFE_i + ProsFE_i + TimeFE_i + \epsilon_i \quad (2)$$

where $Guilty_i$ is the dependent variable, a binary indicator for guilty or non-guilty outcomes and $FemaleJudge_i$ represents a binary variable for judges' gender. The set of control variables, $Controls_i$, is based on attributes of the jury pool.[24] Furthermore, $DistrictsFE_i$, $ProsFE_i$, and $TimeFE_i$ stand for fixed effects related to districts, prosecutors, and time. Given that trial distribution is not uniform across judges, to prevent potential over-rejections (as highlighted in Flana-

[24] I also implement the analysis using the characteristics in the seated juries rather than the jury pools with similar results. The results are presented in Table A2 in Section A1 of the Appendix.

gan (2018), Cameron et al. (2008), Carter et al. (2017), and Cameron and Miller (2015)), I adopt heteroscedasticity-robust standard errors instead of clustered standard errors at the judge level.[25] The outcomes for Eq. 2 are outlined in Table 4.[26]

[TABLE 4 APPROXIMATELY HERE]

In Table 4, the coefficients in Columns (1)-(4) for female judges are always positive and statistically significant, indicating that female judges systematically increase the probability of guilty outcomes. In terms of magnitude, having a female judge increases of 14.7 percentage points the probability of a guilty verdict in Column (1) and of 27.7 percentage points in Column (4), my benchmark model. The magnitudes of the findings are coherent with the previous literature. For example, Anwar et al. (2012) find that there is a 16-percentage point conviction gap between all white juries and juries where there is at least one black member in the jury pool. Furthermore, other judges characteristics do not seem to have a statistically significant impact on guilty verdict, with some slightly significant exceptions for certain types of crimes as shown in Table A1 in the Online Appendix.

To assess whether the estimated gender effect is statistically significant as a result of pure chance, I implement a permutation test, following the previous literature (e.g. Nagler et al. (2020), Bertrand et al. (2004) and Fujiwara and Wantchekon (2013)). This test randomly reassigns judges' gender (the treatment) in the sample and re-estimates β using this placebo assignment multiple times (1,000 in this case).[27] The randomization inference test for the benchmark model (Column (4) of Table 4) indicates that the estimated coefficient is statistically significant at 1% and larger in magnitude than almost all simulated effects, as shown in Figure 2.[28]

[FIGURE 2 APPROXIMATELY HERE]

5.2 Robustness checks

To check that the results in Table 4 are reliable and not driven by some outliers, I implement a series of robustness checks. First, given that the dependent variable is a dummy, I replicate the main analysis using a Logit model. Second, there is also heterogeneity across districts dimensions, with some districts having more than 20 trials and some having less than 5. To control that the results are not driven by very big or very small districts, I create an average measure of trials per

[25] I also conduct the analysis using cluster standard errors at judge level with similar results. Available upon request.

[26] In Appendix A1, I provide comprehensive results that include coefficients for all control variables (Table A1).

[27] To implement this analysis, I use the randomization inference test (*ritest*) proposed by Heß (2017).

[28] I replicate this test also for the other specifications presented in Table 4 with similar results. Graphs available upon request.

district at semester level and remove districts with the top 5% and the low 5% of this distribution. Third, different judges have different caseloads and the main results could be driven by a specific group of judges with a particular big or small caseloads. Similarly to districts, I create an average measure for the judges workload at semester level and remove judges in the top 5% and low 5% of the caseloads average distribution.[29] The results corresponding to the benchmark model are presented in Table 5.[30]

[TABLE 5 APPROXIMATELY HERE]

Column (1) of Table 5 is the benchmark specification (Column (4) of Table 4). Column (2) of Table 5 reports the margin estimated by a Logit model. This specification has a significant drop in observations, but the finding remains positive and statistically significant.[31]

In Columns (3) and (4) of Table 5, I remove the districts that have very few or very high number of trials. Specifically, I remove the low 5% (Column (3)) and high 5% (Column (4)) of the distribution of average trials per districts. The removal of "small" an "big" districts do not seem to affect the coefficients for female judges, which remain positive and significant. Columns (3) and (4) in Table 5 indicate that the results are not driven by very "small" or "big" districts.

In Columns (5) and (6) of Table 5 I evaluate the effect of judges' workloads. Specifically, I remove the low 5% percentiles (Column (5)) and high 5% (Column (6)) of the distribution of average workload per judge. The coefficients for female judges are similar to the benchmark model and they remain positive and significant. These findings indicate that the results in Table 4 do not seem to be driven by judges with very high/low workloads.

5.2.1 Sensitivity checks

Given the relatively low number of judges in the sample, I implement a sensitivity check by removing each judge from the sample and replicate the analysis for the benchmark model (Column (4) Table 4). The coefficients from the estimation of the female coefficients in these new samples are presented in Fig. 3.

[Figure 3 APPROXIMATELY HERE]

While there is some variations across the different samples, all the coefficients remain positive and statistically significant. Overall, it does not seem that the main findings are driven by a single judge.

[29] In some cases the data removed are a little bit higher due to the distribution of the averages.

[30] The results with all the coefficients and all the specifications are presented in Tables A3, A4, A4, A7 in Section A2 in the Appendix. Overall, all the coefficients are positive and statistically significant.

[31] Although the findings for the Logit model are larger, once I replicate the linear probability model in the Logit sub-sample, I obtain a similar magnitude. Findings available upon request.

5.3 Heterogeneity checks

As indicated in Section 2, judges are quite different across many dimensions. Previous literature indicates that ethnicity can have an impact on jurors decisions (e.g., Anwar et al. (2012) and Flanagan (2015)) and judges (Spohn (1991)). Moreover, some works indicate that political affiliation of judges could play a role (e.g., Johnson (2014), Spitzer and Talley (2013)). Finally, judges are elected officials and it is possible that the electoral cycle has some effects on their behaviours (Coviello and Gagliarducci (2017)). To evaluate this possible heterogeneity, in Table 6 I replicate the analysis by interacting female judges with a dummy for judges' ethnicity (Column (2)), judges' political affiliation (Column (3)) and electoral cycle (Column (4)). In Column (1) of Table 6 I report the benchmark model (Column (4) Table 4).[32]

[TABLE 6 APPROXIMATELY HERE]

The findings presented in Columns (2)-(4) of Table 6 are interesting. In Column (2) of Table 6 the interaction between female judges and non-white judges is negative but non-significant. On the contrary, if the judges is a white woman, there is an increase in the probability of a guilty verdict of 37% percentage points. This is in line with some previous works indicating that white jurors are more likely to convict (Flanagan (2015)). In Column (3) of Table 6, the interaction between Republican and female judges is non-significant. Similarly, interaction between independent and female judges is non-significant while the result for democratic female judges is positive and significant. These results are quite surprising. As mentioned in section 1, the previous literature indicated that Republican judges tend to be harsher than Democratic judges. However, Table 6 indicates that it is not always true and that gender could play an important role. Finally, in Column (4) of Table 6 I examine the possible presence of an electoral cycle effect. The interaction between the dummy for next year elections and female judges is non-significant, suggesting that career concerns do not seem to play a statistically significant role.[33]

In Section A3.1 of the Appendix, I also replicate the analysis by interacting female judges with different types of crimes, different jury characteristics and different defendants' characteristics. Generally, the results tend to be non significant with two interesting exceptions: the interaction between *Female Judge* and *Def. Non-White* (Column (2)) and the interaction between *Female Judge* and *Jury pool: other (Prop.)* (Column (3)). Both coefficients are positive and statistically significant.

[32] The results with all the coefficients are presented in Tables A8 in Section A3 of the Appendix.

[33] I also replicate the analysis using a dummy for this year elections and set of dummies for the electoral cycle, with similar findings. Results available upon request.

6 Why does judges' gender play a role?

The previous sections document the importance of judges' gender on jury trials outcomes. In this section, I speculate over the possible mechanisms behind these findings. First, I investigate the role of judges' gender in the jury selection process (Section 6.1). Second, in Section 6.2, I evaluate the possible differences in attitudes and behaviours of female judges.

6.1 Jury selection process and seated jury composition

A possible mechanism behind the results in Table 4 is that, when there is a female judge, the jury selection processes systematically select "harsher" jurors. To understand how it is possible, it is crucial to investigate the role of judges (and other actors) during these proceedings. During the jury selection process, judges, defence attorneys, and prosecutors remove those jurors that they deem to be biased. Potential jurors can be excluded either *for cause* or thanks to *peremptory challenges*. A *removal for cause* is when the judge decides to struck a juror for apparent bias or hardship.[34] These removals are unlimited but are left to judges' discretion. Prosecutors and defence attorneys can exclude jurors through the *peremptory challenges*. These removals do not require any explanations, but they are limited in numbers and cannot be based on ethnicity or gender (*Batson v. Kentucky*, 476 U.S. 79 [1986]; *J.E.B. v. Alabama*, 511 U.S. 127 [1994]).[35] In North Carolina, prosecutors and defence attorneys have six peremptory challenges plus one for every alternative juror each.

If female and male judges behave differently in their removals' choices during the jury selection process, this could influence the seated juries' composition, and, as a consequence, the verdicts. To evaluate the relationship between removals, characteristics of jurors and judges, I implement a series of OLS regressions.[36] As dependent variables, I use a dummy for being remove from the jury pool. The regression is the following:[37]

$$Struck_i = \alpha + \beta_1 X_i FemaleJudge_i + \gamma_1 X_i + TrialFE_i FemaleJudge_i + \epsilon_i \tag{3}$$

Where $Struck_i$ represents a dummy for being removed form the jury pool, X_i represents the characteristics of potential jurors, $TrialFE_i$ represents Trials Fixed Effects and $FemaleJudge_i$ represents a dummy for female judges. The summary statistics for these variables are presented

[34] Some examples are medical or financial hardship.

[35] If one side is suspected of gender/race discrimination, the opposing side may object using the so-called Batson's challenges. In practice, successful Batson challenge are extremely rare.

[36] Additionally, there is a literature about the role of number of jurors in jury outcomes. I replicate my analysis using the choice in the number of alternative jurors. However, the results are not statistically significant. Available upon request.

[37] I also replicate the analysis following Flanagan (2015) with similar results. Available upon request.

in Table A10 in Section A4. I also replicate the analysis in the subsample of removed jurors using as dependent variables a series of dummies for being removed by the judges, by the prosecutors or by the defence attorneys. The findings for Eq. (3) are presented in Table 7.

[TABLE 7 APPROXIMATELY HERE]

The findings in Table 7 are quite interesting. In Column (1), there are some statistically significant interactions. When there is a female judge, there seems to be a positive and significant relation between the interactions between *female judge* and *race other juror* and *female judge* and *race unknown juror* and negative and significant interactions between *female judge* and *Pol.Aff. unknown juror*. Overall, it seems that in some cases, when a female judge is present some difference in the striking behaviours are present. However, according to the previous literature, none of these characteristics is systematically associated with more guilty verdicts.

In Columns (2)-(4), I examine in more details if it is possible to attribute this striking patterns to a specific actor. Overall, it seems that female judges are slightly less likely to remove female jurors (column (2)) but overall there seems to be not a very significant difference in *removal by court* between female and male judges. Similarly, prosecutors and defence attorneys overall seem to behave in similar ways, with 3 exceptions. When there is a female judges, prosecutors are more likely to remove female jurors and (slightly) less likely to remove white jurors (Column (3)) and defence are more likely to remove white jurors (Column (4)).

After exploring the existence of differences in the patterns of strikes, I am examining whether these differences result in actual variations in the composition of seated juries. In Table 8, I am testing whether the differences between seated juries and jury pools, based on the gender of judges, are statistically significant.

[TABLE 8 APPROXIMATELY HERE]

Overall, the results in Column (3) and (4) indicate that the average composition between seated juries and jury pools is not statistically significant and there does not seem to be a some strong difference across the gender of the judge. How can the results between Table 7 and Table 8 be reconciled? While some statistically significant effects are present in Table 7, they are not strong enough to generate a statistically significant difference in Table 8. However, the differences in Table 8, even if not statistically significant, tend to behave go in the direction suggested by Table 7.

To conclude, there is not enough empirical evidence to suggest that, when there is a female judges, the *voir dire* process selects stricter juries, at least with respect to the observable characteristics. However, it is possible that stricter juries are selected based on *unobservable characteristics*.

These characteristics may not be readily evident in the data collection process, but they could become apparent during jury selection, such as through the tone of voice or the attitudes of potential jurors. While an all republican (or all democrats) jury might raise some eyebrows (and increase the possibility for the judge to be reversed later), female judges might use the *unobservable characteristics* to select harsher juries without the risks of impropriety or appeal. Unfortunately, I cannot test this hypothesis with this dataset, but it could be a new avenue of research.

6.2 Judges Behaviours

The previous literature indicates that differences in judges behaviours could play a role in jury decisions. According to existing works, even the most restraint judges have opinions and "expectations" about the trials. Through their attitudes and non-verbal behaviours, the judges might "leak" his/her ideas towards the jury (e.g., Lenehan and O'Neill (1981), Ekman and Friesen (1969) and Burnett and Badzinski (2005)). According to Hart (1995) and Hart (1992), juries, generally, know what the judges think and tend to side with judges' positions. If female judges, on average, have stricter attitude or a more likely to expect a guilty verdict, these attitudes and expectations might "leak" to the juries and help explain the findings in Table 4.[38] In this Section, I try to investigate the existence of different attitudes/preferences between men and women towards the judiciary system in general (Section 6.2.1) and in the sample (Section 6.2.2).

6.2.1 GSS analysis

As proposed Anwar et al. (2014), I explore the attitude of US women towards the judiciary system by using the *General Social Survey* (GSS)[39] between 2000-2018. Specifically, I select the following the answers "too harsh", "'right" and "too lenient" to the sentence: "Courts dealing with criminals". I implement a linear probability model to explore the relationship between these variables and gender. The results are presented in Table 9.[40]

[TABLE 9 APPROXIMATELY HERE]

The findings in Table 9 indicate that women have harsher attitudes (Column (3)) and less likely to agree with the idea that courts are too harsh (Column (1)) or the right level of harshness (Column (2)). These results indicate that women in the US, in general, are more likely to side with a harsher justice system. Possible explanations could be that women are more risk adverse (e.g.,

[38] Another possible explanation is that jurors react different to similar behaviours based on judges' gender. Educational literature suggests that students react differently to female and male teachers (e.g., Carrell et al. (2010) and Dee (2007)). Unfortunately, I cannot test this hypothesis with this dataset, but it could be a new avenue of research.

[39] Source: GSS.

[40] Summary statistics and full regressions are presented in Table A11 and Table A12 in Section A5 in the Appendix.

Croson and Gneezy (2009) and Byrnes et al. (1999)) or more likely to be (or being perceived as) the victim of certain type of crimes, such as sex crimes, and less likely to be the perpetrator. For example, according to the US Bureau of Justice Statistics, in 2022, the share of violent incidents involving a male offender was 79% while the share involving a female offender was 17%.[41] On the contrary, women were the victims in 3,201,730 (51.4%) violent incidents while man were victims in 3,028,420 (48.6%) violent incidents.

6.2.2 Judges Toughness

The results in Section 6.2.1 are extremely interesting and encouraging. However, female judges are not the average "Jane". They are an extremely selected group of women and there might be some differences with respect to the average woman. Unfortunately, the dataset does not provide information about judges' attitudes during the trials. However, it does provide information about sentencing in case of a guilty verdict. Following the approach proposed by Leslie and Pope (2017), I calculate a leave-out-mean measuring the degree to which each judge deviates from the average crime sentencing. To control for systematic differences across time and place, I use the residuals from regressing sentences on time and district fixed effects. Similar to Leslie and Pope (2017), I calculate this measure with respect to crime types:

$$leniency_j = \frac{Res_{jc}}{N_{jc}} - \frac{Res_c}{N_c} \tag{4}$$

where $leniency_j$ is the measure for leniency for judge j, Res_{jc} the residuals of the regressions for judge j and type of crime c, N_{jc} the number of trials of judge j for type of crime c, Res_c the residuals of the regressions for type of crime c, N_c the number of trials for type of crime c. To compare female and male judges, I plot the measure for female and male judges in Figure 4.

[Figure 4 APPROXIMATELY HERE]

The results in Fig. 4 are quite interesting. The distribution for female judges indicates that women in the sample are, on average, less lenient that their male counterpart, suggesting a tougher attitude, in line with the GSS findings.

In conclusion, both Sections show the existence of some differences in the attitudes of female and male judges during trials which aligns with existing literature. As mentioned earlier, these variations might potentially influence jury decisions, as suggested by Hart (1992). It's important to note that this is not definitive evidence but rather a preliminary descriptive analysis. A more comprehensive research study on this topic is needed.

[41] Source: National Crime Victimization Survey, [United States], 2022 (ICPSR 38603).

7 Conclusion

This article investigates the impact of female judges in jury trials' outcomes. The analysis is implemented in North Carolina between 2010-2012 and the research design relies on fixed effects and judges' rotation across districts. The findings indicate that female judges are more likely to preside jury trials that end with a guilty verdict. The results do not depend on the choice of model and they are robust to a series of checks. I also implement a series of heterogeneity checks based on judge characteristics. The findings indicate that female democratic judges and white judges play an important role in the main results.

Finally, I explore the possible mechanisms behind my findings. The data does not seem to support the presence of an effect of judges' gender on the jury selection process. On the contrary, they indicate that female judges are tougher than their male counterpart. These findings seem to point to the existence of an indirect effect base on the influence of judge on the jury.

References

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson (2012). "The impact of jury race in criminal trials". In: *The Quarterly Journal of Economics* 127.2, pp. 1017–1055.

— (2014). "The role of age in jury selection and trial outcomes". In: *The Journal of Law and Economics* 57.4, pp. 1001–1030.

Bandyopadhyay, Siddhartha and Bryan C McCannon (2014). "The effect of the election of prosecutors on criminal trials". In: *Public Choice* 161.1-2, pp. 141–156.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004). "How much should we trust differences-in-differences estimates?" In: *The Quarterly journal of economics* 119.1, pp. 249–275.

Blanck, Peter David, Robert Rosenthal, and Doris Hazzard Cordell (1985). "THE APPEARANCE OF JUSTICE: JUDGES'VERBAL AND NONVERBAL BEHAVIOUR IN CRIMINAL JURY TRIALS". In: *Stanford Law Review*.

Burnett, Ann and Diane M Badzinski (2005). "Judge nonverbal communication on trial: Do mock trial jurors notice?" In: *Journal of Communication* 55.2, pp. 209–224.

Byrnes, James P, David C Miller, and William D Schafer (1999). "Gender differences in risk taking: A meta-analysis." In: *Psychological Bulletin* 125.3, p. 367.

Cameron, A Colin and Douglas L Miller (2015). "A practitioner's guide to cluster-robust inference". In: *Journal of human resources* 50.2, pp. 317–372.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller (2008). "Bootstrap-based improvements for inference with clustered errors". In: *The review of economics and statistics* 90.3, pp. 414–427.

Carrell, Scott E, Marianne E Page, and James E West (2010). "Sex and science: How professor gender perpetuates the gender gap". In: *The Quarterly journal of economics* 125.3, pp. 1101–1144.

Carter, Andrew V, Kevin T Schnepel, and Douglas G Steigerwald (2017). "Asymptotic behavior of at-test robust to cluster heterogeneity". In: *Review of Economics and Statistics* 99.4, pp. 698–709.

Coviello, Decio and Stefano Gagliarducci (2017). "Tenure in office and public procurement". In: *American Economic Journal: Economic Policy* 9.3, pp. 59–105.

Croson, Rachel and Uri Gneezy (2009). "Gender differences in preferences". In: *Journal of Economic Literature* 47.2, pp. 448–74.

Dee, Thomas S (2007). "Teachers and the gender gaps in student achievement". In: *Journal of Human resources* 42.3, pp. 528–554.

Ekman, Paul and Wallace V Friesen (1969). "Nonverbal leakage and clues to deception". In: *Psychiatry* 32.1, pp. 88–106.

Fischman, Joshua B and Max M Schanzenbach (2011). "Do standards of review matter? The case of federal criminal sentencing". In: *The Journal of Legal Studies* 40.2, pp. 405–437.

Flanagan, Francis X (2015). "Peremptory challenges and jury selection". In: *The Journal of Law and Economics* 58.2, pp. 385–416.

— (2018). "Race, gender, and juries: Evidence from North Carolina". In: *The Journal of Law and Economics* 61.2, pp. 189–214.

Foresta, Alessandra (2022). *Beyond reasonable doubt: the impact of politically independent jurors on jury trials in North Carolina*. Tech. rep. Discussion paper, University of York.

Fox, Richard and Robert Van Sickel (2000). "Gender dynamics and judicial behavior in criminal trial courts: An exploratory study". In: *Justice System Journal* 21.3, pp. 261–280.

Fujiwara, Thomas and Leonard Wantchekon (2013). "Can informed public deliberation overcome clientelism? Experimental evidence from Benin". In: *American Economic Journal: Applied Economics* 5.4, pp. 241–55.

Gilligan, Carol (1982). *In a different voice: Psychological theory and women's development*. Harvard University Press.

Halverson, Andrea M, Mark Hallahan, Allen J Hart, and Robert Rosenthal (1997). "Reducing the biasing effects of judges' nonverbal behavior with simplified jury instruction." In: *Journal of Applied Psychology* 82.4, p. 590.

Hart, Allen J (1995). "Naturally occurring expectation effects." In: *Journal of Personality and Social Psychology* 68.1, p. 109.

Hart, Allen Jay (1992). "On the sobriety of judges: Nonverbal influence in the courtroom." In:

Heß, Simon (2017). "Randomization inference with Stata: A guide and software". In: *The Stata Journal* 17.3, pp. 630–651.

Hoekstra, Mark and Brittany Street (2021). "The effect of own-gender jurors on conviction rates". In: *The Journal of Law and Economics* 64.3, pp. 513–537.

Johnson, Brian D (2014). "Judges on trial: A reexamination of judicial race and gender effects across modes of conviction". In: *Criminal Justice Policy Review* 25.2, pp. 159–184.

Lenehan, Gregory E and Patrick O'Neill (1981). "Reactance and conflict as determinants of judgment in a mock jury experiment". In: *Journal of Applied Social Psychology* 11.3, pp. 231–239.

Leslie, Emily and Nolan G Pope (2017). "The unintended impact of pretrial detention on case outcomes: Evidence from New York City arraignments". In: *The Journal of Law and Economics* 60.3, pp. 529–557.

Nagler, Markus, Marc Piopiunik, and Martin R West (2020). "Weak markets, strong teachers: Recession at career start and teacher effectiveness". In: *Journal of Labor Economics* 38.2, pp. 453–500.

Schanzenbach, Max M and Emerson H Tiller (2008). "Reviewing the sentencing guidelines: Judicial politics, empirical evidence, and reform". In: *The University of Chicago Law Review* 75.2, pp. 715–760.

Songer, Donald R, Sue Davis, and Susan Haire (1994). "A reappraisal of diversification in the federal courts: Gender effects in the courts of appeals". In: *The Journal of Politics* 56.2, pp. 425–439.

Spitzer, Matthew and Eric Talley (2013). "Left, right, and center: Strategic information acquisition and diversity in judicial panels". In: *The Journal of Law, Economics, & Organization* 29.3, pp. 638–680.

Spohn, Cassia (1991). "Decision making in sexual assault cases: Do black and female judges make a difference?" In: *Women & Criminal Justice* 2.1, pp. 83–105.

Steffensmeier, Darrell and Chris Hebert (1999). "Women and men policymakers: Does the judge's gender affect the sentencing of criminal defendants?" In: *Social forces* 77.3, pp. 1163–1196.

Tables

Table 1: Descriptive Statistics

	mean	sd	min	max
Defendant characteristics:				
Def. woman	0.1184	0.32	0.00	1.00
Def. white	0.3357	0.47	0.00	1.00
Def. black	0.6325	0.48	0.00	1.00
Def. other	0.0318	0.18	0.00	1.00
Def. over 30	0.6378	0.48	0.00	1.00
Trials' characteristics:				
Murder charge	0.0230	0.15	0.00	1.00
Robbery charge	0.0830	0.28	0.00	1.00
Drug charge	0.1926	0.39	0.00	1.00
Sex charge	0.1184	0.32	0.00	1.00
Other violent crime charge	0.1413	0.35	0.00	1.00
Property crime charge	0.3004	0.46	0.00	1.00
Other crime charge	0.0795	0.27	0.00	1.00
Weapon charge	0.0618	0.24	0.00	1.00
Dependent variable:				
Guilty	0.7279	0.45	0.00	1.00
Jury pools' characteristics:				
Jury Pool: women (Prop.)	0.5102	0.11	0.14	0.79
Jury Pool: men (Prop.)	0.4588	0.11	0.18	0.81
Jury Pool: unknown gender (Prop.)	0.0311	0.04	0.00	0.25
Jury Pool: black (Prop.)	0.1672	0.14	0.00	0.67
Jury Pool: white (Prop.)	0.6147	0.18	0.05	1.00
Jury Pool: other (Prop.)	0.0200	0.03	0.00	0.19
Jury Pool: unknown race (Prop.)	0.1981	0.12	0.00	0.90
Jury Pool: Republicans (Prop.)	0.2611	0.12	0.00	0.63
Jury Pool: Democrats (Prop.)	0.3218	0.14	0.00	0.83
Jury Pool: Indipendent (Prop.)	0.1683	0.09	0.00	0.48
Jury Pool: unknown political (Prop.)	0.2488	0.12	0.00	0.90
Judge characteristics:				
Experience (per 100)	0.2981	0.08	0.13	0.44
Non-White Judge	0.1537	0.36	0.00	1.00
Nr. trials per Judge	3.4753	2.39	1.00	10.00
Independent Judge	0.1148	0.32	0.00	1.00
Republican Judge	0.2668	0.44	0.00	1.00
Democratic Judge	0.6184	0.49	0.00	1.00
Election (Next year)	0.0813	0.27	0.00	1.00
Non-Elected judges	0.1343	0.34	0.00	1.00
Female Judge	0.0777	0.27	0.00	1.00
Observations	566			
N. District	33			
Year	2010-2012			

Notes: *Def.* is an abbreviation for defendant. *Prop.* is an abbreviation for proportion. *Experience (per 100)* is the number of years since the J.D. of the judge and it is divided by 100. *Nr. trials per Judge* indicates the number of trials per different judges per semester.

Table 2: Differences between Judges by gender

	Female (Mean)	Male (Mean)	Diff.	Std. Error	Obs.
Experience (per 100)	0.2341	0.3035	0.0694***	0.0121	566
Non-White Judge	0.5909	0.1169	-0.4741***	0.0531	566
Nr. trials per Judge	2.2500	3.5785	1.3285***	0.3718	566
Independent Judge	0.0909	0.1169	0.0259	0.0501	566
Republican Judge	0.2045	0.2720	0.0675	0.0695	566
Democratic Judge	0.7045	0.6111	-0.0934	0.0763	566
Election (Next year)	0.2045	0.0709	-0.1337***	0.0426	566
Non-Elected judges	0.1818	0.1303	-0.0515	0.0536	566

Notes: * p < 0.1, ** p < 0.05, *** p < 0.01. *Experience (per 100)* is in year divided by 100. *Nr. trials per Judge* indicates the number of trials per different judges. *Non-White Judge*, *Democratic Judge*, *Republican Judge*, *Election (Next year)*, *Female Judge* and *Non-Elected Judge* are a series of dummies for ethnicity, political affiliation, elections, gender and non-elected judges.

Table 3: Relationship between Judges demographics and Trials Characteristics

dep. var.:	Female Judge					
	(1)			(2)		
	b	se	t	b	se	t
Murder charge	-0.100**	(0.05)	[-1.97]	-0.0421	(0.06)	[-0.65]
Robbery charge	-0.0519	(0.06)	[-0.93]	-0.000463	(0.05)	[-0.01]
Drug charge	0.00143	(0.06)	[0.03]	0.0202	(0.05)	[0.38]
Sex charge	0.0155	(0.06)	[0.24]	0.0708	(0.07)	[1.05]
Other violent crime charge	-0.0113	(0.06)	[-0.20]	0.00430	(0.05)	[0.09]
Property crime charge	-0.0173	(0.05)	[-0.33]	0.0193	(0.04)	[0.43]
Other crime charge	0.0438	(0.07)	[0.63]	0.0550	(0.08)	[0.67]
Jury Pool: women (Prop.)	0.336	(0.25)	[1.35]	0.0932	(0.28)	[0.33]
Jury Pool: men (Prop.)	0.449*	(0.26)	[1.74]	0.243	(0.29)	[0.83]
Jury Pool: black (Prop.)	-0.100	(0.19)	[-0.54]	0.0545	(0.30)	[0.18]
Jury Pool: white (Prop.)	-0.295*	(0.16)	[-1.87]	-0.164	(0.25)	[-0.67]
Jury Pool: other (Prop.)	-0.209	(0.36)	[-0.58]	-0.269	(0.49)	[-0.55]
Jury Pool: Republicans (Prop.)	0.0462	(0.14)	[0.34]	0.0477	(0.21)	[0.23]
Jury Pool: Democrats (Prop.)	0.137	(0.13)	[1.03]	0.174	(0.22)	[0.79]
Jury Pool: Indipendent (Prop.)	0.109	(0.16)	[0.67]	0.0877	(0.25)	[0.35]
Def. woman	0.00424	(0.04)	[0.12]	0.0285	(0.03)	[0.84]
Def. non-white	0.0199	(0.03)	[0.77]	-0.0206	(0.03)	[-0.74]
Def. over 30	0.00575	(0.02)	[0.25]	0.00647	(0.02)	[0.31]
Constant	-0.181	(0.22)	[-0.83]	-0.0829	(0.25)	[-0.34]
Observations			566			566
F-stat			1.767			0.588
Time, Prosecutor, District FE			No			Yes

Notes: OLS regressions. *b* represent the estimates, *se* the robust standard errors in parenthesis and *t* the t-statistics in square brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. *Female Judge* is a dummy for female judges. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

Table 4: Main results

dep. var.:	Guilty			
	(1)	(2)	(3)	(4)
Female Judge	0.147***	0.207**	0.266***	0.277***
	(0.06)	(0.09)	(0.10)	(0.10)
	[2.65]	[2.25]	[2.64]	[2.72]
Experience (per 100)			-2.554	-2.876
			(2.33)	(2.32)
			[-1.10]	[-1.24]
Experience sq. (per 100)			0.000474	0.000518
			(0.00)	(0.00)
			[1.19]	[1.29]
Non-White Judge			-0.0786	-0.112
			(0.08)	(0.08)
			[-0.94]	[-1.36]
Nr. trials per Judge			0.0168	0.0175
			(0.01)	(0.01)
			[1.33]	[1.38]
Independent Judge			-0.0554	-0.0245
			(0.08)	(0.08)
			[-0.69]	[-0.29]
Republican Judge			-0.0000290	-0.00358
			(0.06)	(0.06)
			[-0.00]	[-0.06]
Election (Next year)			0.0130	0.0327
			(0.10)	(0.09)
			[0.14]	[0.35]
Non-Elected judges			-0.0103	-0.0382
			(0.07)	(0.07)
			[-0.15]	[-0.55]
Time, Prosecutor, District FE	No	Yes	Yes	Yes
Controls	No	No	No	Yes
Observations	566	566	566	566
N. District	33	33	33	33
Mean Guilty	0.728	0.728	0.728	0.728
Mean Fem Judge	0.0777	0.0777	0.0777	0.0777

Notes: Robust standard errors in parenthesis and t statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Female Judge* is a dummy for female judges. *Controls* includes: Murder charge, Robbery charge, Drug charge, Sex charge, Other violent crime charge, Property crime charge, Other crime charge, Jury Pool: women (Prop.), Jury Pool: men (Prop.), Jury Pool: black (Prop.), Jury Pool: white (Prop.), Jury Pool: other (Prop.), Jury Pool: Republicans (Prop.), Jury Pool: Democrats (Prop.), Jury Pool: Independent (Prop.), Def. woman, Def. white, Def. black and Def. over 30. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

Table 5: Robustness checks

dep. var.:	Guilty					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Judge	0.277***	0.587***	0.332***	0.272**	0.369***	0.280**
	(0.10)	(0.18)	(0.12)	(0.12)	(0.12)	(0.11)
	[2.72]	[3.18]	[2.85]	[2.23]	[3.10]	[2.56]
Observations	566	348	530	479	507	511
N. District	33	24	23	32	31	33
Logit		Yes				
Removal: "Small" Districts			Yes			
Removal: "Big" Districts				Yes		
Removal: "Small" Workload					Yes	
Removal: "Big" Workload						Yes
Time, Prosecutor, District FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parenthesis and t statistics in square brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. *Guilty* is a dummy for a guilty verdict. *Female Judge* is a dummy for female judges. In Column (1) I present the benchmark model as in Column (4) of Table (4). In Column (2) I present the margins from a Logit regression. *Removal: "Small" Districts* (Column (3)) and *Removal: "Big" Districts* (Column (4)) indicate the removal of the lower 5% and higher 5% of districts in the distribution of the average number of trials per districts. *Removal: "Small" Workload* (Column (5)) and *Removal: "Big" Workload* (Column (6)) indicate the removal of the lower 5% and higher 5% of judges in the distribution of the average number of trials per judge. *Controls* includes: Experience (per 100), Nr. trials per Judge, Non-White Judge, Independent Judge, Republican Judge, Election (Next year), Non-Elected Judge, Murder charge, Robbery charge, Drug charge, Sex charge, Other violent crime charge, Property crime charge, Other crime charge, Jury Pool: women (Prop.), Jury Pool: men (Prop.), Jury Pool: Black (Prop.), Jury Pool: White (Prop.), Jury Pool: Other (Prop.), Jury Pool: Republicans (Prop.), Jury Pool: Democrats (Prop.), Def. woman, Def. white, Def. black and Def. over 30. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

Table 6: Heterogeneity checks over judges characteristics

dep. var.:	Guilty			
	(1)	(2)	(3)	(4)
Female Judge	0.277***	0.368***	0.243*	0.287**
	(0.10)	(0.13)	(0.13)	(0.13)
	[2.72]	[2.92]	[1.86]	[2.18]
Non-White Judge		-0.0826		
		(0.09)		
		[-0.90]		
Female Judge x Non-White Judge		-0.198		
		(0.21)		
		[-0.93]		
Republican Judge			0.00994	
			(0.06)	
			[0.17]	
Independent Judge			-0.0105	
			(0.09)	
			[-0.12]	
Female Judge x Republican Judge			0.0347	
			(0.21)	
			[0.17]	
Female Judge x Independent Judge			-0.103	
			(0.35)	
			[-0.29]	
Election (Next year)				-0.0609
				(0.12)
				[-0.51]
Election (Next year) x Female Judge				0.0745
				(0.18)
				[0.42]
Observations	566	566	566	471
N. District	33	33	33	32
Year, Prosecutor, District FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes

Notes: Robust standard errors in parenthesis and t statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Female Judge* is a dummy for female judges. *Controls* includes: Experience (per 100), Nr. trials per Judge, Non-White Judge, Democratic Judge, Republican Judge, Election (Next year), Non-Elected Judge, Murder charge, Robbery charge, Drug charge, Sex charge, Other violent crime charge, Property crime charge, Other crime charge, Jury Pool: women (Prop.), Jury Pool: men (Prop.), Jury Pool: Black (Prop.), Jury Pool: White (Prop.), Jury Pool: Other (Prop.), Jury Pool: Republicans (Prop.), Jury Pool: Democrats (Prop.), Jury Pool: Independents (Prop.), Def. woman, Def. white, Def. black and Def. over 30. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*. In Column (4) I excluded non-elected judges.

Table 7: Jury Selection Process and Female Judges

dep. var.:	Remove	Remove: judge	Remove: pros.	Remove: defence
	(1)	(2)	(3)	(4)
Female Juror	-0.0205**	0.0168	-0.0466***	0.0297**
	(0.01)	(0.01)	(0.01)	(0.01)
	[-2.15]	[1.37]	[-3.67]	[2.20]
Gender Unknown Juror	0.0739**	0.0668*	0.00425	-0.0755**
	(0.03)	(0.03)	(0.04)	(0.04)
	[2.53]	[1.91]	[0.12]	[-2.11]
Race other Juror	0.00221	0.0515	-0.0553	-0.00397
	(0.04)	(0.05)	(0.05)	(0.04)
	[0.06]	[1.07]	[-1.13]	[-0.09]
White Juror	0.0101	-0.0356*	-0.211***	0.249***
	(0.02)	(0.02)	(0.02)	(0.02)
	[0.66]	[-1.80]	[-9.91]	[12.25]
Race Unknown Juror	0.0111	0.0162	-0.192***	0.162***
	(0.02)	(0.03)	(0.03)	(0.03)
	[0.49]	[0.55]	[-6.16]	[5.07]
Democrat Juror	0.0459***	0.0269	0.0510***	-0.0767***
	(0.01)	(0.02)	(0.02)	(0.02)
	[3.30]	[1.52]	[2.80]	[-3.78]
Independent Juror	0.0319**	0.0402**	0.0101	-0.0503**
	(0.01)	(0.02)	(0.02)	(0.02)
	[2.18]	[2.17]	[0.55]	[-2.39]
Pol.Aff. Unknown Juror	0.0225	0.0173	0.0510**	-0.0556**
	(0.02)	(0.02)	(0.02)	(0.03)
	[1.20]	[0.73]	[2.11]	[-2.01]
Female Juror x Female Judge	0.00374	-0.0735*	0.109***	-0.0353
	(0.03)	(0.04)	(0.04)	(0.04)
	[0.11]	[-1.77]	[2.60]	[-0.81]
Female Judge x Gender Unknown Juror	0.0125	0.101	-0.104	0.00784
	(0.10)	(0.11)	(0.11)	(0.12)
	[0.12]	[0.88]	[-0.99]	[0.07]
Female Judge x Race other Juror	0.227**	-0.0817	-0.0802	0.170
	(0.11)	(0.13)	(0.14)	(0.12)
	[2.02]	[-0.62]	[-0.59]	[1.38]
Female Judge x White Juror	0.0688	-0.0491	-0.142**	0.189***
	(0.05)	(0.07)	(0.07)	(0.07)
	[1.32]	[-0.70]	[-2.13]	[2.73]
Female Judge x Race Unknown Juror	0.219***	-0.0434	-0.107	0.165
	(0.08)	(0.10)	(0.10)	(0.11)
	[2.86]	[-0.42]	[-1.08]	[1.49]
Female Judge x Democrat Juror	-0.0263	0.00870	-0.0638	0.0539
	(0.05)	(0.07)	(0.06)	(0.08)
	[-0.49]	[0.13]	[-1.03]	[0.71]
Female Judge x Independent Juror	-0.0809	-0.0605	0.115	-0.0548
	(0.06)	(0.07)	(0.07)	(0.08)
	[-1.46]	[-0.85]	[1.58]	[-0.69]
Female Judge x Pol.Aff. Unknown Juror	-0.193***	-0.0874	0.0107	0.0640
	(0.07)	(0.09)	(0.09)	(0.11)
	[-2.82]	[-0.95]	[0.12]	[0.59]
Constant	0.385***	0.208***	0.401***	0.177***
	(0.02)	(0.02)	(0.02)	(0.02)
	[22.17]	[9.39]	[17.02]	[7.52]
Trail FE X Female Judges	Yes	Yes	Yes	Yes
Observations	12,427	5,126	5,126	5,126
N. Trials	566	555	555	555
N. District	33	33	33	33

Notes: robust standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. *Excluded category*: whether the prospective jurors was seated on the jury. *Remove* is a dummy for being removed; *Remove: judge*, *Remove: pros* and *Remove: defence* are dummies for being remove by the judge, prosecutors and defence respectively. *Excluded categories*: male jurors, black jurors, republican jurors.

Table 8: Seated juries vs Jury pools

	Female Judge			Male Judge		
	Jury Pool	Seated Jury		Jury Pool	Seated Jury	
	Mean	Mean	Diff.	Mean	Mean	Diff.
	(1)	(2)	(3)	(4)	(5)	(6)
Female Jurors	0.5029***	0.5113***	0.0084	0.5108***	0.5202***	0.0094
	(0.0177)	(0.0232)	(0.0292)	(0.0046)	(0.0061)	(0.0077)
Male Jurors	0.4677***	0.4650***	-0.0027	0.4580***	0.4528***	-0.0051
	(0.0184)	(0.0226)	(0.0292)	(0.0047)	(0.0061)	(0.0076)
Black Jurors	0.2232***	0.2400***	0.0168	0.1625***	0.1619***	-0.0006
	(0.0242)	(0.0301)	(0.0387)	(0.0058)	(0.0065)	(0.0087)
White Jurors	0.5302***	0.5307***	0.0005	0.6218***	0.6258***	0.0040
	(0.0306)	(0.0333)	(0.0452)	(0.0076)	(0.0084)	(0.0113)
Other Jurors	0.0215***	0.0158***	-0.0058	0.0199***	0.0199***	0.0000
	(0.0046)	(0.0054)	(0.0071)	(0.0015)	(0.0018)	(0.0024)
Republican Jurors	0.2190***	0.2103***	-0.0087	0.2647***	0.2755***	0.0108
	(0.0198)	(0.0203)	(0.0284)	(0.0052)	(0.0063)	(0.0082)
Democrat Jurors	0.3643***	0.3578***	-0.0065	0.3182***	0.3119***	-0.0063
	(0.0237)	(0.0290)	(0.0375)	(0.0062)	(0.0070)	(0.0094)
Independent Jurors	0.1584***	0.1656***	0.0072	0.1691***	0.1677***	-0.0014
	(0.0141)	(0.0158)	(0.0212)	(0.0041)	(0.0050)	(0.0065)

Notes: robust standard errors in parenthesis. * p < 0.1, ** p < 0.05, *** p < 0.01. All variables are dummies.

Table 9: GSS regressions results

dep. var.:	Courts are:		
	Too harsh	Right	Too lenient
	(1)	(2)	(3)
Female	-0.030***	-0.026***	0.032***
	(0.00)	(0.00)	(0.01)
	[-7.96]	[-5.82]	[5.32]
Observations	26586	26586	26586
Controls	Yes	Yes	Yes

Notes: robust standard errors in parenthesis and t-statistics in brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. Dependent variables are a dummies. *Controls*: years fixed effects, a dummy for unaffiliated and democrats, a dummy for white, a variable about the higher year of education completed, a variable for the number of children and a dummy for married. *Omitted category*: Republicans.

Figures

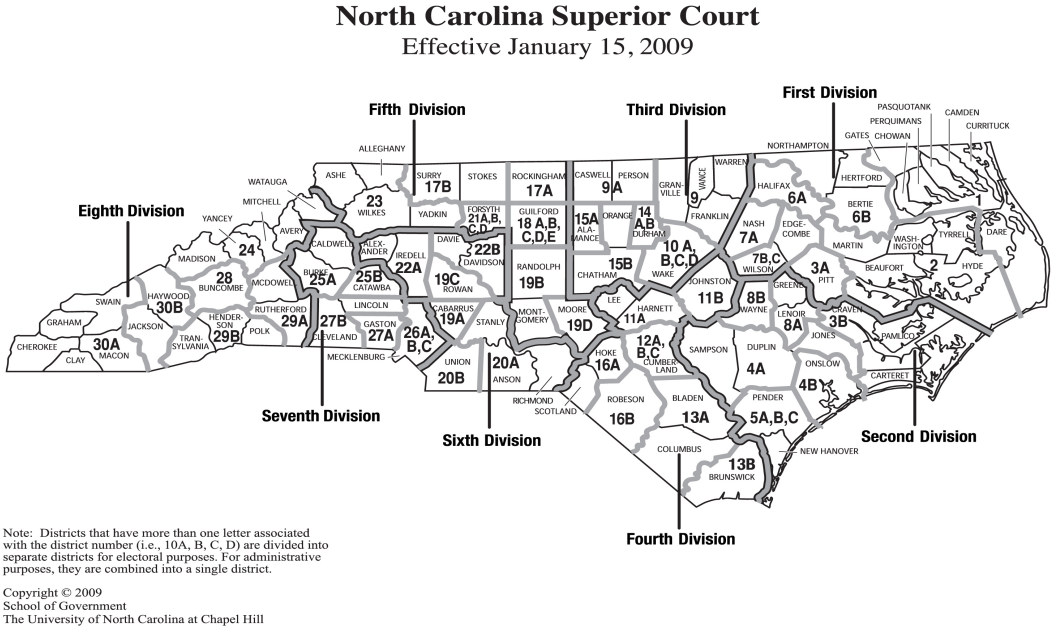

Figure 1: North Carolina Superior Court Map

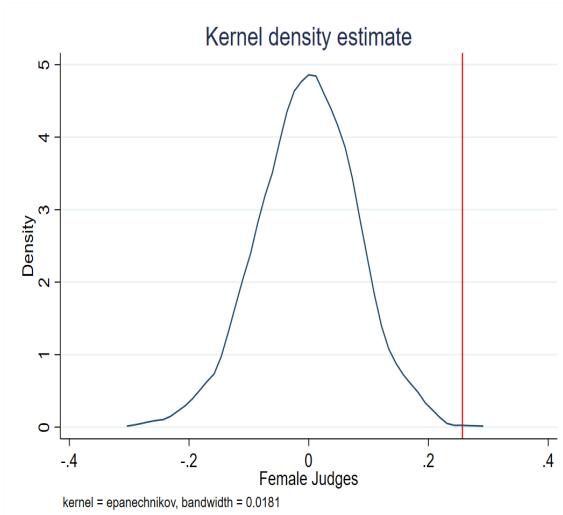

Figure 2: Permutation test. Shown it the Kernel density plot of a randomization inference test for simulated judges' gender assigned using 1000 replications. The red vertical line shows the benchmark model estimated in Column (4) of Table 4.

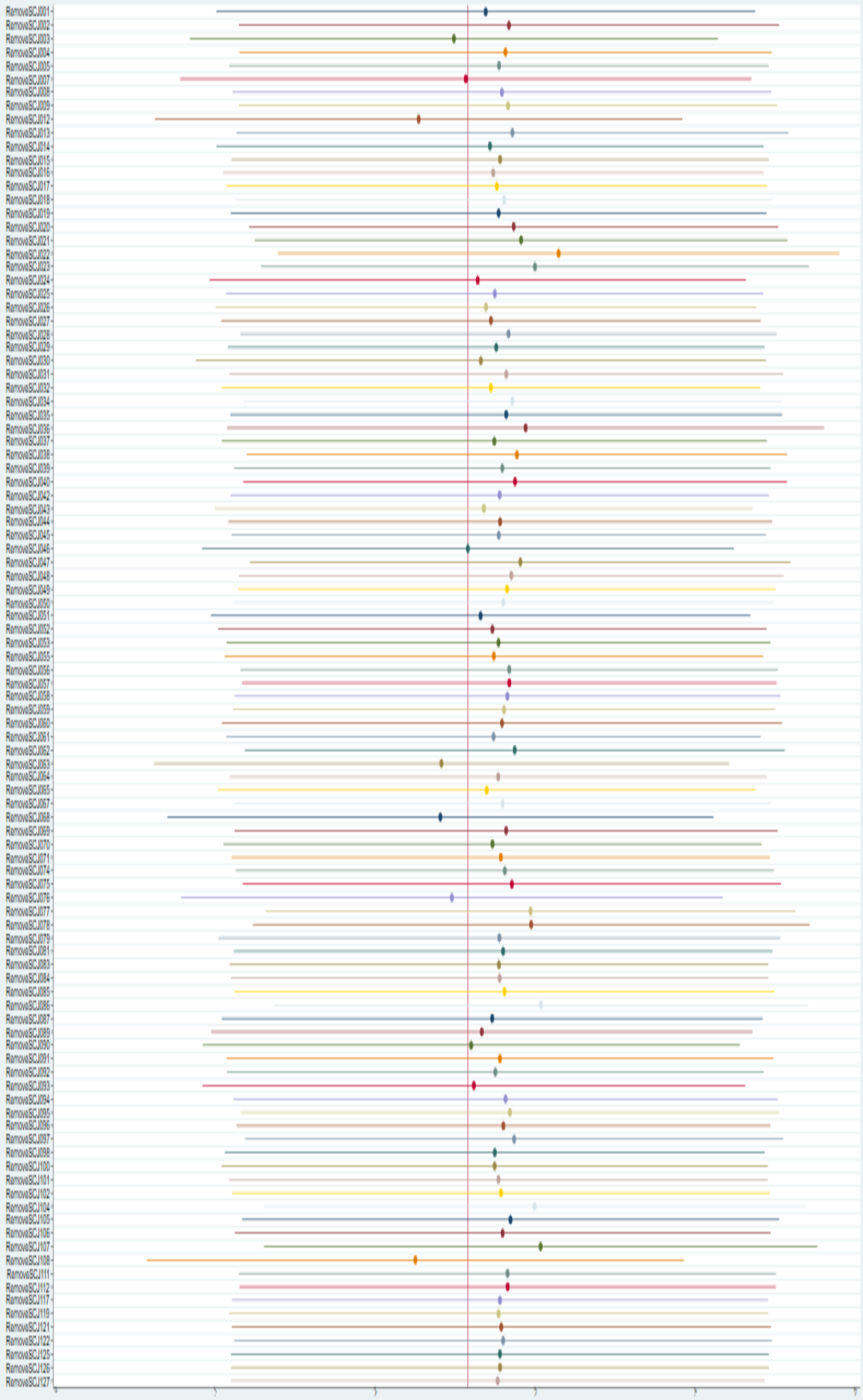

Figure 3: Sensitivity check: re-estimations of the benchmark model (Table 4 Col (4)) by excluding one different judge each time. 90% Confidence intervals. The red link is the coefficient estimated in the benchmark model (Table 4 Col (4)).

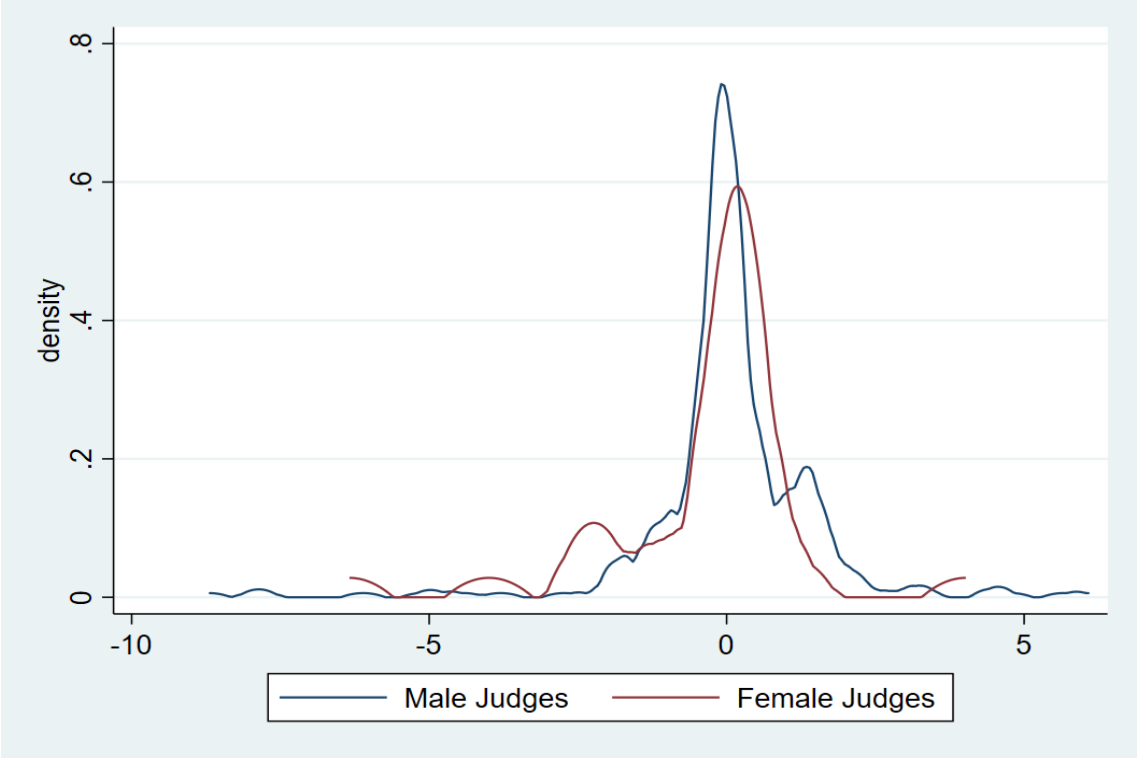

Figure 4: Leniency measure based on Leslie and Pope (2017).

Appendices

A1 Main results: Additional results

Table A1: Main results

dep. var.:	Guilty											
	(1)			(2)			(3)			(4)		
	b	se	t	b	se	t	b	se	t	b	se	t
Female Judge	0.147***	(0.06)	[2.65]	0.207**	(0.09)	[2.25]	0.266***	(0.10)	[2.64]	0.277***	(0.10)	[2.72]
Experience (per 100)							-2.554	(2.33)	[-1.10]	-2.876	(2.32)	[-1.24]
Experience sq. (per 100)							0.000474	(0.00)	[1.19]	0.000518	(0.00)	[1.29]
Non-White Judge							-0.0786	(0.08)	[-0.94]	-0.112	(0.08)	[-1.36]
Nr. trials per Judge							0.0168	(0.01)	[1.33]	0.0175	(0.01)	[1.38]
Independent Judge							-0.0554	(0.08)	[-0.69]	-0.0245	(0.08)	[-0.29]
Republican Judge							-0.0000290	(0.06)	[-0.00]	-0.00358	(0.06)	[-0.06]
Election (Next year)							0.0130	(0.10)	[0.14]	0.0327	(0.09)	[0.35]
Non-Elected judges							-0.0103	(0.07)	[-0.15]	-0.0382	(0.07)	[-0.55]
Jury Pool: women (Prop.)										0.145	(0.60)	[0.24]
Jury Pool: men (Prop.)										-0.193	(0.60)	[-0.32]
Jury Pool: black (Prop.)										-0.297	(0.46)	[-0.64]
Jury Pool: white (Prop.)										0.160	(0.37)	[0.43]
Jury Pool: other (Prop.)										-0.923	(0.87)	[-1.06]
Jury Pool: Republicans (Prop.)										0.171	(0.36)	[0.48]
Jury Pool: Indipendent (Prop.)										-0.260	(0.38)	[-0.69]
Jury Pool: Democrats (Prop.)										0.214	(0.34)	[0.63]
Def. woman										-0.0440	(0.07)	[-0.65]
Def. non-white										0.0283	(0.05)	[0.52]
Def. over 30										-0.00254	(0.05)	[-0.05]
Murder charge										0.0830	(0.24)	[0.35]
Robbery charge										0.181	(0.13)	[1.43]
Drug charge										0.209*	(0.12)	[1.69]
Sex charge										0.197	(0.12)	[1.62]
Other violent crime charge										0.0928	(0.12)	[0.79]
Property crime charge										0.221**	(0.11)	[2.03]
Other crime charge										0.369**	(0.14)	[2.59]
Constant	0.716***	(0.02)	[36.26]	0.712***	(0.02)	[36.36]	0.978***	(0.33)	[2.98]	0.747	(0.63)	[1.18]
Time, Prosecutor, District FE			No			Yes			Yes			Yes
Observations			566			566			566			566
N. District			33			33			33			33
Mean Guilty			0.728			0.728			0.728			0.728
Mean Fem Judge			0.0777			0.0777			0.0777			0.0777

Notes: OLS regressions. *b* represent the estimates, *se* the robust standard errors in parenthesis and *t* the t-statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

Table A2: Main results using controls for the seated juries' characteristics.

dep. var.:	Guilty											
	(1)			(2)			(3)			(4)		
	b	se	t	b	se	t	b	se	t	b	se	t
Female Judge	0.147***	(0.06)	[2.65]	0.207**	(0.09)	[2.25]	0.266***	(0.10)	[2.64]	0.281***	(0.11)	[2.65]
Experience (per 100)							-2.554	(2.33)	[-1.10]	-2.707	(2.30)	[-1.18]
Experience sq. (per 100)							0.000474	(0.00)	[1.19]	0.000497	(0.00)	[1.26]
Non-White Judge							-0.0786	(0.08)	[-0.94]	-0.110	(0.09)	[-1.29]
Nr. trials per Judge							0.0168	(0.01)	[1.33]	0.0176	(0.01)	[1.42]
Independent Judge							-0.0554	(0.08)	[-0.69]	-0.0363	(0.08)	[-0.44]
Republican Judge							-0.0000290	(0.06)	[-0.00]	0.00249	(0.06)	[0.04]
Election (Next year)							0.0130	(0.10)	[0.14]	0.0335	(0.09)	[0.36]
Non-Elected judges							-0.0103	(0.07)	[-0.15]	-0.0238	(0.07)	[-0.35]
Seated Jury: women (Prop.)										0.576	(0.51)	[1.13]
Seated Jury: men (Prop.)										0.370	(0.51)	[0.72]
Seated Jury: Black (Prop.)										-0.274	(0.35)	[-0.78]
Seated Jury: White (Prop.)										-0.119	(0.30)	[-0.40]
Seated Jury: Other (Prop.)										-0.739	(0.60)	[-1.22]
Seated Jury: Republicans (Prop.)										0.491*	(0.29)	[1.71]
Seated Jury: Independent (Prop.)										0.00868	(0.29)	[0.03]
Seated Jury: Democrats (Prop.)										0.167	(0.28)	[0.60]
Def. woman										-0.0392	(0.07)	[-0.60]
Def. non-white										0.0313	(0.05)	[0.58]
Def. over 30										-0.00960	(0.05)	[-0.19]
Murder charge										0.0767	(0.25)	[0.31]
Robbery charge										0.149	(0.13)	[1.13]
Drug charge										0.223*	(0.12)	[1.80]
Sex charge										0.206*	(0.12)	[1.67]
Other violent crime charge										0.0973	(0.12)	[0.83]
Property crime charge										0.224**	(0.11)	[2.03]
Other crime charge										0.357**	(0.14)	[2.51]
Constant	0.716***	(0.02)	[36.26]	0.712***	(0.02)	[36.36]	0.978***	(0.33)	[2.98]	0.279	(0.57)	[0.49]
Time, Prosecutor, District FE			No			Yes			Yes			Yes
Observations			566			566			566			566
N. District			33			33			33			33
Mean Guilty			0.728			0.728			0.728			0.728
Mean Fem Judge			0.0777			0.0777			0.0777			0.0777

Notes: OLS regressions. *b* represent the estimates, *se* the robust standard errors in parenthesis and *t* the t-statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Seated Jury: unknown gender (Prop.)*, *Seated Jury: unknown race (Prop.)* and *Seated Jury: unknown political (Prop.)*.

A2 Robustness checks: Additional results

Table A3: Logit model regressions

dep. var.:	Guilty											
	(1)			(2)			(3)			(4)		
	b	se	t	b	se	t	b	se	t	b	se	t
Female Judge	0.181**	(0.09)	[2.06]	0.421**	(0.18)	[2.36]	0.492***	(0.18)	[2.67]	0.587***	(0.18)	[3.18]
Experience (per 100)							-3.310	(3.11)	[-1.07]	-3.917	(3.13)	[-1.25]
Experience sq. (per 100)							0.000645	(0.00)	[1.17]	0.000738	(0.00)	[1.33]
Non-Elected judges							-0.0504	(0.10)	[-0.50]	-0.0637	(0.09)	[-0.71]
Non-White Judge							-0.0926	(0.10)	[-0.92]	-0.120	(0.10)	[-1.24]
Nr. trials per Judge							0.0233	(0.02)	[1.54]	0.0263*	(0.02)	[1.67]
Independent Judge							-0.0954	(0.10)	[-0.93]	-0.101	(0.10)	[-1.04]
Republican Judge							0.00827	(0.07)	[0.12]	0.0217	(0.08)	[0.29]
Election (Next year)							0.0208	(0.13)	[0.16]	0.00140	(0.12)	[0.01]
Jury Pool: women (Prop.)										0.473	(0.80)	[0.59]
Jury Pool: men (Prop.)										-0.0398	(0.82)	[-0.05]
Jury Pool: black (Prop.)										-0.287	(0.57)	[-0.50]
Jury Pool: other (Prop.)										-1.003	(1.08)	[-0.93]
Jury Pool: white (Prop.)										0.192	(0.50)	[0.38]
Jury Pool: Republicans (Prop.)										0.353	(0.49)	[0.72]
Jury Pool: Indipendent (Prop.)										-0.586	(0.53)	[-1.11]
Jury Pool: Democrats (Prop.)										0.225	(0.49)	[0.46]
Def. woman										-0.0388	(0.08)	[-0.48]
Def. non-white										0.0563	(0.07)	[0.85]
Def. over 30										0.0113	(0.06)	[0.19]
Murder charge										0.170	(0.29)	[0.58]
Robbery charge										0.243	(0.15)	[1.62]
Drug charge										0.338**	(0.15)	[2.25]
Sex charge										0.321**	(0.16)	[2.07]
Other violent crime charge										0.199	(0.13)	[1.49]
Property crime charge										0.349**	(0.14)	[2.52]
Other crime charge										0.625***	(0.16)	[3.80]
Time, Prosecutor, District FE			Yes			Yes			Yes			Yes
Observations			566			348			348			348
N. District			33			24			24			24

Notes: Margins. *b* represent the estimates, *se* the robust standard errors in parenthesis and *t* the t-statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

Table A4: Main regression removing the district in the low 5% of the average trial distribution.

dep. var.:	Guilty											
	(1)			(2)			(3)			(4)		
	b	se	t	b	se	t	b	se	t	b	se	t
Female Judge	0.128**	(0.06)	[2.00]	0.218**	(0.10)	[2.12]	0.315***	(0.11)	[2.75]	0.326***	(0.12)	[2.82]
Experience (per 100)							-2.038	(2.44)	[-0.84]	-2.217	(2.42)	[-0.91]
Experience sq. (per 100)							0.000403	(0.00)	[0.96]	0.000422	(0.00)	[1.01]
Non-White Judge							-0.109	(0.09)	[-1.19]	-0.131	(0.09)	[-1.46]
Nr. trials per Judge							0.0208	(0.01)	[1.60]	0.0229*	(0.01)	[1.76]
Independent Judge							-0.0441	(0.08)	[-0.53]	-0.0180	(0.09)	[-0.21]
Republican Judge							0.00832	(0.06)	[0.14]	0.00598	(0.06)	[0.10]
Election (Next year)							-0.00171	(0.11)	[-0.02]	0.00404	(0.10)	[0.04]
Non-Elected judges							-0.0148	(0.08)	[-0.20]	-0.0460	(0.07)	[-0.63]
Jury Pool: women (Prop.)										0.109	(0.61)	[0.18]
Jury Pool: men (Prop.)										-0.211	(0.61)	[-0.35]
Jury Pool: black (Prop.)										-0.337	(0.52)	[-0.65]
Jury Pool: white (Prop.)										0.146	(0.41)	[0.35]
Jury Pool: other (Prop.)										-0.720	(0.90)	[-0.80]
Jury Pool: Republicans (Prop.)										0.232	(0.39)	[0.60]
Jury Pool: Indipendent (Prop.)										-0.346	(0.42)	[-0.83]
Jury Pool: Democrats (Prop.)										0.340	(0.39)	[0.86]
Def. woman										-0.0763	(0.07)	[-1.09]
Def. non-white										0.00496	(0.06)	[0.09]
Def. over 30										0.00996	(0.05)	[0.19]
Murder charge										0.0960	(0.25)	[0.39]
Robbery charge										0.200	(0.13)	[1.53]
Drug charge										0.230*	(0.13)	[1.78]
Sex charge										0.220*	(0.13)	[1.70]
Other violent crime charge										0.131	(0.12)	[1.10]
Property crime charge										0.230**	(0.11)	[2.03]
Other crime charge										0.403***	(0.15)	[2.71]
Constant	0.710***	(0.02)	[34.67]	0.704***	(0.02)	[34.81]	0.867**	(0.35)	[2.51]	0.600	(0.64)	[0.94]
Time, Prosecutor, District FE			Yes			Yes			Yes			Yes
Observations			530			530			530			530
N. District			23			23			23			23

Notes: *b* represent the estimates, *se* the robust standard errors in parenthesis and *t* the t-statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

Table A5: Main regression removing the district in the top 5% of the average trial distribution.

dep. var.:	Guilty											
	(1)			(2)			(3)			(4)		
	b	se	t	b	se	t	b	se	t	b	se	t
Female Judge	0.141**	(0.06)	[2.24]	0.210*	(0.11)	[1.88]	0.264**	(0.12)	[2.21]	0.268**	(0.12)	[2.21]
Experience (per 100)							-1.805	(2.59)	[-0.70]	-2.097	(2.59)	[-0.81]
Experience sq. (per 100)							0.000352	(0.00)	[0.79]	0.000395	(0.00)	[0.89]
Non-White Judge							-0.0760	(0.09)	[-0.82]	-0.108	(0.09)	[-1.18]
Nr. trials per Judge							0.0148	(0.01)	[1.06]	0.0171	(0.01)	[1.21]
Independent Judge							0.0354	(0.09)	[0.39]	0.0568	(0.09)	[0.61]
Republican Judge							0.0138	(0.06)	[0.21]	0.00145	(0.07)	[0.02]
Election (Next year)							0.0226	(0.11)	[0.20]	0.0335	(0.11)	[0.32]
Non-Elected judges							-0.0131	(0.07)	[-0.18]	-0.0407	(0.07)	[-0.55]
Jury Pool: women (Prop.)										0.166	(0.75)	[0.22]
Jury Pool: men (Prop.)										-0.179	(0.75)	[-0.24]
Jury Pool: black (Prop.)										-0.272	(0.50)	[-0.55]
Jury Pool: white (Prop.)										0.00596	(0.42)	[0.01]
Jury Pool: other (Prop.)										-1.029	(1.09)	[-0.95]
Jury Pool: Republicans (Prop.)										0.258	(0.41)	[0.62]
Jury Pool: Indipendent (Prop.)										-0.119	(0.43)	[-0.27]
Jury Pool: Democrats (Prop.)										0.251	(0.39)	[0.65]
Def. woman										-0.0290	(0.07)	[-0.42]
Def. non-white										0.0383	(0.06)	[0.67]
Def. over 30										-0.0112	(0.06)	[-0.19]
Murder charge										0.00310	(0.29)	[0.01]
Robbery charge										0.127	(0.14)	[0.92]
Drug charge										0.181	(0.13)	[1.35]
Sex charge										0.161	(0.13)	[1.20]
Other violent crime charge										0.0454	(0.13)	[0.36]
Property crime charge										0.198*	(0.12)	[1.68]
Other crime charge										0.360**	(0.15)	[2.38]
Constant	0.701***	(0.02)	[32.06]	0.695***	(0.02)	[31.55]	0.847**	(0.36)	[2.33]	0.646	(0.78)	[0.82]
Time, Prosecutor, District FE			Yes			Yes			Yes			Yes
Observations			479			479			479			479
N. District			32			32			32			32

Notes: *b* represent the estimates, *se* the robust standard errors in parenthesis and *t* the t-statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

Table A6: Main regression removing the judge in the low 5% of the average trial distribution.

dep. var.:	Guilty											
	(1)			(2)			(3)			(4)		
	b	se	t	b	se	t	b	se	t	b	se	t
Female Judge	0.152**	(0.06)	[2.48]	0.232**	(0.11)	[2.09]	0.294**	(0.12)	[2.44]	0.367***	(0.12)	[3.10]
Experience (per 100)							-0.829	(2.62)	[-0.32]	-1.455	(2.60)	[-0.56]
Experience sq. (per 100)							0.000200	(0.00)	[0.45]	0.000297	(0.00)	[0.67]
Non-White Judge							-0.0454	(0.10)	[-0.46]	-0.0826	(0.09)	[-0.89]
Nr. trials per Judge							0.0240*	(0.01)	[1.81]	0.0243*	(0.01)	[1.81]
Independent Judge							-0.121	(0.09)	[-1.41]	-0.0888	(0.09)	[-0.97]
Republican Judge							0.0166	(0.06)	[0.26]	0.00444	(0.06)	[0.07]
Election (Next year)							-0.00862	(0.10)	[-0.08]	0.00547	(0.10)	[0.06]
Non-Elected judges							0.0306	(0.08)	[0.36]	0.0000146	(0.08)	[0.00]
Jury Pool: women (Prop.)										0.145	(0.62)	[0.23]
Jury Pool: men (Prop.)										-0.146	(0.62)	[-0.24]
Jury Pool: black (Prop.)										-0.329	(0.51)	[-0.64]
Jury Pool: white (Prop.)										0.265	(0.40)	[0.66]
Jury Pool: other (Prop.)										-0.806	(0.90)	[-0.89]
Jury Pool: Republicans (Prop.)										0.0910	(0.38)	[0.24]
Jury Pool: Indipendent (Prop.)										-0.487	(0.41)	[-1.20]
Jury Pool: Democrats (Prop.)										0.119	(0.36)	[0.33]
Def. woman										-0.0786	(0.07)	[-1.11]
Def. non-white										0.0287	(0.06)	[0.49]
Def. over 30										-0.0136	(0.05)	[-0.25]
Murder charge										0.0864	(0.27)	[0.31]
Robbery charge										0.203	(0.13)	[1.55]
Drug charge										0.199	(0.13)	[1.48]
Sex charge										0.164	(0.13)	[1.26]
Other violent crime charge										0.0501	(0.12)	[0.41]
Property crime charge										0.231**	(0.12)	[2.00]
Other crime charge										0.459***	(0.15)	[3.04]
Constant	0.709***	(0.02)	[34.33]	0.703***	(0.02)	[33.62]	0.679*	(0.38)	[1.81]	0.509	(0.64)	[0.79]
Time, Prosecutor, District FE			Yes			Yes			Yes			Yes
Observations			521			507			507			507
N. District			31			31			31			31

Notes: *b* represent the estimates, *se* the robust standard errors in parenthesis and *t* the t-statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

Table A7: Main regression removing judges in the top 5% of the average trial distribution.

dep. var.:	Guilty											
	(1)			(2)			(3)			(4)		
	b	se	t	b	se	t	b	se	t	b	se	t
Female Judge	0.154***	(0.06)	[2.76]	0.230**	(0.10)	[2.34]	0.275***	(0.10)	[2.62]	0.274**	(0.11)	[2.52]
Experience (per 100)							-2.016	(2.55)	[-0.79]	-2.350	(2.53)	[-0.93]
Experience sq. (per 100)							0.000389	(0.00)	[0.90]	0.000428	(0.00)	[0.99]
Non-White Judge							-0.0818	(0.09)	[-0.95]	-0.111	(0.08)	[-1.32]
Nr. trials per Judge							0.00791	(0.02)	[0.50]	0.00789	(0.02)	[0.48]
Independent Judge							-0.0332	(0.08)	[-0.40]	-0.0133	(0.09)	[-0.15]
Republican Judge							0.0296	(0.06)	[0.49]	0.0263	(0.06)	[0.42]
Election (Next year)							0.0787	(0.10)	[0.77]	0.0838	(0.10)	[0.84]
Non-Elected judges							0.00233	(0.07)	[0.03]	-0.0218	(0.07)	[-0.31]
Jury Pool: women (Prop.)										-0.00461	(0.63)	[-0.01]
Jury Pool: men (Prop.)										-0.183	(0.63)	[-0.29]
Jury Pool: black (Prop.)										-0.266	(0.49)	[-0.55]
Jury Pool: white (Prop.)										0.183	(0.38)	[0.48]
Jury Pool: other (Prop.)										-1.124	(0.90)	[-1.25]
Jury Pool: Republicans (Prop.)										0.112	(0.36)	[0.31]
Jury Pool: Indipendent (Prop.)										-0.239	(0.40)	[-0.60]
Jury Pool: Democrats (Prop.)										0.189	(0.35)	[0.54]
Def. woman										-0.0499	(0.07)	[-0.71]
Def. non-white										0.0231	(0.06)	[0.39]
Def. over 30										0.00376	(0.05)	[0.07]
Murder charge										0.116	(0.24)	[0.49]
Robbery charge										0.197	(0.13)	[1.56]
Drug charge										0.263**	(0.12)	[2.11]
Sex charge										0.194	(0.13)	[1.54]
Other violent crime charge										0.135	(0.12)	[1.14]
Property crime charge										0.225**	(0.11)	[2.06]
Other crime charge										0.430***	(0.15)	[2.92]
Constant	0.709***	(0.02)	[34.08]	0.709***	(0.02)	[34.04]	0.914**	(0.36)	[2.57]	0.755	(0.67)	[1.13]
Time, Prosecutor, District FE			Yes			Yes			Yes			Yes
Observations			522			511			511			511
N. District			33			33			33			33

Notes: *b* represent the estimates, *se* the robust standard errors in parenthesis and *t* the t-statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories: *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

A3 Heterogeneity checks: Additional results

Table A8: Heterogeneity checks over judges characteristics with controls

dep. var.:	Guilty											
	(1)			(2)			(3)			(4)		
	b	se	t	b	se	t	b	se	t	b	se	t
Female Judge	0.277***	(0.10)	[2.72]	0.368***	(0.13)	[2.92]	0.243*	(0.13)	[1.86]	0.287**	(0.13)	[2.18]
Experience (per 100)	-2.876	(2.32)	[-1.24]	-3.054	(2.33)	[-1.31]	-3.113	(2.39)	[-1.30]	-4.302*	(2.58)	[-1.66]
Experience sq. (per 100)	0.000518	(0.00)	[1.29]	0.000548	(0.00)	[1.36]	0.000566	(0.00)	[1.37]	0.000791*	(0.00)	[1.76]
Non-White Judge	-0.112	(0.08)	[-1.36]									
Nr. trials per Judge	0.0175	(0.01)	[1.38]	0.0185	(0.01)	[1.47]	0.0189	(0.01)	[1.51]	0.0123	(0.02)	[0.81]
Democratic Judge	0.0245	(0.08)	[0.29]									
Republican Judge	0.0209	(0.09)	[0.24]	-0.00630	(0.06)	[-0.11]				0.00475	(0.06)	[0.07]
Election (Next year)	0.0327	(0.09)	[0.35]	0.0299	(0.09)	[0.32]	-0.00746	(0.09)	[-0.08]			
Non-Elected judges	-0.0382	(0.07)	[-0.55]	-0.0544	(0.07)	[-0.77]	-0.0298	(0.07)	[-0.43]			
Jury Pool: women (Prop.)	0.145	(0.60)	[0.24]	0.174	(0.60)	[0.29]	0.131	(0.59)	[0.22]	0.393	(0.73)	[0.54]
Jury Pool: men (Prop.)	-0.193	(0.60)	[-0.32]	-0.160	(0.59)	[-0.27]	-0.196	(0.59)	[-0.33]	-0.101	(0.72)	[-0.14]
Jury Pool: other (Prop.)	-0.923	(0.87)	[-1.06]	-0.971	(0.87)	[-1.11]	-0.995	(0.86)	[-1.15]	-0.748	(0.97)	[-0.77]
Jury Pool: black (Prop.)	-0.297	(0.46)	[-0.64]	-0.300	(0.47)	[-0.64]	-0.287	(0.47)	[-0.61]	-0.424	(0.52)	[-0.82]
Jury Pool: white (Prop.)	0.160	(0.37)	[0.43]	0.177	(0.37)	[0.47]	0.122	(0.37)	[0.33]	0.0235	(0.41)	[0.06]
Jury Pool: Republicans (Prop.)	0.171	(0.36)	[0.48]	0.169	(0.36)	[0.47]	0.211	(0.36)	[0.59]	0.423	(0.41)	[1.02]
Jury Pool: Democrats (Prop.)	0.214	(0.34)	[0.63]	0.222	(0.34)	[0.65]	0.220	(0.34)	[0.64]	0.456	(0.40)	[1.15]
Jury Pool: Indipendent (Prop.)	-0.260	(0.38)	[-0.69]	-0.262	(0.38)	[-0.69]	-0.238	(0.38)	[-0.63]	0.0209	(0.43)	[0.05]
Def. woman	-0.0440	(0.07)	[-0.65]	-0.0402	(0.07)	[-0.60]	-0.0448	(0.07)	[-0.66]	-0.0244	(0.08)	[-0.32]
Def. Non-White	0.0283	(0.05)	[0.52]	0.0311	(0.05)	[0.57]	0.0238	(0.05)	[0.44]	0.00735	(0.06)	[0.11]
Def. Over 30	-0.00254	(0.05)	[-0.05]	-0.00325	(0.05)	[-0.06]	-0.00246	(0.05)	[-0.05]	-0.00390	(0.06)	[-0.07]
Murder charge	0.0830	(0.24)	[0.35]	0.0784	(0.24)	[0.33]	0.0759	(0.25)	[0.30]	0.0408	(0.26)	[0.16]
Robbery charge	0.181	(0.13)	[1.43]	0.180	(0.13)	[1.43]	0.177	(0.13)	[1.37]	0.135	(0.14)	[0.97]
Drug charge	0.209*	(0.12)	[1.69]	0.212*	(0.12)	[1.72]	0.198	(0.13)	[1.56]	0.152	(0.13)	[1.14]
Sex charge	0.197	(0.12)	[1.62]	0.200	(0.12)	[1.65]	0.193	(0.12)	[1.54]	0.176	(0.14)	[1.28]
Other violent crime charge	0.0928	(0.12)	[0.79]	0.0977	(0.12)	[0.83]	0.0902	(0.12)	[0.76]	0.0909	(0.12)	[0.73]
Property crime charge	0.221**	(0.11)	[2.03]	0.216**	(0.11)	[1.99]	0.214*	(0.11)	[1.93]	0.193*	(0.12)	[1.66]
Other crime charge	0.369**	(0.14)	[2.59]	0.375***	(0.14)	[2.62]	0.352**	(0.15)	[2.39]	0.225	(0.15)	[1.46]
Non-White Judge				-0.0826	(0.09)	[-0.90]						
Female Judge x Non-White Judge				-0.198	(0.21)	[-0.93]						
Independent Judge				-0.0220	(0.08)	[-0.26]				0.0121	(0.12)	[0.10]
Republican Judge							0.00994	(0.06)	[0.17]			
Independent Judge							-0.0105	(0.09)	[-0.12]			
Female Judge x Republican Judge							0.0347	(0.21)	[0.17]			
Female Judge x Independent Judge							-0.103	(0.35)	[-0.29]			
Election (Next year)										-0.0609	(0.12)	[-0.51]
Election (Next year) x Female Judge										0.0745	(0.18)	[0.42]
Constant	0.722	(0.64)	[1.13]	0.724	(0.63)	[1.15]	0.775	(0.63)	[1.24]	0.714	(0.77)	[0.93]
Observations			566			566			566			471
N. District			33			33			33			32
Time, Prosecutor, District FE			Yes			Yes			Yes			Yes

Notes: Robust standard errors in parenthesis and t statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Female Judge* is a dummy for female judges. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. *Excluded category*: weapon.

A3.1 Additional Heterogeneity checks

Many works in the existing literature suggests that there is a lot of heterogeneity based on the types of crimes and jurors' and defendants' characteristics (e.g., Anwar et al., 2012 and Hoekstra and Street (2021)). First, to test the possible effect of different crimes, I generate a set of dummies for the different types of crimes: *Violent crimes* ,[42] *Property Crimes* and *Other Crimes*.[43] In Column (2) of Table A9, I interact the variable *Female Judge* with the different types of crime. Secondly, in Columns (3) and (4) of Table A9 I interact the variable *Female Judge* with jurors' characteristics

[42] I define as *Violent Crimes* as a dummy equal to one if there is one of the following offences: murder, robbery, drug, sex and other violent crimes.

[43] *Other crimes* include drug and other crimes.

and defendants' characteristics, respectively. In Column (1) I present the benchmark model.

Overall, there are only two interesting results. In Column (2), there is a positive and significant of the interaction between *Female Judge* and *Def. Non-White* while in Column (3) there is a positive and significant of the interaction between *Female Judge* and *Jury pool: other (Prop.)*.

Table A9: Additional Heterogeneity checks: jurors' and defendants' characteristics and types of crime.

dep. var.:	Guilty											
	(1)			(2)			(3)			(4)		
	b	se	t	b	se	t	b	se	t	b	se	t
Female Judge	0.277***	(0.10)	[2.72]	0.0452	(0.16)	[0.28]	1.423	(1.80)	[0.79]	0.195	(0.16)	[1.26]
Experience (per 100)	-2.876	(2.32)	[-1.24]	-2.565	(2.33)	[-1.10]	-1.888	(2.42)	[-0.78]	-3.073	(2.36)	[-1.30]
Experience sq. (per 100)	0.000518	(0.00)	[1.29]	0.000471	(0.00)	[1.17]	0.000360	(0.00)	[0.86]	0.000559	(0.00)	[1.38]
Non-White Judge	-0.112	(0.08)	[-1.36]	-0.125	(0.08)	[-1.50]	-0.0952	(0.08)	[-1.15]	-0.110	(0.08)	[-1.31]
Nr. trials per Judge	0.0175	(0.01)	[1.38]	0.0177	(0.01)	[1.41]	0.0186	(0.01)	[1.46]	0.0176	(0.01)	[1.37]
Independent Judge	-0.0245	(0.08)	[-0.29]	-0.0295	(0.08)	[-0.35]	-0.0503	(0.09)	[-0.58]	-0.0222	(0.08)	[-0.26]
Republican Judge	-0.00358	(0.06)	[-0.06]	0.000820	(0.06)	[0.01]	-0.00542	(0.06)	[-0.09]	-0.00411	(0.06)	[-0.07]
Election (Next year)	0.0327	(0.09)	[0.35]	0.0249	(0.09)	[0.27]	-0.00662	(0.10)	[-0.07]	0.0157	(0.10)	[0.16]
Non-Elected judges	-0.0382	(0.07)	[-0.55]	-0.0297	(0.07)	[-0.43]	-0.0385	(0.07)	[-0.56]	-0.0431	(0.07)	[-0.62]
Jury Pool: women (Prop.)	0.145	(0.60)	[0.24]	0.179	(0.60)	[0.30]	0.185	(0.63)	[0.30]	0.204	(0.59)	[0.35]
Jury Pool: men (Prop.)	-0.193	(0.60)	[-0.32]	-0.129	(0.60)	[-0.22]	-0.0474	(0.63)	[-0.08]	-0.147	(0.59)	[-0.25]
Jury Pool: other (Prop.)	-0.923	(0.87)	[-1.06]	-0.859	(0.88)	[-0.98]	-1.287	(0.93)	[-1.38]	-1.055	(0.88)	[-1.20]
Jury Pool: black (Prop.)	-0.297	(0.46)	[-0.64]	-0.237	(0.47)	[-0.50]	-0.222	(0.53)	[-0.42]	-0.395	(0.47)	[-0.84]
Jury Pool: white (Prop.)	0.160	(0.37)	[0.43]	0.179	(0.38)	[0.47]	0.194	(0.41)	[0.47]	0.0715	(0.38)	[0.19]
Jury Pool: Republicans (Prop.)	0.171	(0.36)	[0.48]	0.155	(0.36)	[0.43]	0.196	(0.37)	[0.53]	0.269	(0.37)	[0.74]
Jury Pool: Democrats (Prop.)	0.214	(0.34)	[0.63]	0.165	(0.35)	[0.48]	0.150	(0.37)	[0.40]	0.318	(0.35)	[0.91]
Jury Pool: Indipendent (Prop.)	-0.260	(0.38)	[-0.69]	-0.278	(0.38)	[-0.74]	-0.303	(0.42)	[-0.73]	-0.124	(0.38)	[-0.33]
Def. woman	-0.0440	(0.07)	[-0.65]	-0.0529	(0.07)	[-0.75]	-0.0569	(0.07)	[-0.82]	-0.0420	(0.06)	[-0.65]
Def. Non-White	0.0283	(0.05)	[0.52]	0.00668	(0.06)	[0.12]	0.0230	(0.05)	[0.42]	0.0164	(0.05)	[0.31]
Def. Over 30	-0.00254	(0.05)	[-0.05]	0.00336	(0.05)	[0.06]	-0.00342	(0.05)	[-0.06]	-0.00329	(0.05)	[-0.06]
Murder charge	0.0830	(0.24)	[0.35]	0.0867	(0.24)	[0.36]	0.0588	(0.24)	[0.24]			
Robbery charge	0.181	(0.13)	[1.43]	0.177	(0.13)	[1.39]	0.193	(0.13)	[1.52]			
Drug charge	0.209*	(0.12)	[1.69]	0.202	(0.12)	[1.63]	0.215*	(0.12)	[1.74]			
Sex charge	0.197	(0.12)	[1.62]	0.185	(0.12)	[1.51]	0.192	(0.12)	[1.55]			
Other violent crime charge	0.0928	(0.12)	[0.79]	0.0872	(0.12)	[0.74]	0.0992	(0.12)	[0.85]			
Property crime charge	0.221**	(0.11)	[2.03]	0.210*	(0.11)	[1.92]	0.222**	(0.11)	[2.02]			
Other crime charge	0.369**	(0.14)	[2.59]	0.349**	(0.15)	[2.37]	0.355**	(0.14)	[2.50]			
Female Judge x Def. woman				-0.143	(0.23)	[-0.63]						
Female Judge x Def. Non-White				0.414**	(0.19)	[2.23]						
Female Judge x Def. Over 30				-0.0274	(0.13)	[-0.21]						
Female Judge x Jury Pool: women (Prop.)							-1.319	(2.02)	[-0.65]			
Female Judge x Jury Pool: men (Prop.)							-1.836	(1.99)	[-0.92]			
Female Judge x Jury Pool: black (Prop.)							-0.255	(1.15)	[-0.22]			
Female Judge x Jury Pool: other (Prop.)							6.421**	(2.64)	[2.43]			
Female Judge x Jury Pool: white (Prop.)							-0.626	(0.90)	[-0.70]			
Female Judge x Jury Pool: Republicans (Prop.)							0.884	(1.23)	[0.72]			
Female Judge x Jury Pool: Democrats (Prop.)							1.008	(1.19)	[0.84]			
Female Judge x Jury Pool: Indepedents (Prop.)							0.563	(1.11)	[0.51]			
Violent Crimes										-0.176**	(0.08)	[-2.25]
Property Crimes										-0.0625	(0.08)	[-0.83]
Female Judge x Violent Crimes										0.201	(0.17)	[1.18]
Female Judge x Property Crimes										0.0803	(0.18)	[0.44]
Constant	0.747	(0.63)	[1.18]	0.669	(0.64)	[1.05]	0.513	(0.66)	[0.78]	0.998	(0.64)	[1.57]
Observations			566			566			566			566
N. District			33			33			33			33
Time, Prosecutor, District FE			Yes			Yes			Yes			Yes
Controls			Yes			Yes			Yes			Yes

Notes: *b* represent the estimates, *se* the robust standard errors in parenthesis and *t* the t-statistics in square brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. *Guilty* is a dummy for a guilty verdict. *Female Judge* is a dummy for female judges. Excluded categories in Columns (1)-(3): *Democratic Judge*, *Weapon charge*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*. *Def.* is an abbreviation for defendant and *Prop.* is an abbreviation for proportion. Excluded categories in Column (4): *Democratic Judge*, *Other charges*, *Jury Pool: unknown gender (Prop.)*, *Jury Pool: unknown race (Prop.)* and *Jury Pool: unknown political (Prop.)*.

A4 Jury selection process and seated jury composition: Additional results

Table A10: Descriptive Statistics

	mean	sd	min	max
Dependent Variables:				
Removed	0.4130	0.49	0.00	1.00
Removed by judge	0.0884	0.28	0.00	1.00
Removed by prosecutor	0.0983	0.30	0.00	1.00
Removed by defence	0.1388	0.35	0.00	1.00
Removed by unknown	0.0875	0.28	0.00	1.00
Potential jurors' characteristics:				
Female juror	0.5109	0.50	0.00	1.00
Male juror	0.4573	0.50	0.00	1.00
Unknwon gender juror	0.0318	0.18	0.00	1.00
White juror	0.6069	0.49	0.00	1.00
Black juror	0.1687	0.37	0.00	1.00
Other race juror	0.0215	0.15	0.00	1.00
Unknown race juror	0.2029	0.40	0.00	1.00
Rep. juror	0.2580	0.44	0.00	1.00
Dem. juror	0.3208	0.47	0.00	1.00
Ind. juror	0.1697	0.38	0.00	1.00
Unknown pol. aff. juror	0.2515	0.43	0.00	1.00
Observations	12427			
N. Trails	566			
N. District	33			
Year	2010-2012			

A5 Judges Behaviours: Additional results

Table A11: Summary Statistics GSS

	count	mean	sd	min	max
Dependent variables:					
Courts: too harsh	26586	0.11	0.31	0.00	1.00
Courts: too lenient	26586	0.51	0.50	0.00	1.00
Courts: No trust	26586	0.01	0.12	0.00	1.00
Courts: stiffer sentences	26586	0.00	0.00	0.00	0.00
Law: Importance of respecting	26586	0.09	0.28	0.00	1.00
No trust: Fed government	26586	0.21	0.41	0.00	1.00
No trust: US supreme court	26586	0.09	0.29	0.00	1.00
No trust: Congress	26586	0.23	0.42	0.00	1.00
Independent variables:					
Unaffiliated	26586	0.42	0.49	0.00	1.00
Republicans	26586	0.24	0.43	0.00	1.00
Democrats	26586	0.33	0.47	0.00	1.00
Female	26586	0.55	0.50	0.00	1.00
White	26586	0.76	0.43	0.00	1.00
Married	26586	0.19	0.39	0.00	1.00
School (years)	26586	13.51	3.04	0.00	20.00
N. children	26586	1.86	1.67	0.00	8.00
Individuals	26586				
Year	2000-2018				

Notes: All variables are dummies, with the exception of *School (years)* and *N. Children.*

Table A12: GSS regressions results with controls.

dep. var.:	Courts are:					
	Too harsh		Right		Too lenient	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.025***	-0.030***	-0.026***	-0.026***	0.029***	0.032***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
	[-6.75]	[-7.96]	[-5.96]	[-5.82]	[4.80]	[5.32]
Republicans		-0.052***		-0.021***		0.107***
		(0.00)		(0.01)		(0.01)
		[-13.11]		[-3.79]		[14.23]
Democrats		0.015***		0.003		-0.003
		(0.00)		(0.01)		(0.01)
		[3.26]		[0.60]		[-0.38]
White		-0.074***		0.020***		0.066***
		(0.01)		(0.01)		(0.01)
		[-14.19]		[3.84]		[9.05]
School (years)		0.001		0.009***		-0.013***
		(0.00)		(0.00)		(0.00)
		[0.75]		[11.22]		[-12.99]
N. children		-0.004***		-0.001		0.006***
		(0.00)		(0.00)		(0.00)
		[-3.04]		[-0.96]		[3.52]
Married		-0.038***		0.007		0.030***
		(0.01)		(0.01)		(0.01)
		[-5.04]		[0.88]		[3.00]
Constant	0.124***	0.235***	0.206***	0.074***	0.607***	0.634***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)
	[16.89]	[16.68]	[22.59]	[4.83]	[53.87]	[32.00]
Observations	26698	26586	26698	26586	26698	26586
Years FE	No	Yes	No	Yes	No	Yes

Notes: robust standard errors in parenthesis and t-statistics in brackets. * p < 0.1, ** p < 0.05, *** p < 0.01. Dependent variables are a dummies. Excluded category: *Republicans*.