ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 035/26

Judicial Gender Match in Juvenile Courts: In-Court Sanctions and Long-Run Socioeconomic Consequences

Ozkan Eren, Randi Hjalmarsson, Orgul Ozturk

www.rfberlin.com JANUARY 2026

Judicial Gender Match in Juvenile Courts: In-Court Sanctions and Long-Run Socioeconomic Consequences

Authors

Ozkan Eren, Randi Hjalmarsson, Orgul Ozturk

Reference

JEL Codes: K40, J16, I20

Keywords: Gender Disparities, Discrimination, Incarceration, Juvenile, Court, Long-Run Consequences, Education

Recommended Citation: Ozkan Eren, Randi Hjalmarsson, Orgul Ozturk (2026): Judicial Gender Match in Juvenile Courts: In-Court Sanctions and Long-Run Socioeconomic Consequences. RFBerlin Discussion Paper No. 035/26

Access

Disclaimer

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

Judicial Gender Match in Juvenile Courts: In-Court Sanctions and Long-Run Socioeconomic Consequences*

Ozkan Eren
University of California - Riverside

Randi Hjalmarsson
University of Gothenburg

Orgul Ozturk†
University of South Carolina

January 23, 2026

Abstract

We exploit conditionally random assignment of case files to judges in a Southern U.S. state to study the effects of same-gender judge assignment on juvenile justice decisions and long-run socioeconomic outcomes. Using a generalized differences-in-differences design, we find evidence of same-gender gaps in sentencing decisions: relative to males, female judges are 18% less likely to incarcerate female than male defendants. A novel rank-order test confirms these findings and indicates that the results are not driven by a small subset of judges. Own-gender judicial assignment, especially for Black children, has lasting effects on socioeconomic outcomes beyond the courtroom: educational attainment increases while adult criminal involvement and welfare use decrease. Further analyses suggest that these long-run improvements are not just driven by differential incarceration decisions.

JEL: K40; J16; I20.
Keywords: Gender Disparities, Discrimination, Incarceration, Juvenile, Court, Long-Run Consequences, Education.

*This study is reviewed by the University of South Carolina IRB. Hjalmarsson is funded by the European Union (ERC, Police-Prisons-Firms, 101093345). Views and opinions expressed are however those of the authors only. Neither the European Union nor the granting authorities can be held responsible for them. We thank seminar participants at the NBER Children and Families Program, the Transatlantic Economics of Crime Workshop, the Catholic University, Georgia State University and the University of South Carolina. We are also grateful to Robert Pettis, Eric Baxley and Ela Ender for their excellent assistance.

†Eren: ozkane@ucr.edu, Hjalmarsson: randi.hjalmarsson@economics.gu.se, Ozturk: odozturk@moore.sc.edu

1 Introduction

A fair and impartial judiciary — one in which court outcomes are only affected by the evidence in the case — is a bedrock principle of the U.S. justice system. The public's perception of such impartiality informs the extent to which they trust the courts. Yet, U.S. opinion surveys reveal a perception of unequal treatment with respect to gender, race, and economic class in the criminal justice system, which has gotten worse over time (Smith et al., 2023).[1] An accumulating body of empirical evidence supports this public perception: research has found in a multitude of contexts that external case factors, such as decision maker characteristics and group identity, indeed impact case outcomes.[2] The resulting erosion of trust in the justice system is a first-order societal concern, as are the potentially costly consequences such high-stakes decisions can have for unfairly treated individuals.

How concerned should society be about unequal treatment in the courtroom? The answer depends on both why such courtroom disparities exist and the extent to which they translate into unequal outcomes and consequences beyond the courtroom for both the defendant and society — a topic on which there is to date little to no empirical evidence. We begin to fill this knowledge gap by studying three core research questions in the context of same-gender judge assignment: (i) Are high-stakes *juvenile* justice decisions affected by factors extraneous to the case – namely whether judges and juvenile defendants are of the same *gender*? In contrast to existing studies on adults (Lim et al., 2016; Ash et al., 2025), we find that judge gender does indeed matter for juvenile sentencing decisions. (ii) Does the (quasi-random) assignment of a same-gender judge impact long-term social and economic juvenile outcomes beyond the courtroom, including education, economic self-sufficiency, and crime? This is, to our knowledge, the first time that in-group courtroom assignments are linked to long-run out-of-court outcomes. (iii) What mechanisms underlie the differential in-court decisions and out-of-court outcomes of same-gender judge assignment?

Youths are a particularly important group to study given the critical developmental juncture at which

[1] For instance, 63% of survey respondents thought the state courts were fair and impartial in 2014 but only 47% in 2022 (State of the State Courts, 2014 and 2023).

[2] External factors include sequential decision-making (Chen et al., 2016; Bindler and Hjalmarsson, 2019), emotions from sports losses (Eren and Mocan, 2018), juror age, politics, and neighborhood (Anwar et al., 2014, 2019, 2022), judge politics (Cohen and Yang, 2019), expected punishment (Bindler and Hjalmarsson, 2018), and media exposure (Lim et al., 2015; Philippe and Ouss, 2018). The contexts across these studies vary in multiple dimensions, including decision maker (judge, jury), country (U.S., France, Sweden), and even time period (historically or today).

these decisions occur; if courtroom decisions affect outcomes such as education, then a lifetime of consequences can spillover onto society and even the next generation.[3] Our emphasis on gender is motivated both by the increasing representation of females amongst juvenile offenders and because there are enough female judges (in contrast to racial minorities) to study the potential long-run consequences of in-group assignment.[4]

To answer these questions, we match South Carolina juvenile justice court registers, in which we observe judge identifiers, to (i) administrative data on long-run socioeconomic outcomes and (ii) judge characteristics. Our core analysis sample contains more than 20,000 family court decisions – made by 84 judges, about 30% of whom are female – from 2002 to 2012. Conditional (on court-by-time fixed effects) random variation in case assignments allows us to overcome the central identification challenge – namely that different judges see *systematically* different cases.

To test for same-gender disparities in judge dispositions of juvenile defendants, we first use a generalized difference-in-differences (DiD) design. Specifically, we regress whether a defendant is incarcerated (as opposed to acquitted or convicted with a non-incarceration sentence) on the interaction of defendant and judge gender, while controlling for court-by-time-fixed effects. We find significant evidence of own-gender gaps in sentencing decisions: female judges are 2.6 percentage points (18%) less likely to incarcerate female than male defendants, relative to male judges. These estimates are robust to defendant and case controls, judge fixed effects, and a full set of interactions between judge gender and other defendant characteristics (suggesting the results are not driven by discrimination on the basis of characteristics correlated with gender). Second, we present a rank-order test that leverages the concept of stochastic dominance to establish a robust ordering across the entire distribution of outcomes. This contrasts the typical rank-order test (e.g., Anwar and Fang, 2006), which can be influenced by a small number of decision-makers given its focus on average sentencing outcomes. We find substantial evidence against the rank-order independence assumption: juvenile gender appears to influence the ranking of incarceration rates by judges' gender, and this violation is not merely driven by a small subset of judges. Findings like these are typically interpreted as evidence of taste-based discrimination (Anwar and Fang, 2006; Alesina and La Ferrara, 2014), though it does not inform

[3]See Oreopoulos and Salvanes (2011) and Hjalmarsson and Lochner (2012).

[4]Females comprised about 35 and 20% of 2019 U.S. juvenile property and violent crime arrests but only 20 and 10%, respectively, in 1980 (Puzzanchera et al., 2022).

us which judges (males, females, or both) are discriminatory.

Our analysis and findings of in-group gender disparities in juvenile incarceration decisions contribute to a growing literature that, in the spirit of Akerlof and Kranton (2000), focuses on whether group identity affects criminal justice decisions.[5] Studies of gender and race have been done for juries and courts (Boyd et al., 2010; Anwar et al., 2012; Depew et al., 2016; Lim et al., 2016; Flanagan, 2018; Anwar et al., 2019; Hoekstra and Street, 2021; Ahrsjö et al., 2024; Ash et al., 2025) as well as police search behavior and use of force (Anwar and Fang, 2006; Antonovics and Knight, 2009; Goncalves and Mello, 2021; Hoekstra and Sloan, 2022).[6] One important contribution of our paper relative to the vast majority of the existing literature is its focus on in-group disparities in *juvenile* court decisions. Despite potentially long-lasting consequences, this setting has received very limited attention to date, largely owing to data availability. A second contribution is that, to our knowledge, Lim et al. (2016) is the only study that causally investigates defendant-judge gender interactions in the U.S. context, using incarcerated adult population from the Texas state district courts. Setting aside concerns about sample selection (Arteaga, 2023), the authors find no evidence that assignment to a same-gender judge affects sentence length. Using data from Indian criminal courts, Ash et al. (2025) similarly finds null effects of gender in-group bias in judicial decisions. Finally and most importantly, our analysis does not stop in the courtroom: rather we are the first to ask what the long-run out-of-court consequences are of same-gender judge assignment and in-group courtroom disparities.

Specifically, applying the same DiD design to analyze long-run out-of-court consequences of same-gender judge assignment, we find that education and welfare outcomes improve. Relative to male judges, female defendants assigned to female judges are more likely than male defendants to take the high school exit exam (10%), enroll in twelfth grade (22%), and graduate high school in four years (16%). Though not significant, there is a sizable negative impact on violent offenses (11%) and they are also 23% less likely to rely on social welfare programs. These long-run effects are concentrated among Black children: in particular, same-gender judge assignment leads to statistically significant declines in both the likelihood of ever being arrested for a violent offense and dependence on social welfare during early adulthood.

These analyses contribute to a literature on the long-run consequences of justice system interactions

[5]Homophily has been studied in varied contexts, including education (Dee, 2005; Hoffmann and Oreopoulos, 2009; Carrell et al., 2010), sports refereeing (Price and Wolfers, 2010; Parsons et al., 2011), and medical care and evaluations (Zeltzer, 2020).

[6]See also Gazal-Ayal and Sulitzeanu-Kenan (2010), Shayo and Zussman (2011), Choi et al. (2022), and Guo et al. (2025) for international evidence on group identity and criminal justice decisions.

more generally. A number of papers investigate the causal effects of incarceration on socioeconomic outcomes by leveraging exogenous variation in incarceration from the quasi-random assignment of cases to judges with differential sentencing inclinations (Kling, 2006; Aizer and Doyle, 2015; Mueller-Smith, 2015; Bhuller et al., 2020; Eren and Mocan, 2021; Norris et al., 2021; Garin et al., 2024). Just two focus on juveniles, despite this critical period, finding incarceration lowers educational attainment and increases adult crime (Aizer and Doyle, 2015; Eren and Mocan, 2021). Recent papers also study other criminal justice interactions, including pretrial detention and misdemeanor prosecution (Dobbie et al., 2018; Agan et al., 2023; Baron et al., 2023) and, for juveniles, treatment by the adult versus the juvenile justice system (Damm et al., 2025; Loeffler and Grunwald, 2015; Loeffler and Chalfin, 2017; Mueller-Smith et al., 2023). Our study thus considers the long-run consequences of a novel dimension of justice system interactions – the characteristics of the judge a juvenile is assigned. While the incarceration decision is still a path through which long-run effects could occur, it is not the only potential mechanism.

This brings us to the mechanisms. Why does same-gender judge assignment improve out of court outcomes? The expected effect is a priori ambiguous given many potentially off-setting channels. First, the differential incarceration decisions associated with same-gender judge assignment could directly affect out-of-court outcomes. The existence and sign of such an effect, however, is not clearly predicted by existing research and/or theory. The few papers that study the effect of juvenile incarceration on post-release outcomes find mixed evidence: Aizer and Doyle (2015) and Eren and Mocan (2021) find, using a judge instrument design, worse post-release outcomes while Hjalmarsson (2009) finds, using a sentencing discontinuity design, improvements. The causal evidence on how incarceration impacts adults is equally mixed (e.g., Mueller-Smith, 2015, Garin et al., 2024).

Moreover, given that these estimates depend on the marginal offender off of which such local average treatment effects are identified, it is not obvious that existing results generalize to the margin under which such in-group biases occur. Second, the extent to which disparate treatment in the courtroom has downstream effects may depend on why the disparities exist. In the case of taste-based discrimination, then Becker's (1957) classical outcomes testing framework would predict unequal long-term outcomes. But if the underlying source of unequal in-group treatment is decision maker differences in their ability to evaluate in-group cases and determine who will be harmed by incarceration, for instance, then such a long-term effect may not occur. A third channel through which in-group decision making disparities can have long-term effects is via an offender's updating of their beliefs concerning expected future punishment. Fourth, judges make many decisions (e.g., conviction and probation terms), and it may not be one single decision – incarceration – that matters; in fact, harsh judge designs often abstract away from these multiple decisions. Finally, even in the absence of unequal treatment, assignment to in-group judges can impact outcomes via channels such as a role model effect, though the judge-defendant interaction is limited in time.

Thus, the final section of the paper speaks to these underlying mechanisms and provides suggestive evidence that the improvement in out-of-court outcomes is not just driven by differential incarceration decisions. In fact, same-gender judge assignment is associated with improved education outcomes, even if the assignment does not result in incarceration. Moreover, we also find significant gender disparities in other sentencing margins, e.g., restitution, by judge's gender.

The reminder of the paper proceeds as follows. Section 2 discusses the institutional setting, while Section 3 discusses the data used in the analysis. We describe the differences-in-differences specification and present tests of random judge assignment in Section 4. Section 5 presents the results for both in-court sentencing and out-of-court socioeconomic outcomes, as well as sensitivity and robustness checks. This section also presents our novel and complementary rank-order test. Section 6 discusses and provides empirical evidence regarding the underlying mechanisms. We conclude in Section 7.

2 The South Carolina Juvenile Justice System

South Carolina uses a system of legislative, merit-based elections for almost all judicial positions. To vie for these positions, candidates must submit comprehensive applications to the Judicial Merit Selection Commission (JMSC), which is comprised of six members from the General Assembly (including three Senators and three Representatives) and four non-legislative members. Legislators then select judges from a pool of up to three candidates deemed qualified for judicial office by the JMSC. Judicial elections take place during a joint session of the General Assembly, with each member of the Senate and House casting one vote and the outcome being determined by a majority rule. Judges are elected to six-year terms and are subject to re-election.

The judicial system is structured into 16 judicial circuits, each comprising a combination of the state's

46 counties. The state's court of general jurisdiction is known as the circuit court. Additionally, family courts hold exclusive jurisdiction over cases involving youth under the age of 17 and matters related to family relationships.[7] Unlike a fixed assignment system, judges in South Carolina do not sit exclusively in one court or county. Instead, they travel throughout different counties within their resident circuits and to circuits outside their residential jurisdiction. The rotation system is neither uniform nor systematic and largely depends on expected caseload requirements, meaning that judge allocations to courthouses are made before the specific cases on the docket are known. Judges are assigned to courts on an approximately weekly term, although it is not uncommon for fewer days of court assignment to occur. Scheduling is under the direction and supervision of the chief justice. In our analysis sample, family court judges, on average, traveled to six counties from three judicial circuits, while counties had seven different family court judges handling juvenile defendant cases during any given year.

Having received a formal complaint, the Office of the Solicitor must decide whether to petition the juvenile case to the court. Prosecutors may choose not to do so because of a lack of sufficient evidence. If certain criteria are met, the prosecutor may also allow the case to be diverted from the juvenile justice system.[8] This means that instead of being prosecuted in the family court, the juvenile is allowed to participate in a diversion program, such as arbitration, a behavior contract, or drug court (Children's Law Center, 2019). Upon successful completion of the diversion program, the charges against the child will be dismissed. Finally, if the prosecutor chooses to prosecute, the case moves to adjudication, which is comparable to trial in adult court. A family court judge reaches a verdict after hearing the evidence presented by the prosecution and defense attorneys.

The judge may determine that the state failed to prove its case beyond a reasonable doubt and find the child not guilty, or the judge may find the child guilty and adjudicate the child delinquent, which is analogous to a conviction in adult criminal court. If the decision is to acquit, the child is released from the court's jurisdiction. Upon a guilty verdict, the judge has to then make a disposition decision.[9] These judgments

[7] South Carolina law defines "child" as a person less than 18 years old. However, in the context of delinquency and criminal acts, "child" refers to a person under age 17. Additionally, this definition excludes 16 years-old individuals charged with several felony offense categories (e.g., second-degree burglary; domestic violence of aggravated nature). The state raised the age of majority from 17 to 18 in 2019.

[8] These criteria include, but are not limited to, being a first-time offender, a nonviolent offender, or drug/alcohol dependent (Children's Law Center 2019).

[9] After adjudicating a child delinquent, the family court judge almost always moves directly to the sentencing phase without ordering a predisposition evaluation report, which includes information about the child and their surroundings, background, and

may include restitution, community service, behavioral/drug assessments and treatment, probation and incarceration (i.e., commitment to the South Carolina Department of Juvenile Justice, SCDJJ, for placement in one of its institutions). The terms of probation are also specified by the judge, and can include, for instance, regular school attendance, random drug testing, and community service. The SCDJJ operates its own school district and committed juveniles must continue their education in the agency's middle and high schools until they reach the minimum drop out age.

In general, the judge should impose the least restrictive sentencing option consistent with the circumstances of the case, the health and safety of the child, and the best interest of the community (Children's Law Center, 2019). To date, South Carolina remains a non-guideline state, meaning that judges exercise large amounts of discretion in their disposition decisions, constrained only by mandatory minimums and statutory maximums (Hester, 2017).

We conclude by presenting statistics that highlight the prevalence of juvenile arrests and prosecutions in the state of South Carolina, demonstrating that South Carolina is fairly representative of the U.S. overall. Specifically, Panel A of Figure A.1 in the Online Appendix displays trends in juvenile arrest rates (combining violent and property crimes) from 2002 to 2012, while Panel B shows the average proportion of juvenile referrals resulting in prosecution over the same period.

3 Data

3.1 Data Sources

This paper matches data from multiple criminal justice and socioeconomic administrative registers. From the South Carolina Department of Juvenile Justice (SCDJJ), we obtained information on all petitioned case files from 2002 to 2012 at the charge, case, and defendant levels. The charge-level data include details about the type and severity (based on state-specific grades, which range from 0 to 9) of offenses and sentencing outcomes for each charge. The case-level data provide information on offense, referral and petition dates, the court where the case was tried, and the date of disposition. Finally, the defendant-level data include

circumstances (Hester 2017). This is because judges already have access to a similar report from the initial screening process or "DJJ intake." If a dispositional hearing is held at a later date, South Carolina Code of Laws require judge who presided over the adjudicatory hearing to also preside over the disposition hearing (South Carolina Family Court Law-Rule 37).

information on the juvenile's gender, race, and age at disposition.

The second data source is the log of judicial rotation assignments from the South Carolina Court Administration, which informs us of the presiding judge for each court and day of the calendar year.[10] We use these logs to construct a family court judge-by-court-by-date of disposition panel. We also supplement these data with manually collected information on judge attributes from a variety of sources, including online searches and short bios of family court judges' available on the South Carolina Judicial Branch website. To link detailed case files to judges, we match administrative data from the SCDJJ to our log panel using court and disposition date. Juvenile defendants can be successfully matched to their disposition judges when cases are handled by a single judge on any day in a given court. In this way, we can match around 38% of all juvenile case files to their exact disposition judge.[11] As discussed below, the quasi-random assignment of case files to judges minimizes concerns about sample selection bias.

To study the effects of courtroom gender interactions on outcomes beyond the courtroom, we use a common and unique individual identification number to link the matched juvenile defendant-judge data with administrative records from: (i) the South Carolina Department of Education (SCDOE), (ii) the South Carolina Department of Social Services (SCDSS), and (iii) the South Carolina State Law Enforcement Division (SLED). Any juvenile defendant who does not have a match is coded as zeros for these outcome variables.

The SCDOE data include information on grade enrollment and on-time high school graduation from 2000 to 2017 as well as high school exit examinations. Public school students are required to pass the high school exit exam to qualify for a diploma. The High School Assessment Program (HSAP) was first administered (in the spring of tenth grade) during the 2005-2006 academic year as the state's designated exit exam. Exit exams were discontinued by the state in 2015.[12] From the SCDSS, we obtain information on enrollment in social programs through 2019. We use information on welfare receipt – both food stamps (renamed Supplemental Nutrition Assistance Program or SNAP in 2008) and Temporary Assistance for

[10]We scraped judicial rotations from the South Carolina Judicial Branch website (https://www.sccourts.org/calendar/).

[11]We also exclude matches in which the disposition judge handled less than 10 case files over the sample period. There are only three judges meeting this criteria.

[12]South Carolina repealed the Common Core Standards in 2015. The decision to abandon exit exams was part of the state's efforts to develop its own educational standards and put in place assessments that will better identify areas of academic strength and weakness. As a result, the HSAP was replaced by end-of-course exams. Scores from these exams count towards students' final grade in gateway subjects such as Algebra, Biology and English.

Needy Families (TANF) – to measure the long-run economic outcomes of our core sample. Finally, from SLED, we obtain adult arrest records for 2000 to 2017. For each offender file in the adult criminal system, we have offense date and type of offense committed.

3.2 Outcome Variables

Our primary sentencing outcome of interest is whether a juvenile defendant was sentenced to incarceration, defined at the case level based on the most serious conviction. We construct analogous indicator variables for other judgments, such as acquittal. Measures of educational outcomes include: (i) HSAP test-taking, defined as a binary indicator equal to one if the child ever took the high school exit exams, (ii) twelfth grade enrollment status, and (iii) on-time high school graduation.[13] We use the SCDSS data to construct two measures of economic self-sufficiency: whether the child ever received food stamps and whether the child ever received TANF as an adult. Finally, we measure criminal activity by defining binary indicators for ever being arrested as an adult, as well as for different types of offenses.

3.3 Sample and Descriptive Statistics

Our core analysis sample consists of 23,569 juvenile case files (from 15,591 unique children) that were prosecuted and successfully matched (exactly) to the disposition judge. These files were handled by 84 judges in 46 family courthouses from 2002 to 2012. Since the juveniles were primarily born between 1988 and 1998, we can observe educational and crime outcomes up until age 29 and welfare participation up to age 31. For example, we can observe the receipt of TANF until age 31 for children born in 1988, age 30 for children born in 1989, and so on. Because we observe all public-school enrollments in South Carolina, concerns about attrition in the education data only arise if children leave the state, attend a private school, or are home-schooled. Endogenous mobility is arguably even less problematic for crime and welfare outcomes since attrition would only be affected by out-of-state mobility. According to data from the American Community Survey, around 9% of adults born in South Carolina between 1988 and 1998 had left the state by 2018; the migration outflow is even lower (7%) when restricting these birth cohorts to have high school

[13] Approximately 12% of juvenile defendants took the HSAP before the disposition date. We exclude these observations from the analysis sample when the outcome of interest is HSAP test-taking. However, keeping them in the sample and imputing the dependent variable either with zero or one for these children yields point estimates that are almost identical to those presented throughout the paper.

or lower levels of education.

Columns 1 and 2 of Table 1 report summary statistics for our matched sample.[14] Panel A presents juvenile and case characteristics. For dispositions, we see that 12% of cases resulted in acquittal, 74% in conviction but not incarceration, and 14% in incarceration.[15] The average age at disposition is around 15, while 28% of cases involve female defendants and 56% and 41% involve Black and White defendants, respectively. Property, violent, and school-related (e.g., disturbing schools; threatening a teacher/principal) offenses comprise 21%, 16%, and 22% of cases respectively; the remainder include a heterogeneous set of other offenses. Finally, 32% of cases are associated with multiple offenses.

Panel A also presents summary statistics for the main out-of-court defendant outcomes. Only 16.6% of defendants enrolled in twelfth grade while 14% graduated high school on-time. Around 80 (14)% used food stamps (TANF) as an adult and 64% were arrested in early adulthood.[16]

Panel B shows that, in the matched sample, female judges handled around 26% of cases, while 95% were decided by White judges and 86% by judges with a South Carolina college degree. At disposition, judges are on average 53 years old.

We next turn to how representative the exactly matched sample is of the universe of South Carolina juvenile court cases. Online Appendix Figure A.2 displays the number and share of case files that were successfully matched in each of South Carolina's 16 judicial districts. This share is less than 20% in three districts and between 40 and 100% in eight districts; there is a significant number of cases from all districts. Panel C of Table 1 looks closer at the underlying county characteristics for the matched and unmatched samples. Cases from less populated counties are over-represented in the matched sample. This is not surprising given that exact matching could only be done on courthouse-days with just one family court judge. Consistent with this, Panel C shows that defendants in the matched sample are adjudicated in counties with on average about seven family court judges while this statistic is nearly double in the unmatched sample.

But, does this non-representativeness in terms of county characteristics (e.g., ruralness and population

[14] Appendix Table A.1 shows statistics by juvenile gender.

[15] Mandatory school attendance comprises a large fraction (around 67%) of the judgments within the subcategory of convictions without probation. The remaining judgments are a very heterogeneous group, ranging from monetary restitution to community service to behavioral/drug assessments. Individuals on probation can also get additional dispositions that characterize the terms of probation, e.g., school attendance.

[16] TANF is restricted to low-income households with dependent children while able-bodied adults without dependents are also eligible for food stamps.

size) impact the representativeness of the sample in terms of juvenile defendant/case characteristics and both sentencing and long-run socioeconomic outcomes? Columns 1 and 3 of Table 1 demonstrate that the matched and unmatched samples are surprisingly similar. To shed further light on the extent to which there are sample selection issues, Online Appendix Table A.2 regresses judge missingness (i.e., whether a case is exactly matched to a judge) on defendant and case characteristics (Column 1) and case outcomes (Column 2). These observable characteristics have little predictive power for either the full sample or the subsamples of male and female defendants (Columns 3 and 4); the p-values for joint significance range from 0.34 to 0.57. These results support the conclusion that failing to match cases to disposition judges is as good as random.

We also show that the baseline results are robust to using an expanded sample of 36,729 cases, which can be matched to courthouse-days with multiple judges who are all of the same gender. Though we cannot match to disposition judge identifiers in such cases, we can assign judge gender. We further use this expanded sample in heterogeneity analyses that depend on sample size for precision.

4 Empirical Methodology

4.1 Difference-in-Differences Approach

To evaluate juvenile defendant-judge gender interaction effects on judicial dispositions, we use a generalized difference-in-differences (DiD) approach and estimate the following equation:

$$JD_{icjt} = \beta_0 + \beta_1 Female\_Def * Female\_Judge_{icjt} + X'_{icjt}\beta_2 + \lambda_j + \gamma_{ct} + \epsilon_{icjt} \tag{1}$$

where JD_{icjt} is the juvenile disposition outcome for defendant i in case c determined by judge j at time t. In most specifications, this equals one if the defendant is adjudicated guilty and incarcerated and zero if the child was acquitted or adjudicated guilty with a non-incarceration sentence (e.g., probation). X'_{icjt} is a vector of observed case and juvenile defendant characteristics (i.e., indicators for gender and race, offense severity, broad offense type and year of birth fixed effects, number of charges, and number of times a defendant is observed in the data).[17] λ_j and γ_{ct} stand for judge and court-by-time fixed effects, respectively, and ϵ_{ijct} is

[17] Robustness checks replace these broad offense type dummies with indicators for each of more than 300 specific offenses.

the error term.

Because case types may vary by court, year-month, and day of the week, we control for court-by-year-month and court-by-day-of-week fixed effects. Court-by-time fixed effects (γ_{ct}) safeguard against potential case selection and ensure that family court judges preside over similar case files, a condition that must hold for causal identification. Below, we empirically assess the quasi-random assignment of case files to judges and provide convincing evidence on the validity of this assumption.

The coefficient β_1 reflects the average change in judicial disposition rates for female defendants, relative to males, from the quasi-random assignment of cases to female versus male judges. Put differently, we estimate whether gender disparities in sentencing outcomes are different between female and male judges. This is necessarily a *relative* measure and we cannot sort out whether a non-zero effect means female judges impose more lenient (harsher) dispositions on female (male) defendants, or whether male judges impose more lenient (harsher) dispositions on males (females). Answering this question in this framework is difficult because it requires a (non-discriminatory) benchmark, which is not easy due to the lack of sentencing guidelines in South Carolina.

Controlling for judge fixed effects addresses the possibility of β_1 being biased due to the non-random sorting of judges across jurisdictions. For example, female judges, with tendencies toward harsher punishments or differential gender attitudes, could systematically choose to serve in judicial circuits where gender (of either the defendant or victim) is a particularly salient case characteristic, such as districts with more sex crimes. In this case, the DiD estimates without judge fixed effects would capture the combined effects of both within-judge behavior and differential sorting. Robust standard errors are two-way clustered at the judge and court level.

Without any further assumptions, the relative gender disparity in judicial dispositions identified in equation (1) includes both direct gender-based discrimination (whether taste-based or statistical) and indirect discrimination via non-gender characteristics correlated with gender (Argyle et al., 2025). If we further assume that, absent direct discrimination, female and male judges on average make decisions that would result in the same disparities for female and male defendants, the gender interaction effect only reflects taste-based and statistical discrimination. This could be violated if, for example, female judges are more likely than males to incarcerate defendants charged with violent crimes and if there is a higher proportion of male defendants charged with these offenses. To isolate direct gender-based discrimination, we present the results controlling for interactions between judge gender and defendant/case characteristics (race, offense severity, offense type, birth year, and number of charges) .

In addition to examining judicial decisions, we use this research design in Section 5.3 to analyze the courtroom gender interaction effects on long-run educational, economic self-sufficiency, and criminal justice outcomes.[18] In other words, we study the reduced form impact of same-gender judge assignment on long-run outcomes. The fact that same-gender judge assignment can affect outcomes through multiple channels — including probation terms, time in confinement, and, in theory, role model influences — makes the reduced form approach appropriate.

4.2 Randomization Tests

In Table 2, we empirically assesses the validity of the quasi-random assignment of case files to judges. The first column uses a linear probability model to test whether case and defendant characteristics are predictive of incarceration, after controlling for court-by-time fixed effects. Each cell represents a separate regression. We find that the decision to incarcerate a juvenile defendant is highly correlated with all observable covariates. The coefficient estimates are also jointly significant from a single regression using all juvenile/case characteristics (p-value=0.00). As shown in Column 2, observable characteristics are also highly correlated with the decision to acquit a juvenile defendant. Columns 3-5 of Table 2 present the point estimates using judge characteristics (gender, race/ethnicity, and age) as dependent variables. None of the nine observable characteristics are significantly related (at even the 10% level) to judge gender (Column 3); the p-value for joint significance is 0.54. These results do not change in a meaningful way when judge gender is replaced with a binary indicator for judge's minority status and age (Columns 4 and 5).

We further complement these randomization tests by analyzing whether judge leniency is correlated with observable characteristics (Column 6). We measure leniency by the leave-one-out average non-incarceration rate among all other juvenile defendants assigned to a judge. Specifically, we construct the leave-one-out measure using the residuals from a regression of the non-incarceration indicator on court-by-time fixed

[18] Since some youth have multiple encounters with the justice system and are observed multiple times, our long-run specifications are estimated using probability weights, where the weight is the inverse of the number of times a juvenile is observed. These models do not control for the number of times a juvenile is observed in the data.

effects.[19] Reassuringly, the coefficient estimates are small in magnitude and not statistically different from zero (p-value=0.73). The p-values for joint significance from analogous regressions for female and male juvenile defendant samples are 0.29 and 0.90, respectively. Online Figure A.3 also displays the distribution of the residualized judge leniency, which ranges from -0.138 to 0.054 with a standard deviation of 0.022.

5 Results

5.1 Juvenile Defendant-Judge Interactions and Sentencing Outcomes

Table 3 presents the baseline sentencing results. The dependent variable in Columns 1-4 is incarceration. Column 1 presents the juvenile defendant-judge female interaction effect on incarceration rates when controlling only for court-by-time fixed effects. Relative to male judges, female judges are 2.4 percentage points less likely to incarcerate female than male defendants. This is roughly equivalent to the baseline sentencing gap between female and male judges and it is 17% of the mean incarceration rate. The estimated coefficient on the female interaction term is robust to the inclusion of defendant and case characteristics in Column 2, consistent with quasi-random assignment of judges to case files. Column 3 includes judge fixed effects to account for potential non-random sorting of female and male judges across counties. The DiD estimate remains stable and significant. In Column 4, we control for interactions between judge gender and juvenile/case characteristics (race, severity of the offense, offense and birth year fixed effects and number of charges at disposition). The point estimate is nearly identical, suggesting limited scope for indirect discrimination on non-gender characteristics and indicating that the estimates likely reflect taste-based and statistical discrimination.

Is this strong defendant-judge gender interaction effect on the likelihood of incarceration (versus probation or acquittal) driven by the sentencing and/or conviction decisions? Column 5 of Table 3 estimates equation (1) when the dependent variable is acquittal, finding a small and statistically insignificant gender interaction effect, i.e., pointing to the disposition as the relevant margin. Finally, as an additional validation check of our design, we show in Column 6 that the defendant-judge gender interaction is not economically or significantly related to the predicted value of juvenile incarceration – obtained from regressing an indicator

[19] To allow for meaningful variation in judge leniency in a given cell, court-by-time fixed effects for leniency specifications include court-by-year-by-quarter fixed effects.

for incarceration on (pre-determined) observable juvenile characteristics.[20]

The defendant-judge gender interaction effect of 2.6 percentage points for incarceration in Column 3 is not small: it roughly equals 45% of the unconditional juvenile female-male gap in incarceration rates (Appendix Table A.1) and 18% of the baseline mean incarceration rate. To gain further perspective, we compare the latter statistic to the effect of being assigned 'harsh' judges on juvenile sentencing outcomes (Aizer and Doyle, 2015). The size of our interaction effect is comparable to being assigned to a judge with a 16 percentage points higher incarceration rate.

There are two contemporaneous papers that investigate the effect of gender interactions on judicial decisions with similar quasi-experimental research designs. Using felony cases from Texas state district courts, Lim et al. (2016) document no evidence that same-gender judge assignment affect sentence length in the sample of incarcerated defendants. Using data from Indian criminal courts, Ash et al. (2025) show that sharing gender with the defendant makes a judge neither more nor less likely to deliver an acquittal. Differences in institutional context (developed versus developing countries), margins of incarceration (extensive versus intensive), and differences in defendant populations (adult versus juveniles) may all contribute to disparate findings across settings.[21]

5.2 Individual Judge Effects and Rank Order Tests

We complement our baseline DiD estimates with a rank-order test that leverages the concept of stochastic dominance to establish a robust ordering across the entire distribution of outcomes. This section shows that the defendant's gender appears to influence the ranking of incarceration rates by judges' gender, and this violation in rank independence is not merely driven by a small subset of judges.

We first use the residuals from a regression of the incarceration indicator on court-by-year-by-quarter fixed effects, along with case, juvenile, and observable judge characteristics, to estimate a judge random effect model. We obtain shrunken estimates for each individual judge by defendant gender.[22] The conditional randomization of cases to judges allows us to causally identify individual judge effects. Panel A of Figure

[20]The estimated gender interaction effect is -0.003 (s.e.=0.002) when using the predicted value of acquittal as the dependent variable.

[21]Differences in effects of judges on extensive versus intensive sentencing margins are not only relevant in the judge-defendant gender context, but also the harsh judge literature more generally, which tends to find that there is variation in judge decisions at some margins, including any incarceration, but not always other margins like sentence length. See Loeffler and Nagin (2022).

[22]We use the Stata package *mixed* to obtain shrunken estimates.

1 displays density plots of these judge effects comparing male and female judges for female defendants. There are fewer female judges in the right tail of the distribution. Panel B presents the corresponding judge effect estimates for male defendants – a noticeable rightward shift is evident in the distribution of female judges compared to male judges.[23]

We next utilize stochastic dominance tests, which offer social welfare comparisons over a large class of preferences, to establish robust ordering over the entire distribution. Such comparisons may provide an important advantage over the usual rank-order test (Anwar and Fang, 2006), which centers on the average outcomes and thus can be impacted by only a few decision-makers. Suppose S_f^F and S_f^M denote the random effects for female and male judges, respectively; these map to the individual judge effect estimates for female defendants (f). Let $F_f^F(s) = \Pr(S_f^F < s)$ represent the cumulative density function (CDF) of S_f^F; $F_f^M(s)$ is similarly defined for S_f^M. Under this notation, S_f^M first-order stochastically dominates (FSD) S_f^F if

$$F_f^M(s) \leq F_f^F(s) \quad \forall s \in S_f \text{ with strict inequality for some } s \tag{2}$$

where S_f denotes the union support of S_f^F and S_f^M. One can similarly formulate the FSD condition for male and female judges using individual judge effects for male defendants (m).

To test for FSD, we utilize the Kolmogorov-Smirnov test criteria and the resampling procedure in Linton et al. (2005). Under this scheme, one can infer FSD to a desirable degree of confidence if the p-values of the test statistics are large, say 0.90 or higher. See Appendix B for details. Assessing the SD tests, we find that the p-value associated with S_f^M FSD S_f^F is 0.81 and that for S_m^F FSD S_m^M is 0.91. Online Appendix A.5 also displays the empirical cumulative distribution functions (CDF). We do not observe any meaningful crossings of the CDFs in either panel.[24]

These results are consistent with same-gender (in-group) bias on the part of judges. In our context, judicial bias implies that the defendant's gender should affect the ranking of incarceration rates over judges' gender. This is exactly what we find. Failure to preserve FSD ranking of incarceration rates by judge's

[23]These conclusions are robust to a standard Empirical Bayes shrinkage correction (Online Figure A.4). Specifically, Empirical Bayes estimates of judge-specific incarceration rates indicate that among female judges, the case-weighted average incarceration rate is 1.8 percentage points for female defendants and 6.5 percentage points for male defendants. Among male judges, the corresponding rates are 2.3 and 5.4 percentage points, respectively. As discussed further below, the gender of the juvenile defendant affects the average ranking of incarceration rates over judge gender, consistent with a violation of the rank order independence.

[24]Point estimates from regressing incarceration on a female judge dummy, with all controls, equals -0.014 (s.e.=0.028) for female defendants and 0.020 (s.e.=0.015) for males, highlighting again the violation of the rank independence condition.

gender (i.e, S_f^M FSD S_f^F, while S_m^F FSD S_m^M) may lend support for taste-based discrimination (Anwar and Fang, 2006; Alesina and La Ferrara, 2014).[25]

5.3 Juvenile Defendant-Judge Gender Interactions and Out-of-Court Longer-Run Outcomes

This section considers the potential out-of-court consequences of these in court juvenile defendant-judge gender interactions. The expected sign of these effects is a priori ambiguous given the multiple potential underlying channels, including: (i) direct short-run incapacitation effects of differential sanctions, (ii) long-run (rehabilitative or criminalizing) sanction effects, (iii) belief updating about expected punishment in the future, and (iv) role model effects. Panel A of Table 4 presents the results of estimating our baseline specification when using three categories – education, crime, and welfare – of out-of-court outcomes in the short and long-run as the dependent variable.

We begin with educational outcomes. The results for these outcomes—took the high school exit exam (spring of tenth grade and beyond), enrollment in twelfth grade, and graduation from high school within four years—are presented in Columns 1–3 of Table 4, respectively. The coefficients on the juvenile defendant-judge gender interaction are positive and significant in all three specifications. Relative to male judges, female defendants assigned to a female judge are: 4.2 percentage points (10%) more likely to take the high school exit exam than male defendants, 3.7 percentage points (22%) more likely to enroll in twelfth grade, and 2.2 percentage points (16%) to graduate high school on time. Column 4 also shows that the schooling index, constructed as the average of z-scores of all non-missing educational outcomes, increases by 9.8% of a standard deviation as a result of same-gender judge assignment.

Column 5 reports the gender interaction effect when the dependent variable is equal to one if the individual was arrested as an adult. The coefficient estimate is both insignificant and small in size, although a two-sided test cannot rule out modest reductions of about 6.5 percentage points (10%). This null effect on crime overall masks some underlying heterogeneity: There is no effect on drug or other offenses, but a sizable negative impact on violent offenses (-0.010, s.e.=0.011, 11%) and positive, statistically significant

[25] Such a rank-order violation can also arise in theory if juvenile defendants respond differently to opposite-gender judges (e.g. they cooperate less in the courtroom), even if judges treat all juveniles equally. We believe the former interpretation – in-group bias – is the more likely. This is because judges determine outcomes after reviewing predisposition and DJJ intake reports about the child's background, environment, and circumstances. Consequently, any observed differences in sentencing ultimately reflect judicial decision-making.

impact on property offenses (0.025, s.e.=0.011, 15%). The property offense effect, which emerges at age 17 and gets smaller relative to the mean as the sample ages, is consistent with an incapacitation effect (or the lack thereof) of the decreased likelihood of incarceration. Consistent with this explanation, the gender interaction effect is 0.008 (s.e.=0.008) when the outcome is defined as ever being arrested for a property offense at age 18 or older.

Columns 6 and 7 of Table 4 present results for our two measures of welfare use – whether the child ever received TANF or food stamps/SNAP as an adult. There is a null (albeit imprecise) effect associated with the use of food stamps but a large reduction in TANF receipt; though not quite achieving significance at traditional levels, we do see that the likelihood of TANF receipt as an adult decreases by 3.2 percentage points or 23%.

Finally, we created a long-run summary measure that allows us to test for co-movements of related outcomes and address multiple hypothesis testing concerns (Kling et al., 2007). The long-run index combines the schooling index from Column 4 with z-scores of indicators for being arrested for a violent offense and any welfare participation, with the latter two reverse-coded so that positive values indicate improvement. The index averages the z-scores, weighting each of the three subcomponents equally. Column 8 of Table 4 displays the results: the estimated gender interaction coefficient indicates that for female defendants, being assigned to a female judge is associated with 7.2% of a standard deviation improvement in long-run outcomes, which is significant at the 10% level. Given their comprehensive nature, we use the schooling and long-run index measures as our outcomes of interest in the subsequent sections.

5.4 Robustness Checks and Heterogeneity

We next implement a series of sensitivity checks to examine the robustness of our main DiD results. Table 5 shows that the results do not substantively change when: (i) using alternative court-by-time fixed effects, (ii) controlling for more than 350 detailed offense types, (iii) excluding non-incarceration convictions with mandatory school attendance (to address concerns about a potential mechanical relationship between mandatory school attendance orders and juvenile educational outcomes), (iv) not weighting the estimates, and (v) expanding the sample of cases by more than 55% by including days when all courthouse judges are the same gender (though this does not allow for the inclusion of judge fixed effects). We will use this expanded sample for additional heterogeneity analyses described below.

Online Appendix Table A.3 shows that judge shopping is not a salient concern, as there is little heterogeneity across different time windows even though shopping is more viable for longer gaps between referral and disposition dates. The statistical inference does not meaningfully change when alternatively clustering at the judge, child, and court-by-year-month levels (Online Appendix Table A.4).

We also conducted two placebo tests. First, we used the referral date, rather than the date of disposition, to link case files to judges. If our results are driven by unobservables that are correlated with judge gender that affect outcomes, then we would expect to see a significant spurious correlation between the female interaction term and sentencing and longer-run outcomes. For example, skilled defense attorneys might be able to delay a case to be heard by a judge of the same-gender, which could erroneously lead to attributing the observed effects to the judge's gender. These results are presented in Table 6. Reassuringly, we do not find any impact of the juvenile defendant-judge gender interaction term, which is based on the gender of the presiding judge on the day of referral, on either the likelihood of incarceration or index measures.[26] Second, we re-estimate our baseline DiD specification after randomly assigning gender to each judge. Figure 2 displays the distribution of the interaction coefficient for 1,000 simulations. Estimates are smaller than the actual one for incarceration in just 18 permutations, while 8 and 41 yield estimates larger than the true schooling and long-run measures, respectively.

Finally, we explored heterogeneity in the effects of courtroom gender interactions by juvenile race. As noted above, to improve efficiency, we use the extended sample that includes court-day observations in which all judges at a given courthouse were of the same gender. We focus on Black and White juvenile defendants since there are few defendants of other races in the juvenile justice system in South Carolina. Table 7 presents the estimates. Recall that these specifications do not control for judge fixed effects; however, given no evidence of non-random sorting of judges, omitting these fixed effects is unlikely to pose a threat to causal identification. As shown in Column 1, relative to male judges, female judges are less likely to incarcerate both Black and White female defendants than male defendants. Yet we find a statistically significant impact on the long-run outcome index only for Black children (Column 2). The coefficient estimates for individual components (Columns 3-6) indicate sizable reductions in both the

[26] A placebo test using the offense date to link case files to judges produces very similar results.

likelihood of ever being arrested for a violent offense (24%) and receiving TANF (37%).[27] Online Appendix A.5 reports corresponding estimates using the baseline sample, conditioning on judge fixed effects as well as interactions between judge race and defendant gender and race. These additional controls account for potential systematic differences in how judges of different races treat juvenile defendants across gender and race groups. Reassuringly, the results are similar to those reported in Table 7.

We also examine heterogeneity along the dimensions of age at first-time disposition (around age 15) and offense type, with the latter analyzed by dropping each of the four broad juvenile offense categories in turn (Online Appendix Figure A.6). For the most part, despite some loss in precision, we do not observe significant heterogeneity along these dimensions. Same-gender judge assignment has similar effects on incarceration decisions across sub-samples as well as out-of-court outcomes measured by the schooling and long-run indices.

6 Mechanisms

Same-gender judge assignment (i) leads to differential incarceration outcomes and (ii) improves long-run outcomes. This section examines whether the latter explicitly operates through reduced incarceration. Such a channel would be consistent with research that finds using a harsh judge design that juvenile incarceration decreases education (Aizer and Doyle, 2015). Judges, however, make multiple decisions and can even impact youths via non-decision channels.

To assess whether incarceration is the sole underlying mechanism, we first follow Black et al. (2022) and re-estimate the long-run specifications for those juveniles *never* sentenced to incarceration. Intuitively, same-gender judge assignment should not significantly affect outcomes among individuals who are not treated. If it does, this suggests that either the exclusion restriction fails or there is selection, or both. Given the compelling evidence of quasi-random assignment of case files to judges (Table 2 and Column 6 of Table 3), one may view the results of the Black et al. (2022) test as evidence on the validity of the exclusion restriction. This intuitive test should, however, be interpreted with caution given concerns about splitting the sample based on an endogenous variable.

[27] We continue to observe no effect of same-gender judge assignment on drug or other offenses, but a statistically significant impact on property offenses (0.015, s.e.=0.007) in the extended sample. There is also no evidence of heterogeneous effects by race for these offense types.

With this proviso in mind, Online Appendix Table A.6 shows that same-gender judge assignment appears to improve educational outcomes. Though statistically insignificant, the point estimate on the long-run index is also sizable. Taken together, these patterns suggest that same-gender judge assignment may influence certain outcomes through channels beyond incarceration. What those channels are, however, are less clear. For instance, in Online Appendix Table A.7, we find that, relative to male judges, female defendants assigned to a female judge are 0.4 percentage points less likely to receive restitution, a difference that is statistically significant at the 5 % level, while we do not find differential effects on acquittal decisions, mandatory school attendance, and other sentencing margins.[28] There are of course other judge decisions (e.g., sentence length) that we cannot observe and channels (e.g., role model) that we cannot directly test.

7 Conclusion

Exploiting the quasi-random allocation of juvenile case files to South Carolina judges, we examine the effects of same-gender judge assignment on both high-stakes juvenile justice decisions and short- and long-run socioeconomic outcomes. Gender congruence in judicial assignment reduces incarceration by about 800 cases or more than 70 per year of our sample.[29] We also document long-run effects: for instance, gender congruence increased twelfth grade enrollment by 700 juveniles. These long-run effects are concentrated among Black children.

Addressing discrimination requires policies and reforms aimed at ensuring a fair and impartial judiciary. Establishing presumptive sentencing guidelines may help limit judicial discretion and thereby reduce unequal in-group disparities on criminal justice decisions and associated consequences beyond the courtroom. Recent research has also shown that making individuals aware of their own biases–for example, using Implicit Association Test (IAT) scores—can influence discriminatory behavior (Alesina et al., 2024). Encouraging judges and other judicial actors to take the IAT to raise awareness of their implicit associations regarding gender and race may be an effective tool for reducing discrimination. Another potential policy dimension to consider is the composition of the judiciary. There are fewer female and minority judges than

[28] There is also no evidence of heterogeneity in the estimated gender interaction effects by juvenile race for other sentencing margins.

[29] We multiply the number of case files (62,502) by the coefficient (-0.013, s.e. = 0.006) on whether the judge and juvenile are same-gender when regressing the incarceration decision on it and the baseline controls.

male and white judges. This raises the question of whether females for instance who do make it into the judiciary are more "selected" (including self-selected) than male judges. Such selection could decrease with increased representation.

References

Agan, A., J. L. Doleac, and A. Harvey (2023). Misdemeanor prosecution. *Quarterly Journal of Economics 138*(3), 1453–1505.

Ahrsjö, U., S. Niknami, and M. Palme (2024, November). Identity in Court Decision-Making. *American Economic Journal: Economic Policy 16*(4), 142–164.

Aizer, A. and J. J. Doyle (2015, May). Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges *. *The Quarterly Journal of Economics 130*(2), 759–803.

Akerlof, G. A. and R. E. Kranton (2000, August). Economics and Identity. *Quarterly Journal of Economics 115*(3), 715–753.

Alesina, A., M. Carlana, E. LaFerrara, and P. Pinotti (2024). Revealing stereotypes: Evidence from immigrants in schools. *American Economic Review 114*, 1916–1948.

Alesina, A. and E. La Ferrara (2014, November). A Test of Racial Bias in Capital Sentencing. *American Economic Review 104*(11), 3397–3433.

Antonovics, K. and B. G. Knight (2009, February). A New Look at Racial Profiling: Evidence from the Boston Police Department. *Review of Economics and Statistics 91*(1), 163–177.

Anwar, S., P. Bayer, and R. Hjalmarsson (2012, May). The Impact of Jury Race in Criminal Trials. *The Quarterly Journal of Economics 127*(2), 1017–1055.

Anwar, S., P. Bayer, and R. Hjalmarsson (2014, November). The Role of Age in Jury Selection and Trial Outcomes. *The Journal of Law and Economics 57*(4), 1001–1030.

Anwar, S., P. Bayer, and R. Hjalmarsson (2019, February). A Jury of Her Peers: The Impact of the First Female Jurors on Criminal Convictions. *The Economic Journal 129*(618), 603–650.

Anwar, S., P. Bayer, and R. Hjalmarsson (2022, June). Unequal Jury Representation and Its Consequences. *American Economic Review: Insights 4*, 159–174.

Anwar, S. and H. Fang (2006, February). An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence. *American Economic Review 96*(1), 127–151.

Argyle, B., S. Indarte, B. Iverson, and C. Palmer (2025). Racial disparities and bias in consumer bankruptcy. Working Paper 33575, National Bureau of Economic Research.

Arteaga, C. (2023). Parental incarceration and children's educational attainment. *Review of Economics and Statistics 105*, 1394–1410.

Ash, E., S. Asher, A. Bhowmick, S. Bhupatiraju, D. Chen, T. Devi, C. Goessmann, P. Novosad, and B. Siddiqi (2025, March). In-Group Bias in the Indian Judiciary: Evidence from 5 Million Criminal Cases. *Review of Economics and Statistics*, 1–45.

Baron, J., B. Jacob, and J. Ryan (2023, January). Pretrial juvenile detention. *Journal of Public Economics 217*, 104798.

Becker, G. S. (1957). *The Economics of Discrimination*. Chicago, IL: University of Chicago Press.

Bhuller, M., G. B. Dahl, K. V. Løken, and M. Mogstad (2020, April). Incarceration, Recidivism, and Employment. *Journal of Political Economy 128*(4), 1269–1324.

Bindler, A. and R. Hjalmarsson (2018, November). How Punishment Severity Affects Jury Verdicts: Evidence from Two Natural Experiments. *American Economic Journal: Economic Policy 10*(4), 36–78.

Bindler, A. and R. Hjalmarsson (2019, December). Path Dependency in Jury Decision Making. *Journal of the European Economic Association 17*(6), 1971–2017.

Black, D. A., J. Joo, R. LaLonde, J. A. Smith, and E. J. Taylor (2022, December). Simple tests for selection: Learning more from instrumental variables. *Labour Economics 79*, 1406–41.

Boyd, C. L., L. Epstein, and A. D. Martin (2010, April). Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science 54*(2), 389–411.

Carrell, S. E., M. E. Page, and J. E. West (2010, August). Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Eonomics 125*(3), 1101–1144.

Chen, D., T. Moskowitz, and K. Shue (2016, February). Decision-Making under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires. Technical Report w22026, National Bureau of Economic Research, Cambridge, MA.

Children's Law Center (2019). The family court process for children charged with criminal and status o¤enses: A brief overview of south carolina's juvenle delinquency proceedings.

Choi, D. D., J. A. Harris, and F. Shen-Bayh (2022, August). Ethnic Bias in Judicial Decision Making: Evidence from Criminal Appeals in Kenya. *American Political Science Review 116*(3), 1067–1080.

Cohen, A. and C. S. Yang (2019, February). Judicial Politics and Sentencing Decisions. *American Economic Journal: Economic Policy 11*(1), 160–191.

Damm, A. P., B. Ø. Larsen, H. S. Nielsen, and M. Simonsen (2025). Lowering the minimum age of criminal responsibility: Consequences for juvenile crime. *Journal of Quantitative Criminology*.

Dee, T. S. (2005, April). A Teacher Like Me: Does Race, Ethnicity, or Gender Matter? *American Economic Review 95*(2), 158–165.

Depew, B., O. Eren, and N. Mocan (2016, February). Judges, Juveniles and In-group Bias. Technical Report w22003, National Bureau of Economic Research, Cambridge, MA.

Dobbie, W., J. Goldin, and C. S. Yang (2018). The effects of pretrial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. *American Economic Review 108*(2), 201–240.

Eren, O. and N. Mocan (2018, July). Emotional Judges and Unlucky Juveniles. *American Economic Journal: Applied Economics 10*(3), 171–205.

Eren, O. and N. Mocan (2021, March). Juvenile Punishment, High School Graduation, and Adult Crime: Evidence from Idiosyncratic Judge Harshness. *Review of Economics and Statistics 103*(1), 34–47.

Flanagan, F. X. (2018, May). Race, Gender, and Juries: Evidence from North Carolina. *The Journal of Law and Economics 61*(2), 189–214.

Garin, A., D. K. Koustas, C. McPherson, S. Norris, M. Pecenco, E. K. Rose, Y. Shem-Tov, and J. Weaver (2024, July). The impact of incarceration on employment, earnings, and tax filing. Working Paper 32747, National Bureau of Economic Research.

Gazal-Ayal, O. and R. Sulitzeanu-Kenan (2010, September). Let My People Go: Ethnic In-Group Bias in Judicial Decisions—Evidence from a Randomized Natural Experiment. *Journal of Empirical Legal Studies 7*(3), 403–428.

Goncalves, F. and S. Mello (2021, May). A few bad apples? racial bias in policing. *American Economic Review 111*(5), 1406–41.

Guo, F., J. Li, M. Ma, and D. Zha (2025). Identity on the Bench: Gender In-Group Bias in the Judiciary.

Hester, R. (2017). Judicial rotation as centripetal force: Sentencing in the court communities of south carolina. *Criminology 55*(1), 205–235.

Hjalmarsson, R. (2009, November). Juvenile jails: A path to the straight and narrow or to hardened criminality? *Journal of Law and Economics 52*(4), 779–809.

Hjalmarsson, R. and L. Lochner (2012). The Impact of Education on Crime: International Evidence. CESifo DICE Report, ifo Institut - Leibniz-Institut für Wirtschaftsforschung an der Universität München, München.

Hoekstra, M. and C. Sloan (2022, March). Does Race Matter for Police Use of Force? Evidence from 911 Calls. *American Economic Review 112*(3), 827–860.

Hoekstra, M. and B. Street (2021, August). The Effect of Own-Gender Jurors on Conviction Rates. *The Journal of Law and Economics 64*(3), 513–537.

Hoffmann, F. and P. Oreopoulos (2009). A Professor Like Me: The Influence of Instructor Gender on College Achievement. *Journal of Human Resources 44*(2), 479–494.

Kling, J. R. (2006, May). Incarceration Length, Employment, and Earnings. *American Economic Review 96*(3), 863–876.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Lim, C. S., B. S. Silveira, and J. M. Snyder (2016, October). Do Judges' Characteristics Matter? Ethnicity, Gender, and Partisanship in Texas State Trial Courts. *American Law and Economics Review 18*(2), 302–357.

Lim, C. S. H., J. M. Snyder, and D. Strömberg (2015, October). The Judge, the Politician, and the Press: Newspaper Coverage and Criminal Sentencing across Electoral Systems. *American Economic Journal: Applied Economics 7*(4), 103–135.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005, July). Consistent Testing for Stochastic Dominance under General Sampling Schemes. *Review of Economic Studies 72*(3), 735–765.

Loeffler, C. E. and A. Chalfin (2017). Estimating the crime effects of raising the age of majority. *Criminology and public policy 16*, 45–71.

Loeffler, C. E. and B. Grunwald (2015). Decriminalizing delinquency: The effect of raising the age of majority on juvenile recidivism. *Journal of Legal Studies 44*.

Loeffler, C. E. and D. S. Nagin (2022, January). The Impact of Incarceration on Recidivism. *Annual Review of Criminology 5*(1), 133–152.

Mueller-Smith, M. (2015, August). The Criminal and Labor Market Impacts of Incarceration. *Working Paper*.

Mueller-Smith, M. G., B. Pyle, and C. Walker (2023, August). Estimating the impact of the age of criminal majority: Decomposing multiple treatments in a regression discontinuity framework. Working Paper 31523, National Bureau of Economic Research.

Norris, S., M. Pecenco, and J. Weaver (2021, September). The Effects of Parental and Sibling Incarceration: Evidence from Ohio. *American Economic Review 111*(9), 2926–2963.

Oreopoulos, P. and K. G. Salvanes (2011, February). Priceless: The Nonpecuniary Benefits of Schooling. *Journal of Economic Perspectives 25*(1), 159–184.

Parsons, C. A., J. Sulaeman, M. C. Yates, and D. S. Hamermesh (2011, June). Strike Three: Discrimination, Incentives, and Evaluation. *American Economic Review 101*(4), 1410–1435.

Philippe, A. and A. Ouss (2018, October). "No Hatred or Malice, Fear or Affection": Media and Sentencing. *Journal of Political Economy 126*(5), 2134–2178.

Price, J. and J. Wolfers (2010, November). Racial Discrimination Among NBA Referees. *The Quarterly Journal of Economics 125*(4), 1859–1887.

Puzzanchera, C., S. Hockenberry, and M. Sickmund (2022, December). Youth and the Juvenile Justice System. Technical report, National Center for Juvenile Justice.

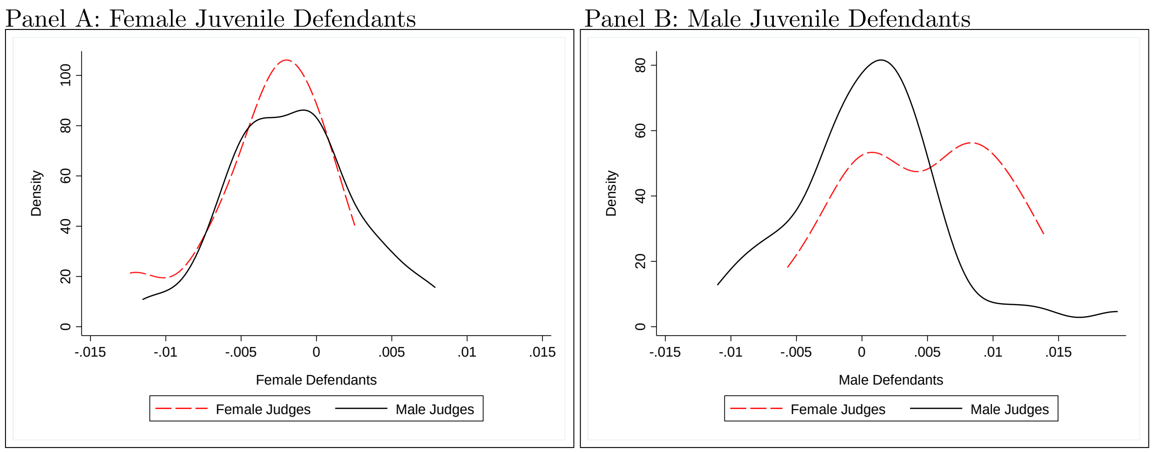
Shayo, M. and A. Zussman (2011, August). Judicial Ingroup Bias in the Shadow of Terrorism *. *The Quarterly Journal of Economics 126*(3), 1447–1484.

Smith, M., M. Hyman, and S. Redfield (2023). Addressing bias among judges. *State Court Report*, https://statecourtreport.org/our–work/analysis–opinion/addressing–bias–among–judges.

Zeltzer, D. (2020, April). Gender Homophily in Referral Networks: Consequences for the Medicare Physician Earnings Gap. *American Economic Journal: Applied Economics 12*(2), 169–197.

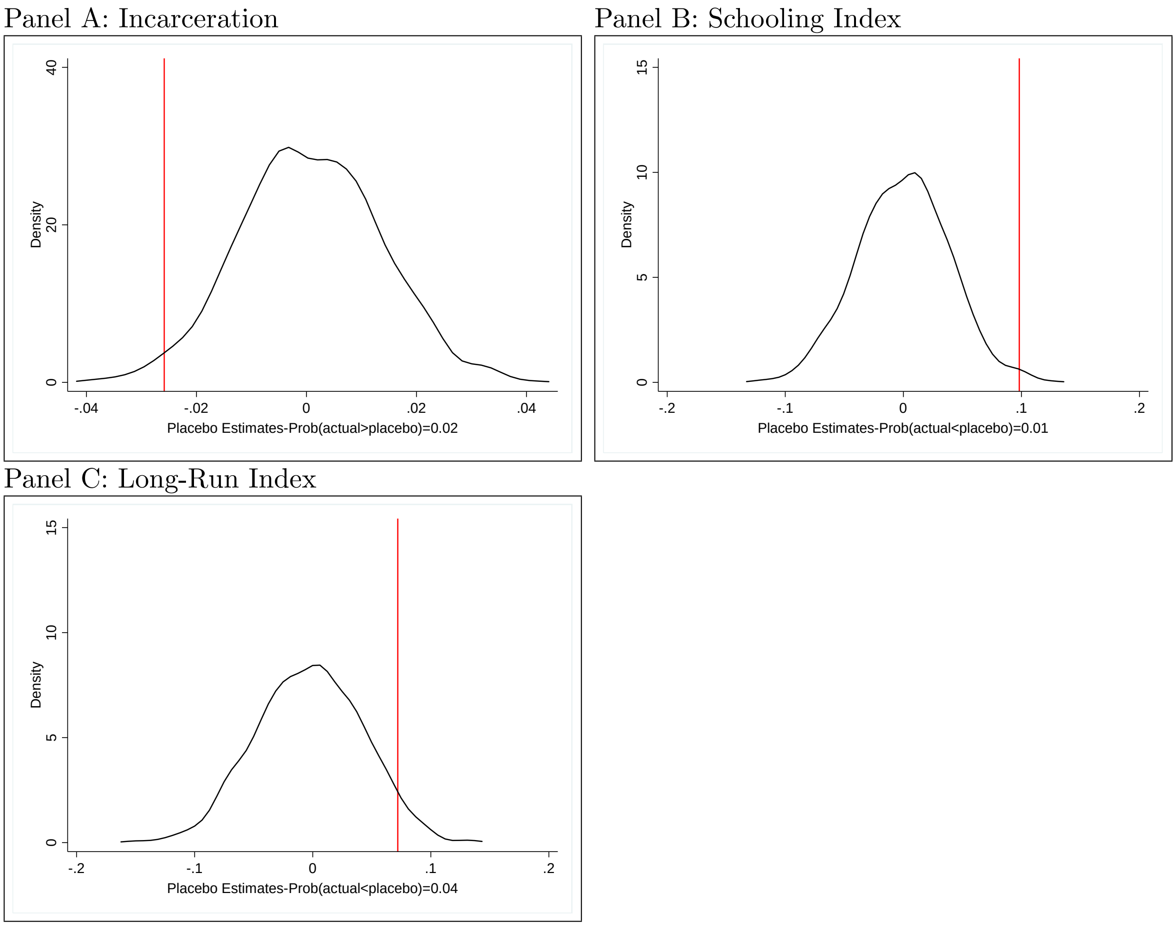
Tables and Figures:

Figure 1: Individual Judge Effects-Shrunken Estimates



Notes: Distribution of the individual judge effects by juvenile defendant's gender. We first regress the incarceration indicator on court-by-time fixed effects, along with case, juvenile, and observable judge characteristics, and retain the residuals. We then use these residuals to estimate a random effect model, obtaining shrunken estimates for each individual judge by defendant gender.

Figure 2: Placebo Estimates of the Effect of Juvenile Defendant-Judge Gender Interactions



Notes: Distribution of the coefficient estimates resulting from 1,000 sets of random assignment of gender to each judge in the analysis sample. The vertical lines denote the actual estimates. The fraction of placebo estimates smaller (larger) than the actual estimate is also reported on the x-axis in Panel A (Panels B and C).

Table 1: **Summary Statistics for Juveniles and Judges**

	Matched Sample		Unmatched Sample	
	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)
Panel A: Juvenile/Case Characteristics				
Incarceration	0.142	0.349	0.156	0.363
Convicted	0.738	0.439	0.755	0.430
Acquitted	0.119	0.324	0.088	0.284
Female	0.281	0.449	0.288	0.453
Black	0.561	0.496	0.577	0.494
White	0.409	0.492	0.382	0.486
Age at Disposition	15.31	1.62	15.48	1.45
Most Serious Charge:				
Property Offense	0.209	0.407	0.188	0.391
Violent Offense	0.156	0.362	0.159	0.367
School-Related Offense	0.221	0.415	0.195	0.396
Charge Severity	3.395	3.490	3.350	3.341
Charged with Multiple Offenses	0.324	0.468	0.333	0.472
Adult Outcomes:				
Enrolled in 12th Grade	0.166	0.372	0.189	0.392
Graduated High School within 4 Years	0.140	0.347	0.151	0.359
TANF Receipt	0.137	0.344	0.131	0.337
SNAP Receipt	0.798	0.402	0.755	0.430
Adult Arrest	0.636	0.481	0.645	0.478
Sample Size	23,569		38,933	
Panel B: Judge Characteristics				
Female	0.264	0.441		
Age at Disposition	53.32	8.70		
White	0.948	0.222		
Graduated from a College in the State	0.855	0.352		
Number of Judges	84			
Panel C: County Characteristics				
Number of Family Court Judges	6.98	3.92	13.92	5.83
Log of County Population	11.10	0.74	12.41	0.60
Unemployment Rate (%)	9.62	3.49	7.43	2.41

NOTES: This table reports summary statistics for the samples of juvenile case files that can and cannot be linked to disposition judges over the period from 2002 to 2012. Juvenile case files can be successfully matched to judges using the disposition date when cases are handled by a single family court judge on any given day. The statistics for adult outcomes in Panel A are weighted by the inverse of the number of times a juvenile defendant is observed in the data.

Table 2: **Randomization and Prediction Tests**

	Incarceration	Acquittal	Female Judge	Minority Judge	Judge Age	Judge Leniency
			Coefficients (Standard Errors)			
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.0457***	-0.0087*	0.0024	0.0001	-0.0100	-0.0002
	(0.0066)	(0.0051)	(0.0026)	(0.0015)	(0.0460)	(0.0002)
Black	0.0340***	0.0084**	0.0009	-0.0006	-0.0354	-0.0002
	(0.0079)	(0.0036)	(0.0033)	(0.0011)	(0.0614)	(0.0003)
White	-0.0339***	-0.0069	-0.0010	0.0002	0.0880	0.0002
	(0.0083)	(0.0043)	(0.0034)	(0.0014)	(0.0543)	(0.0003)
Age at Disposition	0.0256***	0.0002	-0.0002	0.0002	0.0494***	0.0000
	(0.0031)	(0.0022)	(0.0007)	(0.0004)	(0.0188)	(0.0001)
Property Offense	-0.0207*	0.0175**	0.0059	0.0020	-0.1290	0.0001
	(0.0110)	(0.0079)	(0.0040)	(0.0016)	(0.0811)	(0.0003)
Violent Offense	-0.0228***	0.0751***	-0.0007	0.0008	0.0569	0.0001
	(0.0068)	(0.0165)	(0.0049)	(0.0017)	(0.0577)	(0.0003)
School-Related Offense	-0.1359***	-0.0483***	-0.0040	-0.0367	-0.037	0.0001
	(0.0137)	(0.0090)	(0.0050)	(0.0024)	(0.0825)	(0.0004)
Offense Severity	0.0091***	0.0071***	-0.0002	0.0003	0.0005	0.0000
	(0.0016)	(0.0016)	(0.0005)	(0.0003)	(0.0086)	(0.0001)
Number of Charges at Disposition	0.0263***	-0.0089***	0.0002	-0.0001	-0.0014	0.0000
	(0.0041)	(0.0030)	(0.0014)	(0.0010)	(0.0238)	(0.0001)
Joint Significance (p-value)	0.00	0.00	0.54	0.93	0.16	0.73
Sample Size	23,569	23,569	23,569	18,090	23,108	23,569

NOTES: Each cell represents a separate regression and all regression estimations in Columns (1)-(5) control for court-by-year-month and court-by-day-of-week fixed effects. The dependent variable in the last column is the leave-one-out average non-incarceration rate. We construct the leave-one-out measure using the residuals from a regression of the non-incarceration indicator on court-by-time fixed effects. To allow for meaningful variation in judge leniency measure, court-by-time fixed effects in Column (6) include court-by-year-by-quarter fixed effects. Standard errors are two-way clustered at the judge and court levels. * significant at 10 %, ** significant at 5%, *** significant at 1%.

Table 3: **The Effect of Juvenile Defendant-Judge Gender Interactions on Judicial Dispositions**

	Incarceration				Acquitted	Predicted Incar.
	Coefficients (Standard Errors)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Judge	0.020	0.021				
	(0.015)	(0.013)				
Female Defendant*Female Judge	-0.024**	-0.026**	-0.026**	-0.024**	0.001	0.002
	(0.012)	(0.012)	(0.012)	(0.011)	(0.009)	(0.004)
Outcome Mean	0.142				0.119	
Sample Size	23,569	23,569	23,569	23,569	23,569	23,569
Controls:						
Court-by-Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Juvenile Characteristics	No	Yes	Yes	Yes	Yes	No
Judge Fixed Effects	No	No	Yes	Yes	Yes	Yes
Judge Gender Interactions	No	No	No	Yes	No	No

NOTES: All regression estimations control for court-by-year-month and court-by-day-of-week fixed effects. Juvenile controls include indicators for gender and race, birth year fixed effects, severity of the offense, offense fixed effects, number of charges at disposition, and number of times a juvenile defendant is observed in the data. The outcome variable in Column (6) is obtained from a regression of incarceration indicator on observable juvenile characteristics. Standard errors are two-way clustered at the judge and court levels. ** significant at 5%.

Table 4: **The Effect of Juvenile Defendant-Judge Gender Interactions on Long-Run Outcomes**

	Took the HS Exit Exam	Enrolled in 12th Grade	Graduate from HS in 4 Years	Schooling Index	Adult Arrest	TANF Receipt	SNAP Receipt	Long-Run Index
			Coefficients (Standard Errors)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Defendant*Female Judge	0.042***	0.037**	0.022*	0.098***	-0.006	-0.032	-0.001	0.072*
	(0.016)	(0.015)	(0.013)	(0.031)	(0.029)	(0.026)	(0.016)	(0.041)
Outcome Mean	0.412	0.166	0.140	0.000	0.636	0.137	0.798	0.000
Sample Size	15,416	23,569	17,629	23,569	23,569	23,569	23,569	23,569

NOTES: All regression estimations control for court-by-year-month, court-by-day-of-week and judge fixed effects. Juvenile controls include indicators for gender and race, birth year fixed effects, severity of the offense, offense fixed effects, and number of charges at disposition. The schooling index in Column (4) is constructed by averaging the z-scores of non-missing subcomponents from Columns (1)-(3). The long-run index in Column (8) is constructed by averaging the z-scores of schooling index, and indicators for being arrested of a violent offense and any welfare participation. All regressions are weighted by the inverse of the times a juvenile defendant is observed in the data. Standard errors are two-way clustered at the judge and court levels. * significant at 10% ** significant at 5%, *** significant at 1%.

Table 5: **Robustness Checks**

	Control Court-by-Year-Week Fixed Effects	Control Court-by-Year-Week-Day Fixed Effects	Add Detailed Offense Fixed Effects	Mandatory Sch Attendance Disp. Excluded	No Weighting	Expanded Sample
			Coefficients (Standard Errors)			
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Incarceration						
Female Defendant*Female Judge	-0.022*	-0.027**	-0.023*	-0.040**	-0.026**	-0.025***
	(0.012)	(0.012)	(0.012)	(0.016)	(0.012)	(0.007)
Sample Size	23,569	23,569	23,569	19,679	23,569	36,729
Panel B: Schooling Index						
Female Defendant*Female Judge	0.103***	0.114***	0.099***	0.089**	0.086**	0.097***
	(0.032)	(0.034)	(0.033)	(0.044)	(0.035)	(0.033)
Sample Size	23,569	23,569	23,569	19,679	23,569	36,729
Panel C: Long-Run Index						
Female Defendant*Female Judge	0.068	0.083*	0.071	0.082	0.061	0.068*
	(0.044)	(0.047)	(0.044)	(0.056)	(0.037)	(0.036)
Sample Size	23,569	23,569	23,569	19,679	23,569	36,729

NOTES: The first column reports defendant-judge female interaction effect from a model which conditions on court-by-year-week and court-by-day-of-week fixed effects, while Column (2) presents the results after controlling for court-by-year-week-day fixed effects. The specifications in all other columns control for court-by-year-month and court-by-day-of-week fixed effects. Column (3) adds detailed offense fixed effects to the baseline specification. There are 350 uniquely defined juvenile offense types in the analysis sample. Column (4) excludes all non-incarceration convictions involving mandatory school attendance. Column (5) reports the estimated effects without probability weights. The last column expands the sample by including days when all judges in a courthouse are of the same gender (without controlling for judge fixed effects). All columns, except the last, control for judge fixed effects. Standard errors are two-way clustered at the judge and court levels. Juvenile controls include indicators for gender and race, birth fixed effects, severity of the offense, offense fixed effects and number of charges at disposition. Panel A additionally controls for the number of times a juvenile defendant is observed in the data. The regressions in Panels B and C are weighted by the inverse of the of times a juvenile defendant is observed in the data. * significant at 10%, ** significant at 5%, *** significant at 1%.

Table 6: **Placebo Test**

	Incarceration	Schooling Index	Long-Run Index
	Coefficients (Standard Errors)		
	(1)	(2)	
Female Defendant*Female Judge	0.003	0.010	0.014
	(0.011)	(0.043)	(0.044)
Outcome Mean	0.149	0.000	0.000
Sample Size	18,353	18,353	18,353

NOTES: Juveniles case files are matched to judges using the referral date. All regression estimations control for court-by-year-month, court-by-day-of-week and judge fixed effects. Juvenile controls include indicators for gender and race, birth year fixed effects, severity of the offense, offense fixed effects, and number of charges at disposition. Standard errors are two-way clustered at the judge and court levels. The regressions in Columns (2) and (3) are weighted by the inverse of the number of times a juvenile defendant is observed in the data.

Table 7: **The Effect of Juvenile Defendant-Judge Gender Interactions on Long-Run Outcomes-By Race: Extended Sample**

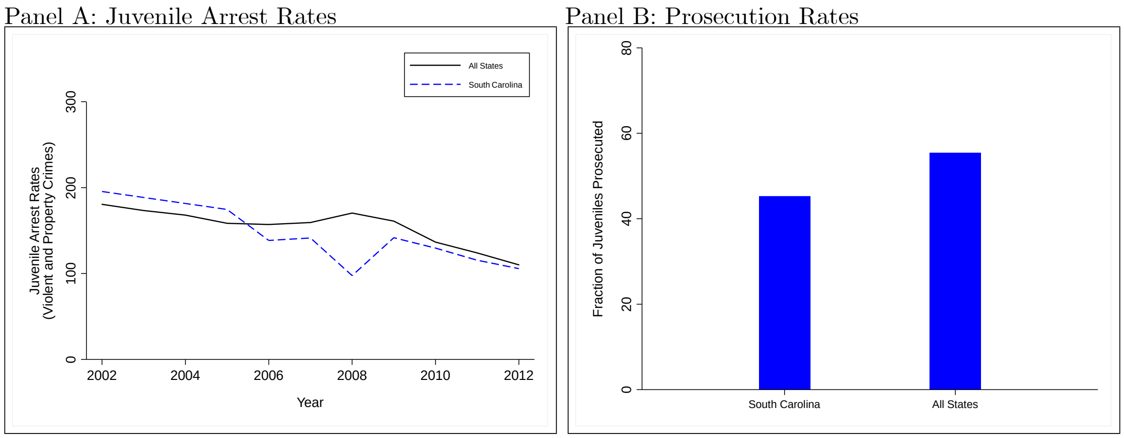
	Incarceration	Long-Run Index	Schooling Index	Violent Offense Arrest	TANF Receipt	SNAP Receipt
			Coefficients (Standard Errors)			
	(1)	(2)	(3)	(4)	(5)	(6)
White Female Defendant*Female Judge	-0.033***	-0.020	0.070	0.016	-0.010	0.016
	(0.012)	(0.050)	(0.061)	(0.011)	(0.020)	(0.021)
Black Female Defendant*Female Judge	-0.018**	0.103**	0.073	-0.027**	-0.058*	-0.008
	(0.009)	(0.044)	(0.046)	(0.013)	(0.030)	(0.024)
Outcome Mean-White Juveniles	0.132	0.095	-0.069	0.061	0.109	0.728
Outcome Mean-Black Juveniles	0.159	-0.089	0.062	0.111	0.158	0.834
Sample Size	36,257	36,257	36,257	36,257	36,257	36,257

NOTES: The sample includes days when all judges in a courthouse are of the same gender. All regression estimations control for court-by-year-month and court-by-day-of-week fixed effects. Juvenile controls include indicators for gender and race, birth year fixed effects, severity of the offense, offense fixed effects, and number of charges at disposition. The long-run index in Column (2) is constructed by averaging the z-scores of schooling index, and indicators for being arrested of a violent offense and any welfare participation. The schooling index in Column (3) is constructed by averaging the z-scores of non-missing subcomponents of educational outcomes. All regressions in Columns (2)-(6) are weighted by the inverse of the times a juvenile defendant is observed in the data. Standard errors are clustered at the court level. * significant at 10% ** significant at 5%, *** significant at 1%.

Online Appendix

Appendix A: Additional Tables and Figures

Figure A.1: Youth Arrest Rates and Fraction of Juveniles Prosecuted

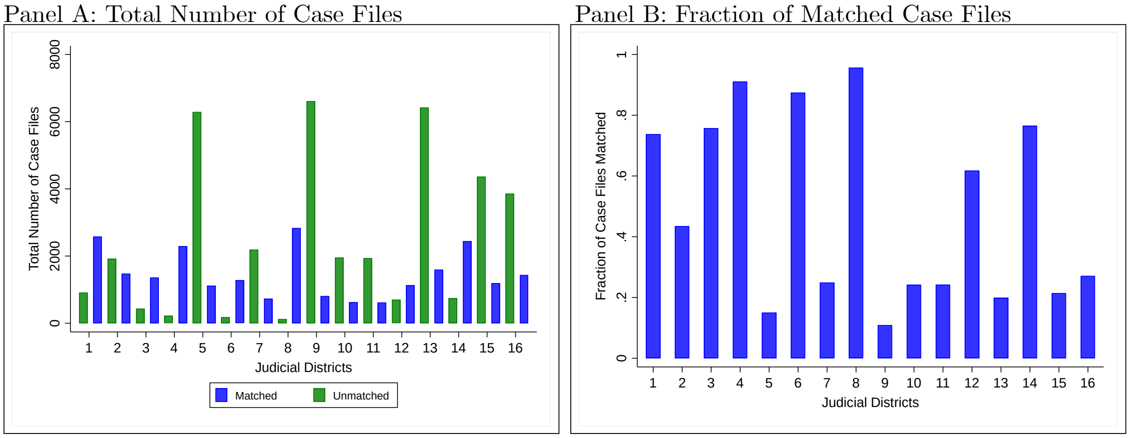


Notes: Panel A plots juvenile arrest rates for violent and property crimes combined from 2002-2012, defined as the number of arrests of individuals under age 18 for every 10,000 persons aged 10-17. Panel B presents the fraction of juvenile referrals that resulted in prosecution over the same period. The youth crime statistics (Panel A) and national estimates on the fraction prosecuted (Panel B) are drawn from the periodical Juvenile Court Statistics reports.

Figure A.2: Matched and Unmatched Case Files by Judicial Districts

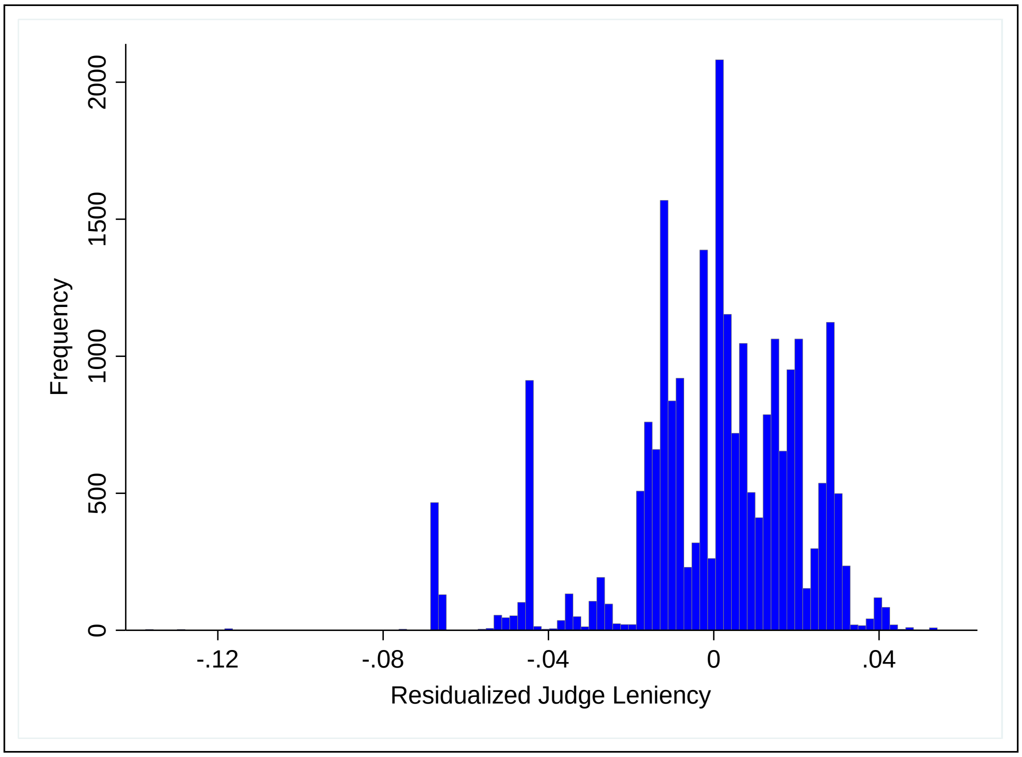
Panel A: Total Number of Case Files

Panel B: Fraction of Matched Case Files



Notes: The figure plots the total number (Panel A) and fraction of case files (Panel B) by judicial districts that can be linked to disposition judges over the period from 2002 to 2012. Juvenile case files can be successfully matched to judges using the disposition date when cases are handled by a single family court judge on any given date. The judicial system in structured into 16 judicial districts, each comprising a combination of state's 46 counties.

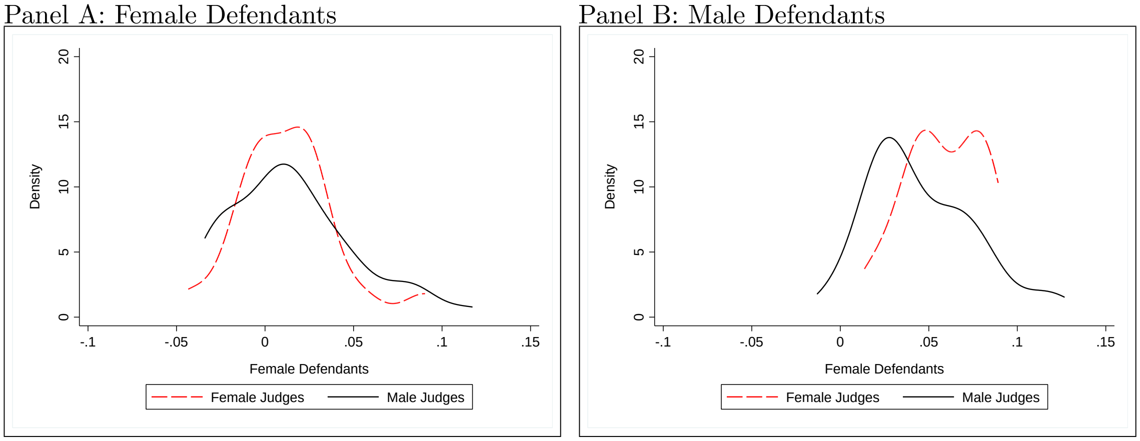
Figure A.3: Distribution of Judge Leniency



Notes: The histogram displays the distribution of the residualized judge leniency, obtained by the leave-one-out average non-incarceration rate among all other juvenile defendants assigned to a judge. We construct the leave-one-out measure using the residuals from a regression of the non-incarceration indicator on court-by-time fixed effects.

Figure A.4: Individual Judge Effects-Empirical Bayes Estimates

Panel A: Female Defendants Panel B: Male Defendants

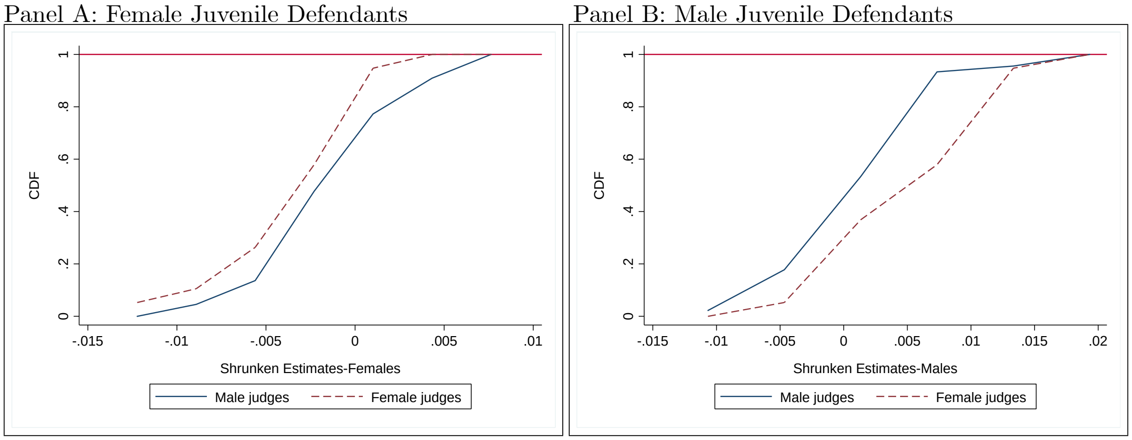


Notes: Distribution of the individual judge effects by juvenile defendant's gender. We regress the incarceration indicator on court-by-time fixed effects, along with case and juvenile characteristics. We then conduct a standard Empirical Bayes shrinkage correction to judge fixed effect estimates.

Figure A.5: Cumulative Distribution Functions: Individual Judge Effects-Shrunken Estimates

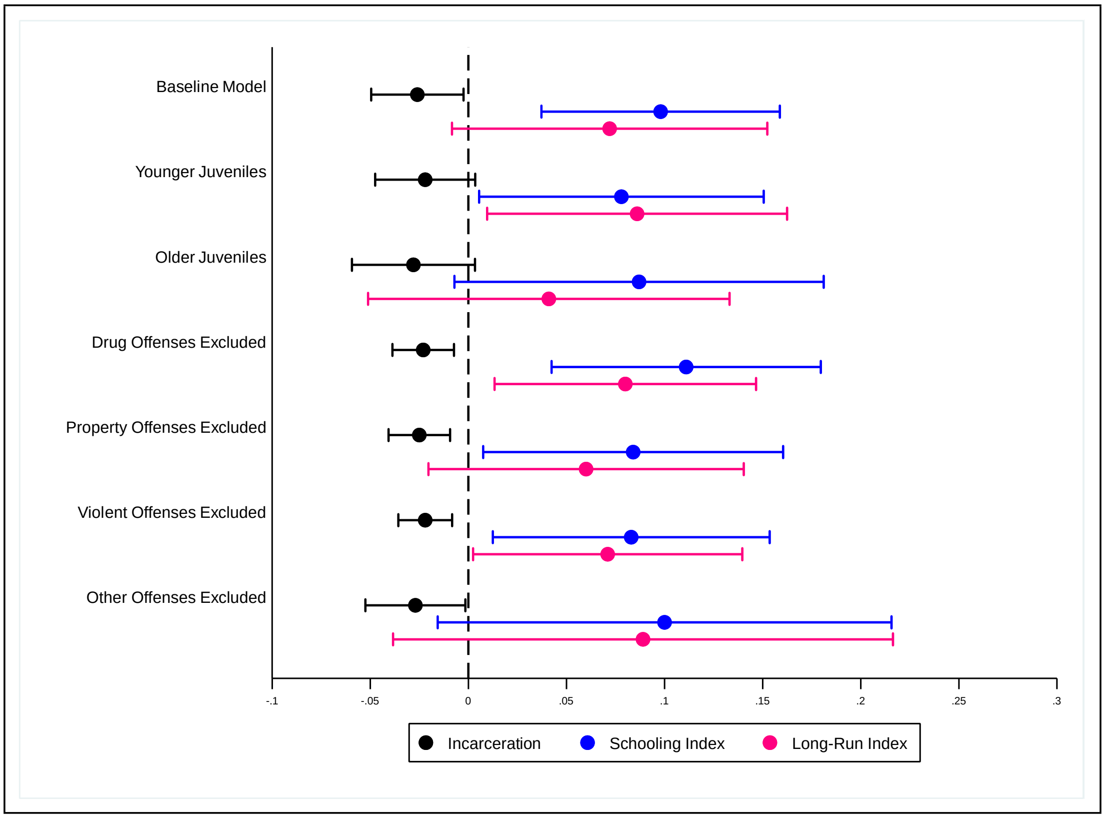
Panel A: Female Juvenile Defendants

Panel B: Male Juvenile Defendants



Notes: Cumulative distribution function of the individual judge effects by juvenile defendant's gender. We first regress the incarceration indicator on court-by-time fixed effects, along with case, juvenile, and observable judge characteristics, and retain the residuals. We then use these residuals to estimate a random effect model, obtaining shrunken estimates for each individual judge by defendant gender.

Figure A.6: Heterogeneous Effects by Age and Offense Types



Notes: The figure presents heterogeneous effects along the dimensions of age at first-time disposition (around age 15) and offense type, with the latter analyzed by dropping each of the four broad juvenile offense categories in turn. The sample includes days when all judges in a courthouse are of the same gender. All regression estimations control for court-by-year-month and court-by-day-of-week fixed effects. The schooling index is constructed by averaging the z-scores of non-missing components of educational outcomes. The long-run index is constructed by averaging the z-scores of schooling index, and indicators for being arrested of a violent offense and any welfare participation. The regressions for schooling and long-run index are weighted by the inverse of times a juvenile defendant is observed in the data. Each row also shows 95% confidence interval based on standard errors clustered at the court level.

Table A.1: **Summary Statistics by Juvenile's Gender**

	Female Juveniles		Male Juveniles	
	Mean	SD	Mean	SD
	(1)	(2)	(3)	(4)
Panel A: Juvenile/County Characteristics:				
Incarceration	0.100	0.300	0.158	0.365
Probation	0.797	0.402	0.715	0.451
Acquitted	0.102	0.303	0.126	0.332
Black	0.487	0.500	0.590	0.491
White	0.484	0.500	0.380	0.485
Age at Disposition	15.20	1.71	15.36	1.58
Most Serious Charge:				
Property Offense	0.102	0.303	0.251	0.434
Violent Offense	0.139	0.346	0.162	0.368
School-Related Offense	0.311	0.463	0.186	0.389
Charge Severity	2.229	2.324	3.852	3.754
Charged with Multiple Offenses	0.236	0.424	0.359	0.480
Panel B: Adult Outcomes:				
Enrolled in 12th Grade	0.205	0.404	0.150	0.357
Graduated High School within 4 Years	0.187	0.390	0.121	0.326
TANF Receipt	0.329	0.470	0.056	0.230
SNAP Receipt	0.873	0.334	0.766	0.423
Adult Arrest	0.490	0.500	0.697	0.460
Sample Size	6,630		16,939	

NOTES: This table reports summary statistics for the samples of juvenile case files that can be linked to disposition judges over the period from 2002 to 2012. Juvenile case files can be successfully matched to judges using the disposition date when cases are handled by a single family court judge on any given day. The statistics for adult outcomes are weighted by the inverse of the number times a juvenile defendant is observed in the data.

Table A.2: **Missingness of Judge and Defendant/Case Characteristics**

	Full Sample	Full Sample	Females	Males
	Coefficients (Standard Errors)			
	(1)	(2)	(3)	(4)
Incarceration		-0.0026	0.0044	-0.0047
		(0.0054)	(0.0052)	(0.0061)
Female	0.0021	0.0020		
	(0.0029)	(0.0029)		
Black	0.0005	0.0005	-0.0031	0.0014
	(0.0021)	(0.0021)	(0.0027)	(0.0026)
Age at Disposition	-0.0002	-0.0001	0.0000	-0.0002
	(0.0007)	(0.0007)	(0.0018)	(0.0009)
Property Offense	0.0018	0.0015	0.0037	0.0010
	(0.0033)	(0.0031)	(0.0074)	(0.0035)
Violent Offense	0.0015	0.0012	0.0036	0.0012
	(0.0034)	(0.0032)	(0.0093)	(0.0040)
School-Related Offense	0.0116**	0.0112**	0.0167**	0.0108**
	0.0053	(0.0052)	(0.0077)	(0.0052)
Offense Severity	-0.0005	-0.0004	-0.0001	-0.0004
	(0.0004)	(0.0004)	(0.0011)	(0.0004)
Number of Charges at Disposition	-0.0007	-0.0006	0.0013	-0.0014
	(0.0010)	(0.0010)	(0.0022)	(0.0011)
Joint Significance (p-value)	0.34	0.42	0.57	0.41
Sample Size	62,502	62,502	17,860	44,642

NOTES: All regression estimations control for court-by-year-month and court-by-day-of-week fixed effects. The dependent variable takes the value one if the juvenile defendant is not successfully matched to the disposition judge. Standard errors are clustered at the court level. ** significant at 5%.

Table A.3: **The Effect of Juvenile Defendant-Judge Gender Interactions on Judicial Dispositions-Different Time Windows**

	Number of Days Since Referral				Days Since Prosecution	
	Days<21	Days≥ 21	Days<10	Days≥ 10	Days<7	Days≥ 7
	Coefficients (Standard Errors)					
	(1)	(2)	(3)	(4)	(5)	(6)
Female Defendant*Female Judge	-0.043*	-0.020*	-0.012	-0.030**	-0.020	-0.029**
	(0.026)	(0.011)	(0.050)	(0.011)	(0.038)	(0.012)
Outcome Mean	0.247	0.118	0.282	0.125	0.226	0.126
Sample Size	4,437	19,132	2,616	20,953	3,692	19,877

NOTES: Columns (1)-(4) partition the analysis sample based on the number of days since initial referral to the juvenile justice system. Columns (5) and (6) divide the sample based on the number of elapsed days relative to prosecution. Juvenile controls include indicators for gender and race, birth year fixed effects, severity of the offense, offense fixed effects, number of charges at disposition, and number of times a juvenile defendant is observed in the data. All columns control for court-by-year-month, court-by-day-of-week and judge fixed effects. Standard errors are two-way clustered at the judge and court levels. * significant at 10%, ** significant at 5%.

Table A.4: **Robustness to Alternative Levels of Clustering**

	Clustered at the Judge Level	Clustered at the Child Level	Clustered at the Unit of Randomization
		Coefficients (Standard Errors)	
	(1)	(2)	(3)
Panel A: Incarceration			
Female Defendant*Female Judge	-0.026**	-0.026**	-0.026**
	(0.013)	(0.011)	(0.011)
Panel B: Schooling Index			
Female Defendant*Female Judge	0.098**	0.098**	0.098**
	(0.044)	(0.042)	(0.043)
Panel C: Long-Run Index			
Female Defendant*Female Judge	0.072	0.072*	0.072*
	(0.049)	(0.040)	(0.040)

NOTES: All regression estimations control for court-by-year-month, court-by-day-of-week and judge fixed effects. Juvenile controls include indicators for gender and race, birth year fixed effects, severity of the offense, offense fixed effects and number of charges at disposition. Panel A additionally controls for the number of times a juvenile defendant is observed in the data. All regressions in Panels B and C are weighted by the inverse of the number of times a juvenile defendant is observed in the data. * significant at 10%, ** significant at 5%.

Table A.5: **The Effect of Juvenile Defendant-Judge Gender Interactions on Long-Run Outcomes-By Race: Matched Sample**

	Incarceration	Long-Run Index	Schooling Index	Violent Offense Arrest	TANF Receipt	SNAP Receipt
	Coefficients (Standard Errors)					
	(1)	(2)	(3)	(4)	(5)	(6)
White Female Defendant*Female Judge	-0.046**	0.085	0.086	0.005	-0.029	0.019
	(0.017)	(0.061)	(0.054)	(0.014)	(0.026)	(0.023)
Black Female Defendant*Female Judge	-0.022*	0.221**	0.054	-0.020	-0.087***	-0.022
	(0.011)	(0.049)	(0.043)	(0.012)	(0.027)	(0.024)
Sample Size	17,874	17,874	17,874	17,874	17,874	17,874

NOTES: The analysis sample is restricted to matched case files with non-missing information on judge's race. All regression estimations control for court-by-year-month, court-by-day-of-week and judge fixed effects, and interactions of judge's race with defendant's gender and race. Juvenile controls include indicators for gender and race, birth year fixed effects, severity of the offense, offense fixed effects, and number of charges at disposition. The long-run index in Column (2) is constructed by averaging the z-scores of schooling index, and indicators for being arrested of a violent offense and any welfare participation. The schooling index in Column (3) is constructed by averaging the z-scores of non-missing subcomponents of educational outcomes. All regressions in Columns (2)-(6) are weighted by the inverse of the times a juvenile defendant is observed in the data. Standard errors are clustered at the court level. * significant at 10% ** significant at 5%, *** significant at 1%.

Table A.6: **The Effect of Juvenile Defendant-Judge Gender Interactions on Long-Run Outcomes-Never Incarcerated Juveniles**

	Schooling Index	Long-Run Index
	Coefficients (Standard Errors)	
	(1)	(2)
Female Defendant*Female Judge	0.097**	0.067
	(0.038)	(0.045)
Outcome Mean	0.051	0.042
Sample Size	17,238	17,238

NOTES: The analysis sample is restricted to juveniles who were never incarcerated. All regression estimations control for court-by-year-month, court-by-day-of-week and judge fixed effects. Juvenile controls include indicators for gender and race, birth year fixed effects, severity of the offense, offense fixed effects, and number of charges at disposition. All regressions are weighted by the inverse of the times a juvenile defendant is observed in the data. Standard errors are two-way clustered at the judge and court levels. ** significant at 5%.

Table A.7: **The Effect of Juvenile Defendant-Judge Gender Interactions on Other Margins**

	Acquitted	Mandatory Sch. Attendance	Monetary Restitution	Other Margins
	Coefficients (Standard Errors)			
	(1)	(2)	(3)	(4)
Female Defendant*Female Judge	0.001	-0.010	-0.004**	0.009
	(0.008)	(0.013)	(0.002)	(0.008)
Outcome Mean	0.103	0.113	0.008	0.068
Sample Size	36,729	36,729	36,729	36,729

NOTES: All regression estimations control for court-by-year-month and court-by-day-of-week fixed effects. Juvenile controls include indicators for gender and race, birth year fixed effects, severity of the offense, offense fixed effects, number of charges at disposition, and number of times a juvenile defendant is observed in the data. Standard errors are clustered at the court level. ** significant at 5%.

Appendix B: Further Estimation Details

Consistent Testing for Stochastic Dominance

Suppose S_f^F and S_f^M denote the random effects for N_0 female and N_1 male judges, respectively; these map to the individual judge effect estimates for female defendants (f). Let $F_f^F(s) = \Pr(S_f^F < s)$ represent the cumulative density function (CDF) of S_f^F; $F_f^M(s)$ is similarly defined for S_f^M. Under this notation, S_f^M first-order stochastically dominates (FSD) S_f^F if

$$F_f^M(s) \leq F_f^F(s) \quad \forall s \in S_f \text{ with strict inequality for some } s \tag{1}$$

where S_f denotes the union support of S_f^F and S_f^M. One can similarly formulate the FSD condition for male and female judges using individual judge effects for male defendants (m). Specifically, S_m^F FSD S_m^M can be written as

$$F_m^F(s) \leq F_m^M(s) \quad \forall s \in S_m \text{ with strict inequality for some } s. \tag{2}$$

To test for FSD, we utilize the following Kolmogorov-Smirnov test criteria:

$$d = \sqrt{\frac{N_0 * N_1}{N_0 + N_1}} \min \sup_{s \in S} [F^F(s) - F^M(s)]$$

where S denotes the union support of S^F and S^M and min is taken over $F^F - F^M$ and $F^M - F^F$, in effect performing two tests in order to leave no ambiguity between the equal and unrankabke cases. Following Linton et al. (2005), our consistent testing procedure calls for

1. Computing the empirical CDFs ($\widehat{F}(\cdot)$) over the support of S for s_k points ($k = 1...K, K$ is 500 in the application).
2. Computing the differences $d_1(s_k) = \widehat{F^F}(s_k) - \widehat{F^M}(s_k)$ and $d_2(s_k) = \widehat{F^M}(s_k) - \widehat{F^F}(s_k)$.
3. Obtaining $\widehat{d} = \sqrt{\frac{N_0 * N_1}{N_0 + N_1}} \min \{\max \{d_1\}, \max \{d_2\}\}$.

If $\widehat{d} \leq 0$ and $\max\{d_1\} < 0$, then S^F is observed to first-order dominate S^M; if $\widehat{d} \leq 0$ and $\max\{d_2\} < 0$, then the reverse is observed. If $\widehat{d} > 0$, then there is no ranking in the first-order sense.

Inference is conducted using a block bootstrapping procedure where resampling is done at the judge level. The bootstrap test statistic, $\widehat{d}^*$, is obtained by estimating a judge random effect model for each bootstrap replication of the original sample.[1] Under this resampling scheme, if $\Pr\left\{\widehat{d}^* \leq 0\right\}$ is large, say 0.90 or higher, and $\widehat{d} \leq 0$, we infer FSD to a desirable degree of confidence.

[1] The number of bootstrap replications is 500.