Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

In-group bias in the Indian judiciary: Evidence from 5 million criminal cases

Elliott Ash, Sam Asher, Aditi Bhowmick, Sandeep Bhupatiraju, Daniel Chen, Tanaya Devi, Christoph Goessmann, Paul Novosad, Bilal Siddiqi*

December 2, 2024

Abstract

We study judicial in-group bias in Indian criminal courts using newly collected data on over 5 million criminal case records from 2010–2018. After classifying gender and religious identity with a neural network, we exploit quasi-random assignment of cases to judges to determine whether judges favor defendants with similar identities to themselves. In the aggregate, we estimate tight zero effects of in-group bias based on shared gender or religion, including in settings where identity may be especially salient, such as when the victim and defendant have discordant identities. Proxying caste similarity with shared last names, we find a degree of in-group bias, but only among people with rare names; its aggregate impact remains small.

JEL codes: J15, J16, K4, O12

*Ash: ETH Zurich, Asher: Imperial College London, Bhowmick: Development Data Lab, Bhupatiraju: World Bank, Chen: Toulouse School of Economics and World Bank, Devi: Harvard, Goessmann: ETH Zurich, Novosad: Dartmouth College, Siddiqi: IDinsight.

We thank Alison Campion, Rebecca Cai, Nikhitha Cheeti, Kritarth Jha, Romina Jafarian, Ornelie Manzambi, Chetana Sabnis, and Jonathan Tan for helpful research assistance. We thank Emergent Ventures, the World Bank Research Support Budget, the World Bank Program on Data and Evidence for Justice Reform, the UC Berkeley Center for Effective Global Action, the DFID Economic Development and Institutions program, and The Bright Initiative for financial support. For helpful feedback we thank participants of the NBER Summer Institute Crime Working Group, Political Economy Seminar at ETH Zurich, Delhi School of Economics Winter School 2020, Texas Economics of Crime Workshop, Midwest International Economic Development Conference, Dis-

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

1 Introduction

This paper examines bias in India's courts, asking whether judges deliver more favorable treatment to defendants who match their identities. The literature suggests that judicial bias along gender, religious, or ethnic lines is pervasive in richer countries, having been identified in a wide range of settings around the world.[1] However, it has not been widely studied in the courts of lower-income countries. In-group bias of this form has been identified in other contexts in India, such as among loan officers (Fisman et al., 2020), election workers (Neggers, 2018), and school teachers (Hanna & Linden, 2012). But the judicial setting is of particular interest, given the premise that individuals who are discriminated against in informal settings can find recourse via equal treatment under the law (Sandefur & Siddiqi, 2015).

We focus on the dimensions of gender, religion, and caste, motivated by growing evidence that India's women, Muslims, and lower castes do not enjoy equal access to economic or other opportunities (Ito, 2009; Bertrand et al., 2010; Hanna & Linden, 2012; Jayachandran, 2015; Borker, 2021; Asher et al., 2024). Women represent half the population but only 27% of district court judges. Similarly, India's 200 million Muslims represent 14% of the population but only 7% of district court judges. We examine whether unequal representation in the courts has a direct effect on the judicial outcomes of women, Muslims and lower castes, in the form of judges delivering

crimination and Diversity Workshop at the University of East Anglia, Seminar in Applied Microeconomics Virtual Assembly and Discussion (SAMVAAD), Women in Economics and Policy seminar series, UC Berkeley Development Economics brown bag series, ACM SIGCAS Conference on Computing and Sustainable Societies (2021), German Development Economics Conference, Evidence in Governance and Politics (EGAP) seminar series, the Yale Race, Ethnicity, Gender, and Economic Justice Virtual Symposium, the Penn Center for the Advanced Study of India, and researchers at the Vidhi Center for Legal Policy.

[1] See, for example, Shayo and Zussman (2011), Didwania (2022), Arnold et al. (2018), Abrams et al. (2012), Alesina and La Ferrara (2014), Anwar et al. (2019) and others below.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

better outcomes to criminal defendants who match their identities.

Our analysis draws upon a new dataset of 5 million criminal court cases covering 2010–2018, constructed from case records scraped from an online government repository for cases heard in India's trial courts.[2] These cases are drawn from a dataset covering the universe of India's 7,000+ district and subordinate trial courts, staffed by over 80,000 judges; for context, the American judiciary is less than half as large, with 31,700 judges (University of Denver Institute for the Advancement of the American Legal System, 2021). We have released an anonymized version of the dataset, opening the door to many new analyses of the judicial process in the world's largest democracy and largest common-law legal system.[3]

An initial challenge with the case data is that it does not include the identity characteristics of judges and defendants. To address this issue, we build a neural-net-based classifier to assign gender and religion based on the text of names. The classifier is trained on a collection of millions of names from the Delhi voter rolls (labeled for gender) and the National Railway Exam (labeled for religion). The deep neural net classifier is sensitive to distinctive sequences of characters in the names, allowing us to classify individuals by gender and religion with over 97% out-of-sample accuracy on both dimensions, significantly higher than the standard approach of fuzzy matching.[4] We apply the trained model to our case dataset to assign identity characteristics to judges, defendants, and victims.

We examine whether judges treat defendants differently when they share the same gender or religion. We focus on the subset of cases filed under India's criminal codes, where acquittal and conviction rates can be interpreted as positive and negative outcomes, respectively. Given the ex-

[2]The eCourts platform can be accessed at https://ecourts.gov.in/.

[3]The data can be accessed at https://www.devdatalab.org/judicial-data. The complete dataset — civil and criminal, without filtering — contains 77 million case records.

[4]The name classifier code is available as an open-source software package, see https://github.com/devdatalab/paper-justice/tree/main/classifier. The trained gender classifier model is also available at that link, while the religion classifier is available to researchers upon request.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

treme delays in India's judicial system (Tata Trusts, 2019; Rao, 2020), we additionally examine whether in-group judge identity affects the court's speed in reaching a decision.

We exploit the arbitrary rules by which cases are assigned to judges, generating as-good-as-random variation in judge identity. Our preferred specification includes court-year-month and charge fixed effects. This approach effectively compares the outcomes of two defendants with the same identity classification, charged under the same criminal section, in the same court and in the same month, but who are assigned to judges with different identities.

In the aggregate, we find that sharing gender or religion with a defendant makes a judge no more or less likely to deliver an acquittal. The confidence intervals rule out effect sizes that are an order of magnitude smaller than nearly all prior estimates of in-group bias based on similar identification strategies in the literature. The exception is Lim et al., 2016, who find little evidence of in-group gender or racial bias among judges in Texas, notably the most statistically powered study in this class before ours. The upper end of the 95% confidence interval in our primary specification rejects a 0.6-percentage-point effect size in the worst case; studies using the same identification strategy in other contexts have routinely found bias effects ranging from 5 to 20 percentage points.[5]

We also examine speed of decision as an outcome. We can again rule out a substantial in-group bias effect — the 95% confidence interval excludes a one percentage point change in either direction in the likelihood that a case is resolved within six months. However, in some specifications, we find that same-gender judges are 0.4 percentage points more likely to conclude a case within this timeframe.

Notwithstanding a null effect of in-group bias on average, bias could be activated in contexts where judge and defendant identity are more salient. We examine four special contexts that the literature suggests may prime in-group bias (Mullen et al., 1992; Shayo & Zussman, 2011; Anwar et al., 2012; Mehmood et al., 2023). First, we examine cases where the defendant and the victim

[5]Judge demographics are not irrelevant to outcomes, however. We find that Muslim judges have a one percentage point higher acquittal rate than non-Muslim judges, and are slightly less likely to resolve a case quickly.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

of the crime have different identities. Sharing an identity with the victim when the defendant is in an out-group could, by creating an external reference point, activate the judge's sense of opposite identity with the defendant. Second, we examine gender bias in criminal cases categorized as crimes against women, which are mostly sexual assaults and kidnappings. Here, the shared identity of gender is intrinsic to the substance of the case and may thus be more salient. Third, we examine whether in-group bias on the basis of religion is activated during or following religious festivals, which may prime religious identity. We continue to find a null bias in all of these settings.[6]

Fourth and finally, we examine in-group bias on the basis of caste. We follow Fisman et al. (2017) and define defendant and judge to be in the same social group when they share the same last name. As above, we find a null relationship between acquittal and assignment to a judge with the same last name. However, we do find an increased probability of acquittal when judge and defendant names match, and the name is uncommon, defined as below-median frequency in the defendant sample. The effect is economically important for these defendents (about a 10% higher chance of acquittal), though not statistically significant in all specifications. The overall same-name effect is small in the aggregate because it applies to a small subset of defendants who both have uncommon names and are lucky enough to be assigned a judge with the same uncommon name. Nevertheless, this effect demonstrates that judges do display some degree of in-group bias, and it may also exist on other markers of caste that we do not observe.

Our estimates do not rule out bias on the basis of identity in a general sense. For example, both Muslim and non-Muslim judges could discriminate against Muslims and both male and female judges could provide unfair judgments to women (as found for Black defendants in U.S. courts by

[6]Our null effects are notable as compared to Mehmood et al. (2023), who find that acquittal rates in Pakistan rise by 23 percentage points (or 40%) during Ramadan, and that they rise by 7 percentage points in India for each additional hour of fasting. Mehmood et al. (2023) do not examine differential outcomes for Muslim and non-Muslim defendants and hence do not study in-group bias. We do not exploit differences in daylight hours in our study because there is little variation in the timing of Ramadan across the 8 years in the study.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Arnold et al. (2018) and Arnold et al. (2022), for example). There could also be bias higher up the judicial pipeline: arrests and/or charges may disproportionately target Muslims, or charges brought by women may not be taken as seriously by the police. Our null estimates are nevertheless notable, as prior studies with very similar designs have found substantial degrees of in-group bias in many other settings.

In Section 6, we discuss several reasons that bias could be small in our setting, given its apparent ubiquity in other judicial settings and other Indian contexts. At face value, the results suggest that rule-of-law institutions and judicial norms effectively prevent favoritism for in-groups. Other factors that might influence the degree of bias include the extent that the context is adversarial or cooperative, the class distance between judge and defendant, or, as suggested by the results on uncommon last names, the overall salience of the shared identity group.

Our finding of in-group bias only in one setting where identity is particularly salient is informative for our understanding of prior work, which consistently finds large in-group effects in the judicial domain. The most similar prior studies focus on the United States and Israel, institutional contexts where race, ethnic, or religious identity may be exceptionally salient. The U.S. incarceration system, in particular, has reproduced many aspects of the slave system that preceded it (Alexander, 2010). With this historical legacy, it is perhaps unsurprising to find that defendant race is a highly salient feature of many U.S. criminal cases.

Another potential contributing factor could be publication bias in the social-science literature on judicial bias, such that contexts *without* in-group bias are not prominently described in completed papers. To assess this possibility, we aggregate the effect sizes and standard errors from earlier papers with highly similar empirical designs to ours. Following the approach from Andrews and Kasy, 2019, we find evidence consistent with a high degree of publication bias. The Andrews and Kasy (2019) estimator suggests that statistically significant findings of in-group bias are about 30 times more likely to make it from conception to publication.

Our study makes four contributions. First, contrary to most of the existing literature, we demonstrate a notable absence of judicial in-group bias in an important low-income-country context with

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

substantial religious, ethnic, and gender-based cleavages. Because the size of our sample is orders of magnitude larger than nearly all prior studies, we are able to measure this (absence of) bias much more precisely than prior work. Second, our finding of in-group bias when the reference group is small and more salient may shed light on contexts where bias may be more or less likely to occur. In particular, the large and significant bias results for Jewish versus Arab defendants in Israel, and Black versus White defendants in the U.S. (described below), are found in contexts where ethnic identity is salient to the extreme, in-groups are well-defined and recognizable, and external cross-group tensions are heightened. Third, we provide evidence that the existing body of knowledge on in-group bias in judicial settings suffers from a substantial degree of publication bias, which has implications well outside the Indian context. Fourth, we have made public a 77 million case dataset which is already enabling a range of future research projects in this domain.

Our results add to the literature on biased decision-making in the legal system. Most prior work is on the U.S. legal system, where disparities have been documented at many levels.[7] The closest

[7]These include racial disparities in the execution of stop-and-frisk programs (Goel et al., 2016), motor vehicle searches by police troopers (Anwar & Fang, 2006), bail decisions (Arnold et al., 2018; Arnold et al., 2022), charge decisions (Rehavi & Starr, 2014), and judge sentence decisions (Mustard, 2001; Abrams et al., 2012; Alesina & La Ferrara, 2014; Kastellec, 2013). African-American judges have been found to vote differently from Caucasian-American judges on issues where minorities are disproportionately affected, such as affirmative action, racial harassment, unions, and search and seizure cases (Scherer, 2004; Chew & Kelley, 2009; Kastellec, 2011). In a similar manner, a number of papers have documented the effect of judges' gender in sexual harassment cases (Boyd et al., 2010; Peresie, 2005). A smaller set of papers use information on both the identity of the defendant and the decision-maker. Anwar et al. (2012) look at random variation in the jury pool and find that having a Black juror in the pool decreases conviction rates for Black defendants. A similar result from Israel is documented by Grossman et al. (2016), who find that the effect of including even one Arab judge on the decision-making panel substantially influences trial outcomes of Arab defendants. Didwania, 2022 find in-group bias in that prosecutors charge

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

paper to ours is Shayo and Zussman (2011), who analyze the effect of assigning a Jewish versus an Arab judge in Israeli small claims court. They find robust evidence of in-group bias: Jewish judges favor Jewish defendants and/or Arab judges favor Arab defendants.

A handful of other studies use quasi-random designs to estimate in-group bias in similar fashion to us. While most of these papers report large and statistically significant pro-in-group effects, one paper finds anti-in-group bias.[8] Of the papers we could find, only Lim et al. (2016) find a null in-group effect of judge ethnicity or gender.

In the Indian context, there is a growing body of evidence on the legal system, mostly focusing on judicial efficacy and economic performance (Chemin, 2009; Rao, 2020), and on corruption in the Indian Supreme Court (Aney et al., 2021). Bharti and Roy (2023) uses similar data to us to examine how judge childhoods affect their future decisions. Beyond the issue of in-group bias, we add to the growing literature on courts in developing countries. Well-functioning courts are widely considered a central component of effective, inclusive institutions, with judicial equity and rule of law seen as key indicators of a country's institutional quality (Rodrik, 2000; Le, 2004; Rodrik, 2005; Pande & Udry, 2005; Visaria, 2009; Lichand & Soares, 2014; Ponticelli & Alencar, 2016; The World Bank Group, 2017). A handful of important cross-country studies have recovered some broad stylized facts on the causes and consequences of different broad features of legal systems (Djankov et al., 2003; La Porta et al., 2004; La Porta et al., 2008). But largely due to a lack of data, there has been a relative paucity of within-country court- or case-level research on the delivery of justice in lower-income settings.

---

same-gender defendants with less severe offenses.

[8]Gazal-Ayal and Sulitzeanu-Kenan (2010) find positive in-group bias in bail decisions when Arab and Jewish defendants are randomly assigned to a judge of the same ethnicity. Knepper (2018) and Sloane (2019) leverage random assignment of cases in the U.S. to judges and prosecutors respectively, finding significant in-group bias in trial outcomes. Depew et al. (2017) exploit random assignment of judges to juvenile crimes in Louisiana and find *negative* in-group bias in sentence lengths and likelihood of being placed in custody.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

The rest of the paper is organized as follows. After outlining the institutional context (Section 2) and data sources (Section 3), we articulate our empirical approach (Section 4). Section 5 reports the results. Section 6 compares the results to the previous literature and concludes. Replication code and data are posted in a public repository, along with a gender classification web app.[9]

2 Background

2.1 Gender, Caste and Religion in India

India's population is characterized by cross-cutting divisions between gender and religion. Women's rights and their status in society are under intense political debate. Women constitute 48% of the population, and remain vulnerable to social practices such as female infanticide, child marriage, and dowry deaths despite existing legislation outlawing all of the above. India accounts for one third of all child marriages globally (Cousins, 2020) and nearly one third of the 142.6 million missing females in the world (Erken et al., 2020).

Muslims in India (14% of the population) have historically had intermediate socioeconomic outcomes worse than upper caste groups but better than lower caste groups (Sachar Committee Report, 2006). However, they have been protected by few of the policies and reservations targeted to Scheduled Castes and Tribes (17% and 9% of the population, respectively). In recent decades, many successful political parties have been accused of implicitly or explicitly discriminating against Muslims. The marginalized statuses of women and Muslims in India motivate our exploration of the role of gender and religion in the context of India's criminal justice system.

2.2 India's Court System

India's judicial system is organized in a jurisdictional hierarchy, similar to other common-law systems. There is a Supreme Court, 25 state High Courts, and 672 district courts below them. Beneath the district courts, there are about 7000 subordinate courts. The district courts and subordinate

[9]The repository can be found at https://github.com/devdatalab/paper-justice/.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

courts (which we study here) collectively constitute India's lower judiciary. These courts represent the point of entry of almost all criminal cases in India.[10]

These courts are staffed by over 80,000 judges. Due to common law institutions where court rulings serve as binding precedent in future cases, judges in India are effectively policymakers. Indian judges are arguably even more powerful than their U.S. counterparts because they do not share decision authority with juries, which were banned in 1959. Therefore, fair and efficient decision-making by judges is a leading issue for governance.

Lower-court judges in India are appointed by the governor in consultation with the state high court's chief justice. At least seven years of legal practice are required as a minimum qualification. The recruitment process entails a written examination and oral interview by a panel of higher-court judges. Judge tenure is in general well-protected, with removal by the governor only possible with the agreement of the high court. Finally, district judges can be promoted to higher offices in the judiciary after specific numbers of years in their post.

There is an active debate in India around reforming the court system. Problems under discussion include a reputation for corruption (Dev, 2019), a substantial backlog of cases (Tata Trusts, 2019), and judicial independence (The Economist, 2024).

2.3 Case Assignment to Judges

The procedure of case assignment to judges is pivotal for this study because our empirical strategy hinges on the exogenous assignment of judges to cases. To better understand the case assignment process, we consulted with several criminal lawyers who practice in India's district courts, senior research fellows at the Vidhi Center for Legal Policy, and several clerks in courts around the country.

Criminal cases are assigned to judges as follows. First, a crime is reported at a particular local police station, where a First Information Report (FIR) is filed. Each police station lies within the territorial jurisdiction of a specific district courthouse, which receives the case. The case is then

[10]We define criminal cases as all cases filed either under the Indian Penal Code Act or the Code of Criminal Procedure Act.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

assigned to a judge sitting in that courthouse. If there is just one judge available to see cases in the courthouse, that judge gets the case.

If there are multiple judges, a rule-based process fully determines the judge assignment. Each judge sits in a specific courtroom in a court for several months at a time. A courtroom is assigned for every police station and every charge. For example, at a given police station, every murder charge will go to the same courtroom. A larceny charge might go to a different courtroom, as might a murder charge reported at a different police station. The police station charge lists leave little room for discretion over which charges are seen by which judges.[11]

Judges typically spend two to three years in a given court, during which they rotate through several of the courtrooms.[12] Given judicial delays, the timing of the first court appearance is unknown when charges are filed. Thus, even if a defendant or prosecutor had discretion over which police station filed the charges, the rotation of judges between courtrooms would make it difficult to target a specific judge.

Finally, the judiciary explicitly condemns the practice of "judge shopping" or "forum shopping," where litigants select particular judges in search of a favorable match. One of the earliest cases in which the Indian Supreme Court condemned the practice of shopping is the case of *M/s Chetak Construction Ltd. v. Om Prakash & Ors.*, 1998(4) SCC 577, where the Court ruled against a litigant trying to select a favorable judge, writing that judge shopping "must be crushed with a heavy hand." This decision has been cited heavily in subsequent judgments.

In U.S. courts, a large share of criminal cases are disposed through plea bargaining, making appearance in court itself an endogenous outcome. This is not a concern in our context. While plea bargaining exists in India, fewer than 0.05% of criminal cases end in plea bargains (National Crime

[11] Since 2013, there has been a random assignment lottery mechanism available through the eCourts platform, but few courts have adopted it to date.

[12] Severe cases (with severity defined by the section or act under which the charge was filed) require judges with higher levels of seniority. Thus, a case in a given district may be eligible to be seen only by a subset of judges in that district.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Records Bureau, 2018).[13]

3 Data

3.1 Case Records

We obtained 77 million case records from the Indian eCourts platform — a public system put in place by the Indian government to host summary data and full text from orders and judgments in courts across the country.[14] The database includes both the PDF documents describing the judge's orders for each case (a series of typically 1–50 page documents), and a set of metadata fields that have been coded by eCourts analysts based on the case documents. We use the coded metadata only; classifying fields from the universe of PDFs will be a multi-year undertaking and is left for future work. The case metadata includes information on the filing, registration, hearing, and decision dates for each case, the petitioner and respondent names, the position of the presiding judge, the acts and sections under which the case was filed, and the final decision or disposition.[15]

The database covers India's lower judiciary, consisting of all courts including and under the jurisdiction of District and Sessions courts and covers the period 2010–2018. Appendix Figure A2 maps the geographic distribution of our sample of courts, which covers the whole country. This paper focuses on cases filed either under the Indian Penal Code or the Code of Criminal Procedure, for two reasons. First, there is only a single litigant, rather than two, providing a clear definition of identity match between judge and defendant. Second, it is relatively straightforward to identify

[13]Plea bargaining has only been available as a resolution mechanism in India since 2007, for cases with a maximum sentence of less than seven years. It is rarely sought by prosecutors. Legal experts suggested to us that it is rarely a good option for defendants with lawyers, who expect reasonable odds of winning at trial, and that unrepresented defendants may not be aware of the possibility of settlement.

[14]https://ecourts.gov.in/ecourts_home/static/about-us.php, accessed Oct 14, 2020

[15]We illustrate such a record in Appendix Figure A1.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

good and bad outcomes for criminal defendants, which is more difficult in civil cases. This constraint filters out 70% of the dataset, leaving us with 23 million criminal case records (Appendix Figure A3).

3.2 Judge Information

We obtained data on judges in all courts in the Indian lower judiciary from the eCourts platform. The data for each judge includes the judge's name, their position or designation, and the start and end date of the judge's appointment to each court.[16]

We joined the case-level data with the judge-level data based on the judge's designation and the initial case filing date. In this process, another 17% of the initial observations are dropped. The remaining dataset where cases are linked to a unique judge consists of 10 million cases. From this subset, we drop all bail decisions, which are a narrow share of the data. We then drop cases where we cannot identify both defendant and judge identity (depending on whether we are analyzing religion or gender, see below). Finally, we drop cases in courts where there is only one judge in a given time period. This leaves 5.7 million cases in the religion analysis and 5.3 million in the gender analysis (Appendix Figure A3).

3.3 Assigning Religion and Gender Identity

Demographic metadata is not included on India's eCourts platform, so we assign gender and religious identity based on names. As described in detail in Appendix B.1, we build a machine learning classifier that predicts gender (male/female) and religion (Muslim/non-Muslim) based on the name text fields. The training data are the Delhi voter rolls (13.7 million names for gender) and the National Railway Exam (1.4 million names for religion). A recurrent neural net architecture reads all the characters in the name and understands them in context, so it is sensitive to nuanced name variations that would confound standard matching algorithms.

[16] See Appendix Figure A4 for a sample page from which we extract the judge data. The data does not include the room in the court to which a judge is assigned.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Accuracy is validated on hold-out test sets of unseen names. Balanced accuracy and F1 scores are near .98 for both gender and religion classifications, which indicates robust predictive performance.

We then apply the classifier to the eCourts data, filtering out incomplete and low-confidence names (below 65% predicted probability). We can classify 96% of judges for gender, 98% of judges for religion, 74% of defendants for gender, and 80% of defendants for religion. Manual checks confirm a 97% accuracy in the eCourts dataset, with additional validation showing strong correlation between the classifier's Muslim share estimates and census data.

3.4 Defining Case Outcomes

We define the defendant's outcome (represented by Y below) as a case-level indicator variable that takes the value one if the disposition is desirable for the defendant and zero otherwise. In the "favors-defendant" category, the main disposition is acquittal, with a couple of other equivalent dispositions such as dismissal. Tabulations on these and other dispositions are shown in Appendix Table A4. The overall acquittal rate — that is, share of dispositions favoring the defendant — is about 15%.

The other 85% of cases, coded as $Y = 0$, represent a somewhat broader "does-not-favor-defendant" category. This group includes, first, the cases where the defendant is convicted, as well as a set of equivalent outcomes indicating a guilty verdict (see Appendix Table A4). Second, it includes cases that do not have a disposition at all because they have not been resolved yet. Third, a case may be closed but with an ambiguously coded disposition, which can indicate that the case was not resolved decisively (e.g., because it was transferred to another court), or else that the eCourts analyst did not write a clear classification even if one was available. An example of the latter is the disposition label "Disposed", which indicates that the case is closed but does not provide information on the direction of the ruling.[17]

[17]When we have looked manually at cases with ambiguous codings, we have found that it is often possible to determine what happened, but doing so manually is not feasible for over a million

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

In our main analysis sample, 58% of cases have decisions; of those, 40% can be unambiguously coded as acquittal or conviction. Our main specification includes all cases — including those without decisions and those with ambiguous codings. Missingness and decision ambiguity will affect our results only if assignment to an in-group judge affects the rate or the subset of cases that are ambiguously coded. Given that the coding appears to be a choice made by the eCourts data entry analysts, this is unlikely to take place. And indeed, we formally test and show that ambiguity and case closure are not affected by assignment to an in-group judge (Appendix Tables A5, A6). Further, we show that our results are robust to alternative specifications along these margins, including (i) coding ambiguous or undecided outcomes as negative rather than positive; (ii) dropping either the ambiguous or the undecided cases from the data and focusing on cases with clear outcomes; or (iii) focusing on courts or charges with lower ambiguity rates (Appendix Tables A7, A8, A9, A10).

Judicial delay is itself a major policy issue in India. Hence, we provide additional results where getting a decision at all is the outcome of interest. For these regressions, we define an outcome indicator for whether a decision is made on a case within six months of the case's filing date, which includes about 30% of cases.

3.5 Summary Statistics

Figure 1 presents descriptive statistics of charges and convictions by gender and religious identity of defendants, respectively.[18] These summary measures are descriptive in nature, but are not directly informative of bias in the judicial system because we do not know the share of defendants who commit crimes or who are guilty when charged.

Figure 1 Panel A shows that the share of women charged under all crime categories is substantially lower than their population share: men are three to six times more likely to be charged with crimes under any classification. Panel B shows that the acquittal rate varies by crime, but overall it is about 1 percentage point higher for women (the "Total" category, at the bottom).

such cases.

[18]The corresponding point estimates are reported in Appendix Tables A11 and A12.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Panel C shows that Muslims are over-represented by 4% in the universe of criminal charges. Representation changes substantially depending on the charge: relative to their population share, Muslims are 34% more likely to be charged with other crimes against women, 23% more likely to be charged with robbery, and 62% more likely to be charged with marriage offenses, but 4% less likely to face charges for murder. Panel D shows that aggregate differences in acquittal rates between Muslims and non-Muslims are negligible.

Table 1 shows descriptive statistics of judges and case outcomes in the analysis sample. About 27% of judges are female and 7% of judges are Muslim. On average, Muslim and female judges have similar acquittal, conviction and rapid decision rates to non-Muslim and male judges. Although there are similar averages across groups, we still observe significant variation across judges after adjusting for court-time and charge fixed effects.[19] The variation in acquittal rates for comparable case portfolios reflects the extensive discretion exercised by judges in this legal context.

Appendix Tables A13 and A14 show the representativeness of our analysis sample (and subsamples used later in the paper) across state and crime categories, relative to the complete dataset of 23 million crime records. With a few exceptions, our analysis datasets are representative of the universe of criminal cases in India.

[19] Appendix Figure A5 shows the substantial variation in the distribution of judge fixed effects for the acquittal rate after residualizing out court-time and charge fixed effects. On a base acquittal rate of 68% (given an unambiguous decision), the IQR of the mean judge fixed effect for acquittal is [-8.0, +10.0] percentage points. Conditional on location-month and charge fixed effects, judge identity explains 10% of the residual variation in the acquittal outcome. This is of course a lower bound on the level of discretion available to a judge, since judges do not always express discretion in the same direction.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

4 Empirical Strategy

4.1 Random Assignment of Judges to Cases

To estimate a causal effect of judge identity, we need to effectively control for any factors other than defendant identity that could affect both judge identity and the case outcome. For example, if Muslim judges could systematically choose to sit in cases with Muslim defendants who had committed less serious crimes, we might mistakenly infer in-group bias even in its absence. Alternately, Muslim defendants and judges are more likely to appear in regions of the country with more Muslims. If those regions are characterized by different crime distributions (with different acquittal rates), we might again mistakenly attribute those differences to in-group bias.

As with much of the prior empirical literature, these concerns are resolved in our context through the random assignment of judges to cases (Section 2). For ease of exposition, we describe the empirical strategy investigating gender bias — the specification and considerations for estimating religious identity bias are identical. Specifications used in subsequent analysis are described with the results.[20]

Our ideal experiment would take two defendants identical in all ways, charged with identical crimes in the same police station on the same date, and then assign them to judges with different identities. In practice, the Indian court system runs this experiment whenever a defendant is charged in a jurisdiction with multiple judges of different identities on the bench. Even if there is bias at other stages of the criminal process (e.g. in who gets charged), that would not undermine our identification strategy given the random assignment of judges.

We use a canonical regression approach to test for the effect of judge identity on case outcomes, as used by Shayo and Zussman's (2011) analysis of judicial in-group bias in Israel. We model

[20]We also explored an event study specification exploiting case timing and changes in the cohort of judges sitting in each court, but we found that recently changed courts are more likely to see younger cases, violating the assumptions required for the event study analysis.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

outcome Y_i (e.g. 1=acquitted) for case i with charge s, filed in court c at time t as:

$$Y_i = \beta_1 \text{judgeMale}_i + \beta_2 \text{defMale}_i + \beta_3 \text{judgeMale}_i * \text{defMale}_i + \phi_{ct(i)} + \zeta_{s(i)} + X_i\delta + \varepsilon_i \quad (1)$$

where judgeMale and defMale are binary variables that indicate the gender of the judge and defendant respectively. The omitted category is the set of cases with female judges and female defendants. $\phi_{ct(i)}$ is a court-month or court-year fixed effect (based on the case filing date), and $\zeta_{s(i)}$ is an act and section fixed effect. X_i includes controls for defendant religion, judge religion, and an interaction term of judge gender and defendant religion. The analysis of religious in-group bias follows the same structure, where the omitted category is the set of Muslim judges and defendants, and X_i represents controls for defendant gender, judge gender, and an interaction term of judge religion and defendant gender.

The charge-section fixed effect ensures that we are comparing defendants charged with similar crimes. The court-time fixed effect ensures that we are comparing defendants who are being charged in the same court at the same time. Our primary specification uses a court-month fixed effect, while a secondary specification uses a court-year fixed effect. The court-year fixed effect allows a larger sample, at some potential bias. Judges on the bench may not hear new cases in some months because they are tied up with previous cases or away from work. It is unlikely that prosecutors or defendants can time their filings to match these absences, nor do we find evidence of disproportionate identity matching in balance tests of either specification (see Appendix B.2). Court-time periods with no variation in judge identity are retained to increase the precision of fixed effects and controls, but they do not directly affect the coefficients of interest. We also test a specification with judge fixed effects, which controls for the average acquittal behavior of each individual judge.[21] Standard errors are clustered at the judge level, since judge assignment is the level of randomization.

[21]This specification is included for completeness, but is unnecessary for identification if judges are indeed randomly assigned.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

There are three causal effects of interest. β_1 describes the causal effect on a female defendant of having a male judge assigned to her case rather than a female judge. $\beta_1 + \beta_3$ describes the causal effect on a *male* defendant of having a male judge assigned to his case. The difference between these effects (β_3) is the own-gender bias — it tells us whether individuals receive better outcomes when a judge matching their gender identity is randomly assigned to their case. Since all three causal effects are of interest, we report all three for each estimation.

About half the time, a case stays in the courts long enough such that the judge making the final decision is different from the one to whom the case was initially (randomly) assigned. For these decisions, we continue to use the identity characteristics of the initially assigned judge. We do not exclude these cases in our primary specification because a rapid decision is itself an outcome. Even if the filing judge does not make the final ruling on a case, they can make key decisions on the case process that influence the decision, such as allowing witnesses, admitting evidence, and determining the schedule on which the case is resolved. Either way, this choice does not drive our results, as we estimate virtually identical effects if we limit the sample to cases decided by the initially assigned judge.

A more subtle identification issue arises in describing these matching-gender and matching-religion effects as capturing "in-group bias." Our interpretation follows the prior empirical literature, where "in-group bias" describes a situation where defendants receive better outcomes when their identity matches the (exogenously assigned) judge's identity. A limitation of this approach, highlighted by Frandsen et al. (2023) and Canay et al. (2023), is that defendants from different identity groups share more characteristics than just their identity, many of which are unobserved. Further, judges from different identity groups might have correlated preferences or biases across those characteristics. For example, suppose that female defendants were more likely to have children, and that female judges were on average more lenient for defendants with children. Our empirical approach would identify in-group bias on the gender dimension. Disentangling these aspects of identity is challenging and admittedly beyond the scope of this paper. However, documenting the contextual variation in where identity matters for outcomes is a valuable first step in addressing

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

these issues. Further, our estimates are informative of the expected impacts of making India's judge body more representative, even if any "bias" found is not driven by identity alone.[22]

5 Results

5.1 Effect of assignment to judge types

The first two rows of Table 2 Panel A present the impact, for female and male defendants respectively, of being randomly assigned to a male judge — these are β_1 and $\beta_1 + \beta_3$ in Equation 1. The third row shows the difference between these two coefficients (β_3), which is the own-gender bias. The outcome variable is an indicator for defendant acquittal. Columns 1–3 show results using court-month fixed effects, while Columns 4–6 use court-year fixed effects. Within each set of three columns, the second column adds additional demographic controls, while the third column adds judge fixed effects.

Male judges acquit at a similar rate to female judges, regardless of defendant gender. The own-gender bias estimate is a tight zero; the effect estimates rule out even a very small in-group bias effect of 0.5 percentage points with 95% confidence.[23] The coefficients are stable across different fixed effect specifications, as is expected given the as-good-as-random assignment of judges to

[22] Another issue is "inframarginality bias", discussed in detail for example by Canay et al. (2023). In brief, if judges are deciding on conviction based on some threshold (say, probability of guilt), then group bias means different thresholds across groups. Then the average group differences captured by our regression estimates are a combination of both differences in judge standards for the marginal cases around the threshold, as well as distributional differences away from the threshold. These issues are much reduced in the context of in-group bias (rather than overall bias), where such confounds would have to occur at the level of judge-defendant interactions. Further, our analysis is implicitly assuming that the inclusion of covariates work to match the defendant risk distributions at the judge-defendant-type level. Finally, again, even in the presence of such bias, our estimates are still informative about what would happen by making the judiciary more representative.

[23] Appendix Table A7 shows bias effects on conviction rates (rather than acquittal rates); the

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

defendants.

Table 2 Panel B shows the effect of filing-judge gender on an indicator for case resolution within six months of being filed. In the specifications without judge fixed effects, we find that cases where judge and defendant match gender are at most 0.5 percentage points more likely to be resolved within six months (on a mean of 28%). The effect is very small in magnitude, and is not robust to the inclusion of judge fixed effects, under which the effect size falls to a statistically insignificant 0.2 percentage points.

Table 3 presents analogous results for Muslim and non-Muslim defendants randomly assigned to Muslim and non-Muslim judges; all panels and columns have the same interpretation as the prior table. In some specifications, we find that non-Muslim judges are less likely to acquit; but this holds equally for Muslim and non-Muslim defendants. The point estimate on in-group bias for acquittals is at most 0.3 percentage points and the estimates rule out an own-religion bias of 0.75 percentage points with 95% confidence.[24] Religious in-group bias is also absent in the speed of judicial decisions, nor is there any evidence that Muslim and non-Muslim judges have different rates of resolving cases (Table 3, Panel B).

Given the important role that lawyers play in the judicial process, in-group bias could also

estimates again are a tight zero. Appendix Table A8 shows estimates when we exclude closed cases for which we are unable to determine the outcome. We prefer the specification in Table 2, because the inability to determine an outcome is itself an outcome. We also find no effect of gender or religious match on whether the outcome is clearly coded as acquittal or conviction (Appendix Table A5). Finally, we show that results are identical when we limit the sample to either courts or charges with below-median rates of the outcome being coded as ambiguous (Appendix Table A18).

[24] Appendix Tables A9 and A10 show results on conviction rates, and on acquittals with ambiguous results dropped. While we find marginally significant bias effects (in the in-group direction) in a handful of specifications, the majority are statistically insignificant, and the point estimate on the bias term is never higher than 0.6 percentage points. Appendix Table A6 shows there is no effect of in-group bias on an indicator for an ambiguous case outcome.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

depend on the identity of the lawyer.[25] The names of lawyers are missing in over 90% of cases, but we are still powered to run a bias test for the set of cases where either lawyer is observed.[26] Appendix Tables A22 and A23 use a similar specification to the judge-defendant models above, and show that there is no evidence of differential outcomes when the judge matches the gender or religious identity of either of the litigant's lawyers.

5.2 Judicial Bias when Identity is Salient

Our estimates thus far show that judges do not provide substantively better outcomes for own-gender and own-religion defendants, on average. This is contrary to most of the work in this space, which finds bias against out-group defendants regardless of the group of the victim or plaintiff, for example (Grossman et al., 2016), (Anwar et al., 2012), (Gazal-Ayal & Sulitzeanu-Kenan, 2010), (Sloane, 2019), and (Didwania, 2022).

Some of the prior literature suggests that various identities can be made more salient by specific contexts or primes. This section examines several circumstances where gender or religious identity may become particularly salient to judges. In each circumstance, we test for additional bias by defining an indicator variable that takes the value one in a condition that activates bias. We interact this variable with every right-hand side variable in Equation 1. If bias is particularly activated in this context, the interaction with the in-group bias term will be positive and significant.

We first examine the subset of cases where the victim and defendant have different identities. In these cases, when the defendant and judge are mismatched, the judge and victim will share the

[25] For example, Marx et al. (2019) find that ethnic patronage in rental contracts depends more on identity match between the slum chief and landlord (where the latter plays a role of arbiter) than on match between the landlord and tenant.

[26] Appendix Table A19 presents a balance test for random judge assignment in the subsample where we observe both defending and petitioning lawyers' gender and religious identities. As in the main sample, the evidence is consistent with random assignment.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

same gender or religious identity.[27] The identity match or mismatch between judge and defendant may be particularly salient in this case (Baldus et al., 1997; ForsterLee et al., 2006; Baumgartner et al., 2015). Column 1 of Table 4 interacts an indicator for defendant-victim gender mismatch with the gender in-group bias indicator. Both the baseline bias effect and the interacted effect are null; judges do not show gender in-group bias even when the defendant and victim have different genders (only one of which is matched by the judge). Similarly, Column 2 shows that there is no additional in-group religion bias when defendant and victim have different religions.[28] Standard errors are larger due to the smaller sample and interaction specification, but the in-group bias effect is less than 1 percentage point in both cases.

We next look at whether male and female judges rule differently on cases classified in the criminal code as crimes against women, where judge and defendant gender identities may be particularly salient. These are about evenly split between sexual assaults and kidnappings.[29] Column 3 of Table 4 shows that the interaction between an indicator for crimes against women and the in-group bias variable is small and statistically insignificant. Male defendants do not receive differential treatment from male and female judges, even in these cases.

Finally, in Table 4 Column 4 we examine whether religious in-group bias emerges during the month of Ramadan, when Muslim religious identity may become particularly salient for both Mus-

[27]In the case of religion, 6% of Indians are neither Muslim nor Hindu, so two non-Muslim individuals are highly likely to be in the same broad religious group but in some cases will not be.

[28]Note that for legibility, the table only lists the in-group bias term and its interaction with the context variable, but all the terms in Equation 1 are interacted with the context variable, as are the fixed effects. Appendix Tables A20 and A21 show all of the coefficients from the regression with court-month fixed effects. Samples are smaller than in the main bias estimation because the identity of the victim can be determined (from the name) in only about half of cases.

[29]One reason "kidnappings" are so common in the data is that this may be the formal charge filed against a man who elopes with a woman. Results are similar if we restrict the interaction term to cover sexual assaults (Appendix Table A24).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

lims and non-Muslims.[30] The interaction between the Ramadan indicator and the in-group bias measure is small and statistically insignificant. Note the difference from (Mehmood et al., 2023), who do find effects of judge religious identity on decisions during Ramadan. Appendix Table A25 shows robustness of the estimates to using court-year instead of court-month fixed effects.[31]

We investigated religious violence, but we could not find datasets on religious violence that were well-matched to our sample period.[32] We show in Appendix Table A27 that there is no differential in-group bias effect in election months. However, there is substantial variation in the extent to which in-group sentiments are primed across elections, so further research here is warranted.

[30]Unlike the sample in Mehmood et al. (2023), our sample only covers eight years, with Ramadan occurring only in the summer. There is thus no substantial time-series variation in daylight hours that can be exploited. Note that for this table only, we use the identity of the judge *deciding* on the case, rather than the judge to whom it was assigned initially. Our implicit assumption is that the effect of Ramadan affects the outcome on the day the decision is reached, rather than on the day the case first appeared before a judge. See Section 4 for more on how we treat cases seen by more than one judge.

[31]Religion is often a salient aspect of elections in India. A rigorous analysis of how elections affect judicial decision-making is beyond the scope of this paper. Our results are unchanged if we limit the sample to 2015–2018 (the post-national BJP period), and are similarly null if we partition the sample into 2-year bins (Appendix Table A26).

[32]The primary dataset used for the study of religious violence is Varshney and Wilkinson (2006), which was updated by Bhalotra et al. (2012) to 2010, the first year in our sample. The ACLED violence database covers India only from 2016; descriptions are in many cases too vague to identify Hindu-Muslim violence specifically or their perpetrators. While we did not find evidence of differential in-group bias in the week or month following local violence, the data quality is too low to treat the analysis as dispositive.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

5.3 In-group Bias on the Basis of Caste

We now consider one of the most important social cleavages in India: caste. Ideally, we would run an equivalent statistical test, where judge and defendant identity sometimes match on the caste dimension and sometimes do not. This is unfortunately infeasible for three reasons. First, unlike gender and religion, there is no classification for caste along which in- and out-groups can be confidently and universally defined. The two major categories of caste, *varna* (four broad hierarchical categories, although hundreds of millions of Indians are *avarna*, or having no *varna*) and *jati* (approximately 5,000 endogamous communities), are both insufficient in characterizing the affinities that people may feel within the caste system. For example, an upper caste person could identify with another upper caste person despite sharing neither *varna* or *jati*. Second, individual names do not identify caste as precisely as they identify Islamic religion or gender identity and the caste significance of names can vary across regions.[33] Due to these limitations and to a lack of training data, we have not been able to develop a reliable correspondence between names and specific castes. Third, there are few district judges in the most identifiable caste categories: Scheduled Castes and Scheduled Tribes.

For these reasons, a direct analysis of in-group caste bias in the Indian judiciary is challenging and imperfect.[34]. Instead, we analyze caste indirectly. Specifically, we follow Fisman et al. (2017) and define individuals as being in the same cultural group if they share a last name. As discussed in that paper and other work, shared last names are a noisy measure of caste similarity for many social groups.

[33]For example, Vahini et al. (2022) find 97% accuracy in predicting religion from Indian names, but only 73% even for broad caste categories like OBC/SC/ST. We have encountered similarly low accuracy in our own exercises. Given the need to match both judge and defendant, over half of our sample would be mis-classified at these rates.

[34]In Appendix B.3, we describe an attempt to analyze in-group bias on the basis of *varna* using a name-*varna* correspondence from the *People of India* project. We found that the data was too noisy and imperfect to be informative.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

The measure is admittedly imperfect. Names are more numerous than castes, so members of the same caste often have different last names. Further, sharing names can indicate greater affinity and closer social proximity than caste. Last names could signal similar socioeconomic status, for example, or shared religion. When a judge and defendant share a last name, they could even be relatives by blood or marriage. Individuals can also share a last name and be in different castes.

To determine whether judges deliver more favorable outcomes to defendants who share their last name, we estimate:

$$Y_i = \beta_1 \text{sameLastName}_i + \phi_{ct(i)} + \zeta_{s(i)} + X_i\delta + \varepsilon_i. \tag{2}$$

where subscripts i, s,c, and t, the court-time ($\phi_{ct(i)}$) and act/section ($\zeta_s(i)$) fixed effects, and the judge/defendant characteristics $X_i\delta$ are all as above. We include additional fixed effects for judge and defendant last names and control for judge and defendant gender and religion. We limit the sample to individuals with last names that match at least one judge in their district at any time.[35]

As above, the act/section fixed effects adjust for judge assignment rules based on the seriousness of the crime. The last name fixed effects adjust for the possibility that individuals from some social groups are more or less likely to be acquitted, and that judges in different social groups may have different average acquittal rates. The identification assumptions for consistent estimation of $\hat{\beta}_1$ are the same as in the prior section, and depend on random assignment of defendants to judges with any given last name within the court-time randomization block.[36]

[35]Without this limitation we have substantially more last name fixed effects in the sample but there is no additional variation in terms of identity match, because the *sameLastName* variable always takes the value 0 for defendants whose last name never appears in the judge list.

[36]For a given last name, we test for balance by regressing an indicator for the judge having that last name on an indicator for a defendant having that last name, with the usual court-month and section fixed effects. We run this test for every last name that has at least one judge-defendant match. Appendix Figure A6 shows that the distribution of coefficients is concentrated around zero

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

The results for last name bias are reported in Table 5. Columns 1 and 2 report unweighted estimates from Equation 2, comparable to the specifications in the previous sections. The point estimate of in-group bias is a precisely estimated zero.

An issue with the unweighted case-level regressions is that the sample is dominated by social groups with common last names, like *Kumar* and *Singh*. These are the names where a defendant-judge name match is the least likely to indicate shared caste. Matching on a common name may not indicate much cultural similarity, and the resulting estimates may not capture the experience of smaller caste groups. To address this issue, we estimate an alternate specification where sample weights treat each defendant last name group equally. Formally, we estimate weighted regressions where the weights are computed as the inverse of the number of defendants in the sample with each given last name. These regressions therefore describe variation in bias across *groups*, rather than across individuals.

The weighted regressions are reported in Columns 3 and 4, corresponding to the respective unweighted regressions in Columns 1 and 2. The weighted regressions show that a judge-defendant name match increases the likelihood of acquittal by about one percentage point ($p = 0.11$ in Column 3). To directly test whether bias is driven by groups with less common names, we add a "rare name" interaction with the last name match indicator, where the "rare name" variable takes the value one if the defendant has a name with a below-median count in the data.[37] Columns 5 and 6 report this specification. The uninteracted coefficient shows an absence of bias for common last names, and the interacted coefficient shows a 2 percentage point in-group bias for individuals with uncommon last names. The effect is not statistically significant in this specification ($p = 0.14$), but the effects

and that nearly all confidence intervals include zero.

[37]Results are similar whether we use the median across individuals or the median across groups. Out of 2,761,382 defendants with last names that appear at least once in the judge sample, 112,934 have rare names based on the individual median, and 1,376,640 have rare names based on the group median. These effects are robust to looser definitions of last name similarity (for example, treating *Patil* and *Patel* as similar).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

in Columns 3–6 are all statistically significant at the 5% level in specifications with court-year fixed effects (Appendix Table A28).[38]

The effect size among individuals with uncommon names is economically relevant, representing about a 10–20% increase in the probability of acquittal. Yet this bias is only seen for the relatively small number of people with less common names. By definition, then, the same-name effect is relevant only for a small share of the population. Groups with rare names are mechanically underrepresented in the population, and the likelihood of matching a judge with the same rare name is even smaller. This bias, therefore, while large in magnitude for some individuals, will be small in aggregate if it operates only at the level of small social groups.[39]

6 Discussion and Conclusion

Courts in developing countries face a number of special challenges, including cultural mismatch from transplanted legal codes, informal justice-system substitutes, citizen skepticism toward formal courts, insufficient human and physical capital investments in the court system, the inability of many individuals to pay for high-quality representation, implicit or explicit bias among members of the judiciary, and corruption (Djankov et al., 2003; La Porta et al., 2008). Yet with a few exceptions (Ponticelli & Alencar, 2016, for example), these characteristics of developing-country courts have been described only anecdotally.

We make progress in this area by analyzing decisions in over 5 million criminal cases in India.

[38]The effect sizes in the court-year specification are slightly larger, but not statistically distinguishable from those in the court-month specification.

[39]Appendix Figure A7 shows how the interaction regression varies as a function of the threshold used to define rare names. The unweighted panel shows that the interaction terms become substantial and significant only for names outside of the 200 most common. About 10% of defendants have names in this category, of whom about 1% get assigned to judges with the same name—representing 2670 cases out of a sample of 2.6 million. Appendix Table A29 shows the list of most common last names in the data.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

We estimate robust, tight zero effects of judicial in-group bias along the dimensions of gender, religion, and caste. We do not find gender- or religion-based bias even in several contexts where these identities may be particularly salient. We do find in-group bias among social groups with shared uncommon last names, suggesting that in-group bias may be magnified when the shared identity group is very small. The aggregate effects of this bias are small, but they may be large for individual defendants.

The systematic null effects are surprising, especially given well-documented gender, caste, and religious in-group bias in non-judicial contexts in India. Two relevant examples are Fisman et al. (2017), who find that credit offers and repayment rates rise when loan officers and clients have the same last name, and Neggers (2018), who finds that random assignment of a minority election worker to a polling station has a large pro-minority effect on vote counts at that station. Our divergent findings raise the question of how these contexts differ from the judicial setting.

One major difference is the judge's incentive structure. Judges expect little direct economic benefit or cost from seeing members of the out-group punished. That "game" is quite different from the cooperative context in Fisman et al. (2017) (where joint gains are possible through a successful loan), or the adversarial context in Neggers (2018) (where only one party can win an election).

A second relevant feature is the competing relevance of other identity factors. The judicial setting may make salient the class, education, or other status differences between judges and defendants, crowding out broader identity characteristics like religion and gender. In contrast, political competition for resources (as in Neggers (2018)) may magnify the salience of these identities. Similarly, Kumar and Sharan (2023) finds that ethnic quotas in local government only improve public service delivery when lower-status groups occupy multiple positions in the political hierarchy. Consistent with this interpretation, our results on matching last names suggest that in-group bias is stronger under more narrow definitions of the in-group.

An example of both of these dynamics outside of judging is Hanna and Linden (2012), who find no evidence of out-group animus (on the caste dimension) in the case of teachers grading student

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

exams. Like judging, grading is a non-adversarial context, where teachers face flat incentives for how students are assessed. Further, there are impactful class and authority differences between teachers and students, which may make differences due to caste less salient.

This discussion highlights the sensitivity of in-group bias to context. Further, it hints at a theoretical grounding for why results on in-group bias vary across different settings. Further empirical research drilling down on these theories will be valuable.

6.1 Considering Publication Bias

In the judicial setting, our null estimates of in-group bias contrast with findings in other jurisdictions, where researchers have tended to find large effects. To compare our estimates to those in the literature, we collect coefficients and standard errors from the studies of judge in-group bias that are most similar to ours. We identify every study we can find that focuses on measuring in-group bias among judges on a race, ethnicity, gender, or religious dimension, and that exploits an as-good-as-random judge or jury assignment mechanism for causal identification.[40] To make the studies comparable, we standardize effect sizes by dividing each in-group bias effect by the sample standard deviation of the outcome variable. As shown in Figure 2 Panel A, our primary effect sizes on religion and gender are the smallest in the literature. The high end of our confidence interval is an order of magnitude smaller than nearly all prior studies.

Another notable pattern in the graph is that the confidence intervals (and hence standard errors) grow with the effect sizes. A positive relationship between effect size and standard errors suggests that there could be publication bias in studies of judicial in-group bias, which would also help explain the distinctiveness of our null finding. To show this more directly, Figure 2 Panel B plots

[40] When papers report multiple specifications for the main effect, we use the effect size described most prominently in the text or described by the authors as the "main specification." When papers have multiple outcomes, we use the outcome most similar to the acquittal or conviction rate, as in this study. If these are unavailable, we use the outcome most prominently described in the paper's abstract and introduction.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

(in black triangles) the effect size of each of the previous studies against the standard error of the main estimated effect. For comparison, the estimates from our study are plotted as red circles. In the absence of publication bias or a design-based mechanical relationship between effect size and precision (such as adaptive sampling), study estimates should form a funnel that is centered around the true estimate.[41] The graphed estimates are evidently asymmetric, with many of the studies falling just outside the boundary defining statistical significance at the 5% level.

To formally test for publication bias in prior studies, we follow the approach of Andrews and Kasy (2019). We estimate a publication function $p(z)$, describing the probability that a study is published as a function of the t-statistic z, the effect size divided by the effect standard error. This function can be identified up to a scale parameter, which we normalize under the assumption that all studies with $z > 1.96$ are published. This estimated function gives us a structural estimate, based on the existing published papers, for the likelihood of publication given a t-statistic z. The method also provides an adjusted effect size based on imputing unpublished studies.

Note that this method does not require all of these studies to estimate the same parameter — this is essential, since the true amount of bias may differ substantially across settings, as we argue above. It only requires the approximate normality of parameter estimates, which is already implicit in the standard error calculations in most of these studies. Each study can be thought of as drawing a single parameter estimate, which is a noisy estimate of the true value in that particular setting. In the absence of publication bias, the expected value of the parameter estimate should not depend on the sample size; this is the null hypothesis of the Andrews and Kasy (2019) test.

Table 6 reports the result of the test for publication bias. Under the assumption that all positive, statistically significant studies are published, Columns 1–3 respectively show the probability that a study will get published, given a t-statistic in the ranges of $(-\infty, -1.96)$, $(-1.96, 0)$, $(0, 1.96)$,

[41] See Egger et al., 1997; Gerber et al., 2001; Levine et al., 2009; Slavin and Smith, 2009; Kühberger et al., 2014; Andrews and Kasy, 2019. A funnel shape is expected because studies with larger standard errors should produce a wider range of estimates that are symmetric around the true value.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

respectively. The estimates imply that studies with negative or statistically insignificant estimates are extremely unlikely to be published. Studies with results like ours — statistically insignificant positive estimates — are *only 3% as likely to be published* as studies with statistically significant positive results.

The estimates from prior literature are consistent with severe publication bias. The Column 5 estimate tells us that if all of these studies were estimating the same parameter, then accounting for publication bias would give us a true effect size of 0.046 SD, a fraction of the average observed effect size of 0.24 SD from the published studies. However, if these studies are all estimating different parameters due to their different contexts, then each of these estimates could be correct; instead, the publication bias exercise would suggest that there is a large volume of potential studies in contexts with no in-group bias which have not made it to the publication phase.[42]

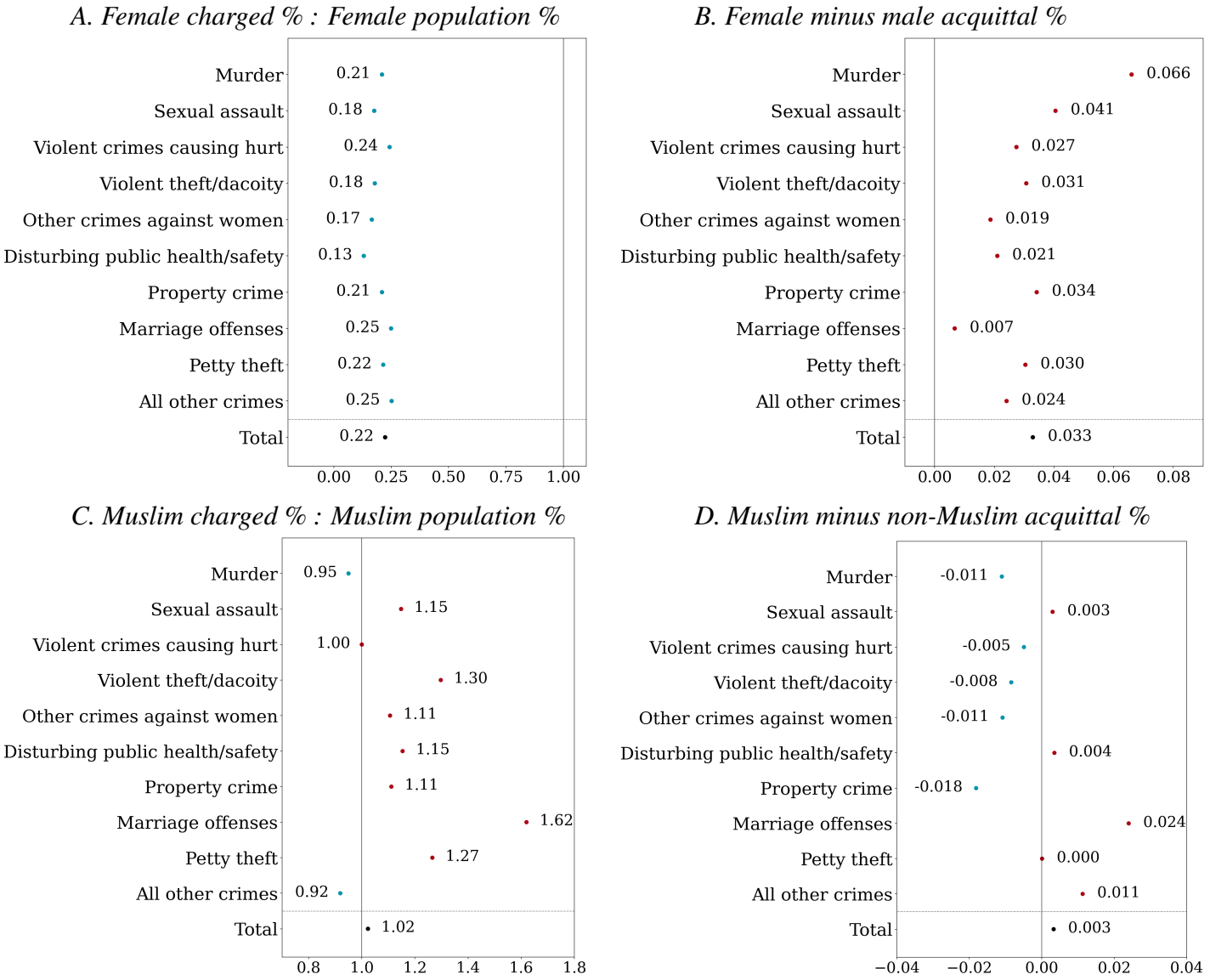
The rest of the literature aside, our finding of a lack of in-group bias in India's lower courts does not rule out bias in the criminal justice system as a whole. Notwithstanding our results on acquittals, the legal system could still be biased against marginalized groups due to unequal geographic distribution of policing, discrimination in investigations, police/prosecutor decisions to file cases, the severity of charges applied, the severity of penalties imposed, the appeals process, civil litigation, or other factors. There could also be absolute bias, where both in- and out-group judges discriminate against out-groups. Our evidence suggests concerns about in-group bias might be better directed to parts of the justice pipeline other than judge acquittal decisions.

More research is sorely needed to create an empirical basis for understanding the judicial process in India and in other developing countries. The expansion of publicly available datasets on judicial systems worldwide will be an important step in making this possible.

[42]Indeed, since posting this paper, we have heard from more than one researcher who abandoned research on in-group bias when their preliminary results suggested a null result. Null or reverse effects of in-group bias do appear in other studies, but these studies focus on different aspects of their contexts and put little emphasis on the null in-group effects (Arnold et al., 2018; Hanna & Linden, 2012).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Figure 1: Summary statistics by crime category and defendant identity



Notes: Panel A shows the imbalance in the per capita rate of criminal charges by gender. The share of cases with female defendants is divided by the share of women in the Indian population for each type of criminal charge. Panel C shows the same result for Muslims. Panel B shows the difference between female and male acquittal rates for each type of crime. Panel D shows the same difference between Muslims and non-Muslims. Crimes are ordered by maximal punishment, from most to least severe.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Table 1: Summary statistics, by judge identity

		Judge gender		Judge religion	
	(1)	(2)	(3)	(4)	(5)
	Total	Female	Male	Muslim	Non-Muslim
Female judge	0.2785	—	0.0000	0.2615	0.2766
	(0.0024)		(0.0000)	(0.0087)	(0.0025)
Muslim judge	0.0697	0.0673	0.0723	—	0.0000
	(0.0013)	(0.0025)	(0.0016)		(0.0000)
Tenure length (Days)	489.8893	490.3807	496.9476	478.4971	491.6112
	(2.2786)	(4.3576)	(2.7237)	(8.8109)	(2.3693)
Decisions					
Decision within 6 months	0.2503	0.2504	0.2465	0.2527	0.2507
	(0.0015)	(0.0030)	(0.0018)	(0.0058)	(0.0016)
Acquitted	0.1920	0.1924	0.1911	0.1972	0.1915
	(0.0012)	(0.0024)	(0.0014)	(0.0047)	(0.0012)
Convicted	0.0352	0.0379	0.0338	0.0369	0.0349
	(0.0005)	(0.0010)	(0.0006)	(0.0020)	(0.0005)
N	37,786	10,091	26,145	2,595	34,632

Notes: Coefficients represent means for each variable in the sample, collapsed to the judge level. Standard errors reported in parentheses.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Table 2: Impact of assignment to a male judge on defendant outcomes

(A) Outcome variable: Acquittal rate						
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	0.0047	0.0037	—	0.0001	-0.0008	—
	(0.0050)	(0.0053)		(0.0027)	(0.0031)	
Male judge on male defendant	0.0059	0.0049	—	0.0011	0.0002	—
	(0.0048)	(0.0050)		(0.0022)	(0.0026)	
Difference = Own gender bias	0.0012	0.0011	0.0002	0.0010	0.0010	-0.0002
	(0.0016)	(0.0016)	(0.0016)	(0.0017)	(0.0017)	(0.0016)
Reference group mean	0.1751	0.1761	0.1761	0.176	0.1771	0.177
Observations	5188580	5094774	5093595	5233366	5139820	5137855
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

(B) Outcome variable: Decision within six months of filing						
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	-0.0066	-0.0041	—	0.0022	0.0034	—
	(0.0101)	(0.0105)		(0.0052)	(0.0060)	
Male judge on male defendant	-0.0112	-0.0083	—	-0.0023	-0.0007	—
	(0.0103)	(0.0107)		(0.0053)	(0.0060)	
Difference = Own gender bias	-0.0045**	-0.0042*	-0.0023	-0.0045**	-0.0040*	-0.0025
	(0.0022)	(0.0021)	(0.0020)	(0.0022)	(0.0022)	(0.0020)
Reference group mean	0.2834	0.2827	0.2828	0.2828	0.2822	0.2822
Observations	4335218	4254502	4253469	4376949	4296455	4294720
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Reference group: Female judges.
Charge section fixed effects have been used across all columns reported.
Specification: $Y_i = \beta_1 \text{judgeMale}_i + \beta_2 \text{defMale}_i + \beta_3 \text{judgeMale}_i * \text{defMale}_i + \phi_{ct(i)} + \zeta_{s(i)} + X_i\delta + \varepsilon_i$

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Table 3: Impact of assignment to a non-Muslim judge on defendant outcomes

(A) Outcome variable: Acquittal rate						
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	-0.0174**	-0.0180**	—	-0.0074*	-0.0078	—
	(0.0086)	(0.0089)		(0.0044)	(0.0052)	
Non-Muslim judge on non-Muslim defendant	-0.0161*	-0.0164*	—	-0.0048	-0.0046	—
	(0.0083)	(0.0086)		(0.0038)	(0.0045)	
Difference = Own religion bias	0.0013	0.0016	0.0019	0.0026	0.0032	0.0026
	(0.0022)	(0.0023)	(0.0021)	(0.0022)	(0.0023)	(0.0021)
Reference group mean	0.1799	0.1831	0.1831	0.1807	0.1839	0.184
Observations	5611751	5178858	5177603	5656115	5224554	5222471
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

(B) Outcome variable: Decision within six months of filing						
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	-0.0182	-0.0201	—	-0.0122	-0.0163*	—
	(0.0174)	(0.0170)		(0.0086)	(0.0090)	
Non-Muslim judge on non-Muslim defendant	-0.0152	-0.0174	—	-0.0087	-0.0130	—
	(0.0172)	(0.0168)		(0.0082)	(0.0086)	
Difference = Own religion bias	0.0030	0.0027	0.0017	0.0035	0.0033	0.0033
	(0.0029)	(0.0030)	(0.0028)	(0.0031)	(0.0031)	(0.0028)
Reference group mean	0.2912	0.2876	0.2877	0.2905	0.287	0.287
Observations	4692802	4327596	4326473	4734172	4370169	4368314
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Reference group: Muslim judges.
Charge section fixed effects have been used across all columns reported.
Specification: $Y_i = \beta_1 \text{judgeNonMuslim}_i + \beta_2 \text{defNonMuslim}_i + \beta_3 \text{judgeNonMuslim}_i * \text{defNonMuslim}_i + \phi_{ct(i)} + \zeta_{s(i)} + X_i\delta + \varepsilon_i$

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Table 4: In-group bias in contexts that activate identity

	(1) Gender	(2) Religion	(3) Gender	(4) Religion	(5) Religion
Ingroup Bias	0.0045	0.0005	0.0002	0.0002	0.0020
	(0.0036)	(0.0048)	(0.0016)	(0.0029)	(0.0033)
Ingroup Bias * Gender mismatch	-0.0059				
	(0.0053)				
Non-Muslim judge and defendant * Mismatch		0.0086			
		(0.0080)			
Ingroup Bias * Crime against women			-0.0090		
			(0.0118)		
Ingroup Bias * Ramadan				0.0013	
				(0.0102)	
Ingroup Bias * Hindu Festival					-0.0078
					(0.0079)
Observations	1748328	1970008	5089229	3052192	3052192
Fixed Effect	Court-month	Court-month	Court-month	Court-month	Court-month
Judge Fixed Effect	Yes	Yes	Yes	Yes	Yes
Sample	All	All	All	All	All

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Regression results on whether in-group bias appears in a set of contexts that may make identity particularly salient. The contexts tested in each column are: (1) the defendant and victim have different religions; (2) the defendant and victim have different genders; (3) the case includes one or more charges considered crimes against women; (4) the judgment takes place during the month of Ramadan; and (5) the judgment takes place on the day of a Hindu festival (Dasara, Diwali, Holi or Rama Navami) or within the six following days. The type of bias considered is based on gender in Columns 1 and 3, and on religion in Columns 2, 4, and 5. Charge section fixed effects have been used across all reported columns.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Table 5: Effect of assignment to judge with same last name on defendant outcomes

	(1) Acquitted	(2) Acquitted	(3) Acquitted	(4) Acquitted	(5) Acquitted	(6) Acquitted
Same last name	-0.0006	-0.0012	0.0088	0.0078	0.0015	0.0009
	(0.0013)	(0.0013)	(0.0055)	(0.0055)	(0.0045)	(0.0045)
Same name * Rare name					0.0212	0.0199
					(0.0142)	(0.0143)
Observations	2081855	2080529	2081855	2080529	2081855	2080529
Fixed Effect	Court-month	Court-month	Court-month	Court-month	Court-month	Court-month
Judge Fixed Effect	No	Yes	No	Yes	No	Yes
Inverse Group Weight	No	No	Yes	Yes	Yes	Yes
Last Name Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

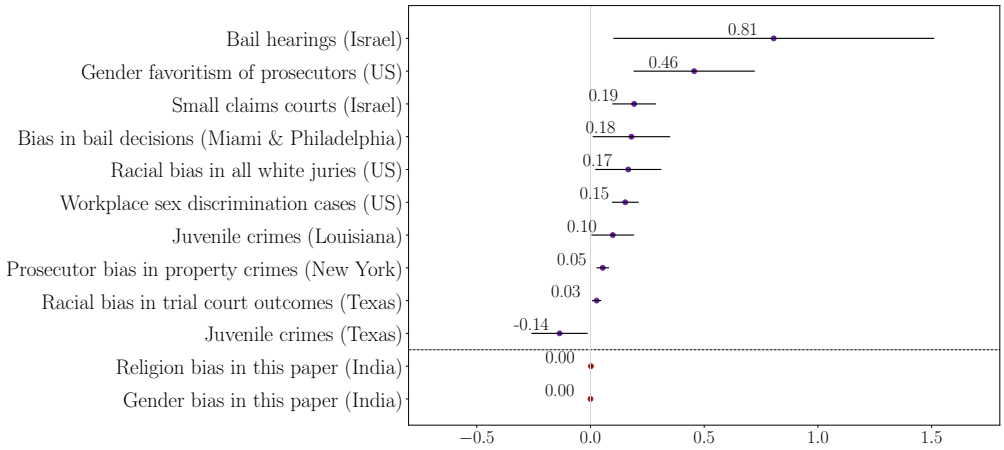
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports results from a test of the effect of assignment to a judge with the same last name as the defendant on likelihood of acquittal (Equation 2). Court-month fixed effects, charge section fixed effects, and judge and defendant last name fixed effects have been used across all columns reported. Standard errors are clustered by judge.

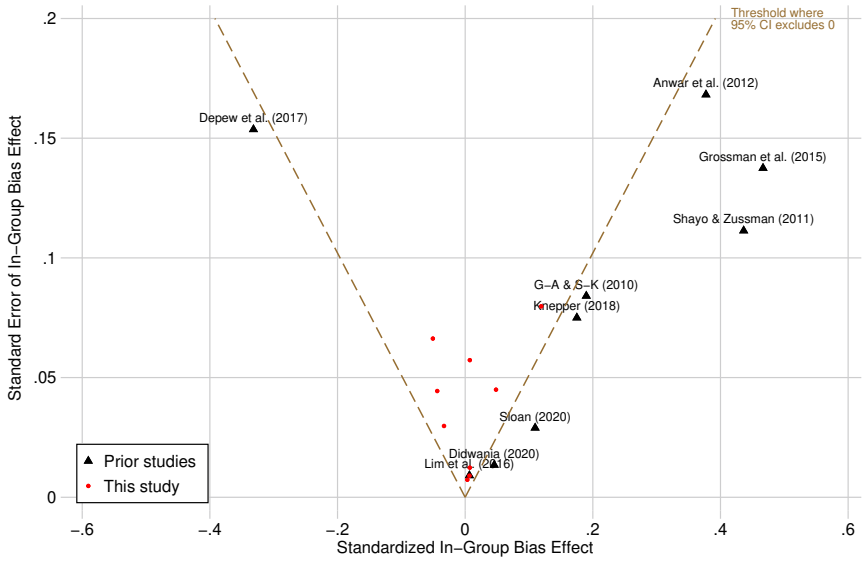
Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Figure 2: Comparison with judicial bias estimates in other contexts

A. Coefficient Plot



B. Standardized Errors vs. Effect Sizes



Notes: This figure shows point estimates of in-group bias from other studies in the relevant literature. From the top, the coefficients of in-group bias (Panel A) correspond to Grossman et al., 2016, Shayo and Zussman, 2011, Anwar et al., 2012, Depew et al., 2017, Gazal-Ayal and Sulitzeanu-Kenan, 2010, Knepper, 2018, Sloane, 2019, Didwania, 2022, Lim et al., 2016, and the main estimates from the present study respectively. Shayo and Zussman, 2017 is excluded because the underlying data and variation overlap substantially with Shayo and Zussman, 2011. Panel B plots reported bias effects (Y axis) against effect standard errors. All effect sizes are standardized (dividing outcome variables by their standard deviation) to allow comparison across studies. From each table in this paper, we chose the specification with court-month and judge fixed effects. For contexts magnifying bias, we show the average effect for the group facing magnified bias. For example, for the Ramadan analysis, we show the sum of the bias coefficient and the bias * Ramadan coefficient, which describes religious in-group bias in the month of Ramadan.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Table 6: Estimates of Publication Bias in Judicial In-Group Bias Studies

	(1)	(2)	(3)	(4)	(5)
	$p(z) = \Pr(\text{Pub} \mid \text{t-stat})$				
	$(-\infty, -1.96]$	$(-1.96, 0]$	$(0, 1.96]$	$(1.96, \infty]$	β^*
Estimate	.0912	0.00	0.029	1.00	0.046
Standard Error	(1.752)	(0.044)	(0.035)	.	(0.020)

Notes: The table summarizes in-group bias in the judicial setting, measured across all papers we could find using randomized assignment of judges and juries, with adjustment for publication bias. Columns 1–4 respectively show the probability that a study gets published, given a t-statistic in the range of $(-\infty, -1.96]$, $(-1.96, 0]$, $(0, 1.96]$, and $(1.96, \infty)$ respectively. β^* in Column 5 gives the true predicted average in-group bias effect after taking publication bias into account and imputing unpublished studies. Estimates were calculated from the papers listed in Figure 2 (not including estimates from this paper), following Andrews and Kasy, 2019.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

References

Abrams, D. S., Bertrand, M., & Mullainathan, S. (2012). Do Judges Vary in Their Treatment of Race? *The Journal of Legal Studies*, *41*(2), 347–383. https://doi.org/10.1086/666006

Alesina, A., & La Ferrara, E. (2014). A Test of Racial Bias in Capital Sentencing. *The American Economic Review*, *104*(11), 3397–3433. https://doi.org/10.1257/aer.104.11.3397

Alexander, M. (2010). *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. The New Press.

Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, *109*(8). https://doi.org/10.1257/aer.20180310

Aney, M. S., Dam, S., & Ko, G. (2021). Jobs for Justice(s): Corruption in the Supreme Court of India. *The Journal of Law and Economics*, *64*(3), 479–511. https://doi.org/10.1086/713728

Anwar, S., Bayer, P., & Hjalmarsson, R. (2012). The Impact of Jury Race in Criminal Trials. *The Quarterly Journal of Economics*, *127*(2), 1–39. https://doi.org/10.1093/qje/qjs014

Anwar, S., Bayer, P., & Hjalmarsson, R. (2019). A Jury of Her Peers: The Impact of the First Female Jurors on Criminal Convictions. *The Economic Journal*, *129*(618), 603–650. https://doi.org/10.1111/ecoj.12562

Anwar, S., & Fang, H. (2006). An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence. *American Economic Review*, *96*(1), 127–151. https://doi.org/10.1257/000282806776157579

Arnold, D., Dobbie, W., & Hull, P. (2022). Measuring Racial Discrimination in Bail Decisions. *American Economic Review*, *112*(9), 2992–3038. https://doi.org/10.1257/aer.20201653

Arnold, D., Dobbie, W., & Yang, C. S. (2018). Racial Bias in Bail Decisions. *The Quarterly Journal of Economics*, *133*(4), 1885–1932. https://doi.org/10.1093/qje/qjy012

Asher, S., Novosad, P., & Rafkin, C. (2024). Intergenerational Mobility in India: New Measures and Estimates across Time and Social Groups. *American Economic Journal: Applied Economics*, *16*(2), 66–98. https://doi.org/10.1257/app.20210686

Baldus, D. C., Woodworth, G., Zuckerman, D., & Weiner, N. A. (1997). Racial Discrimination and the Death Penalty in the Post-Furman Era: An Empirical and Legal Overview with Recent Findings from Philadelphia. *Cornell L. Rev.*, *83*, 1638.

Baumgartner, F. R., Grigg, A. J., & Mastro, A. (2015). #BlackLivesDon'tMatter: Race-Of-Victim Effects in US Executions, 1976–2013. *Politics, Groups, and Identities*, *3*(2), 209–221. https://doi.org/10.1080/21565503.2015.1024262

Bertrand, M., Hanna, R., & Mullainathan, S. (2010). Affirmative Action in Education: Evidence from Engineering College Admissions in India. *Journal of Public Economics*, *94*(1-2), 16–29. https://doi.org/10.1016/j.jpubeco.2009.11.003

Bhalotra, S., Clots-Figueras, I., Iyer, L., & Cassan, G. (2012). Politician Identity and Religious Conflict in India. *Working Paper*. https://www.theigc.org/sites/default/files/2014/10/Bhalotra-et-al-2012-Working-Paper.pdf

Bharti, N. K., & Roy, S. (2023). The Early Origins of Judicial Stringency in Bail Decisions: Evidence from Early Childhood Exposure to Hindu-Muslim Riots in India. *Journal of Public Economics*, *221*. https://doi.org/10.1016/j.jpubeco.2023.104846

Borker, G. (2021). Safety First: Perceived Risk of Street Harassment and Educational Choices of Women. *World Bank Policy Research Working Paper No. 9731*. https://hdl.handle.net/10986/36004

Boyd, C., Epstein, L., & Martin, A. D. (2010). Untangling the Causal Effects of Sex on Judging. *American Journal of Political Science*, *54*(2), 389–411. https://doi.org/10.1111/j.1540-5907.2010.00437.x

Canay, I. A., Mogstad, M., & Mountjoy, J. (2023). On the Use of Outcome Tests for Detecting Bias in Decision Making. *The Review of Economic Studies*, *91*(4), 2135–2167. https://doi.org/10.1093/restud/rdad082

Chaturvedi, R., & Chaturvedi, S. (2024). It's All in the Name: A Character-Based Approach to Infer Religion. *Political Analysis*, *32*(1), 34–49. https://doi.org/10.1017/pan.2023.6

Chemin, M. (2009). Do Judiciaries Matter for Development? Evidence from India. *Journal of Comparative Economics*, *37*(2), 230–250. https://doi.org/10.1016/j.jce.2009.02.001

Chew, P. K., & Kelley, R. E. (2009). Myth of the Color-Blind Judge: An Empirical Analysis of Racial Harassment Cases. *Wash. U. L. Rev.*, *86*, 1117.

Cousins, S. (2020). 2.5 Million more Child Marriages due to COVID-19 Pandemic. *The Lancet*, *396*(10257), 1059. https://doi.org/10.1016/S0140-6736(20)32112-7

Depew, B., Eren, O., & Mocan, N. (2017). Judges, Juveniles, and In-Group Bias. *The Journal of Law and Economics*, *60*(2), 209–239. https://doi.org/10.1086/693822

Dev, A. (2019). *India's Supreme Court Is Teetering on the Edge*. Retrieved June 20, 2024, from https://www.theatlantic.com/international/archive/2019/04/india-supreme-court-corruption/587152/

Didwania, S. H. (2022). Gender Favoritism among Criminal Prosecutors. *The Journal of Law and Economics*, *65*(1), 77–104. https://doi.org/10.1086/718463

Djankov, S., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2003). Courts. *Quarterly Journal of Economics*, *118*(2), 453–517. https://doi.org/10.1162/003355303321675437

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in Meta-Analysis Detected by a Simple, Graphical Test. *BMJ*, *315*(7109), 629–634.

Erken, A., Chalasani, S., Diop, N., Liang, M., Weny, K., Baker, D., Baric, S., Guilmoto, C., Luchsinger, G., Mogelgaard, K., & et al. (2020). *State of the World Population 2020*. United Nations Population Fund. https://www.unfpa.org/sites/default/files/pub-pdf/UNFPA_PUB_2020_EN_State_of_World_Population.pdf

Fisman, R., Paravisini, D., & Vig, V. (2017). Cultural Proximity and Loan Outcomes. *American Economic Review*, *107*(2), 457–92. https://doi.org/10.1257/aer.20120942

Fisman, R., Sarkar, A., Skrastins, J., & Vig, V. (2020). Experience of Communal Conflicts and Intergroup Lending. *Journal of Political Economy*, *128*(9), 3346–3375. https://doi.org/10.1086/708856

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

ForsterLee, R., ForsterLee, L., Horowitz, I. A., & King, E. (2006). The Effects of Defendant Race, Victim Race, and Juror Gender on Evidence Processing in a Murder Trial. *Behavioral Sciences & the Law*, *24*(2), 179–198. https://doi.org/10.1002/bsl.675

Frandsen, B., Lefgren, L., & Leslie, E. (2023). Judging Judge Fixed Effects. *American Economic Review*, *113*(1), 253–77. https://doi.org/10.1257/aer.20201860

Gazal-Ayal, O., & Sulitzeanu-Kenan, R. (2010). 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. https://doi.org/10.1111/j.1740-1461.2010.01183.x

Gerber, A. S., Green, D. P., & Nickerson, D. (2001). Testing for Publication Bias in Political Science. *Political Analysis*, *9*(4), 385–392. https://doi.org/10.1093/oxfordjournals.pan.a004877

Goel, S., Rao, J. M., Shroff, R., et al. (2016). Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-And-Frisk Policy. *The Annals of Applied Statistics*, *10*(1), 365–394. https://doi.org/10.1214/15-AOAS897

Grossman, G., Gazal-Ayal, O., Pimentel, S. D., & Weinstein, J. M. (2016). Descriptive Representation and Judicial Outcomes in Multiethnic Societies. *American Journal of Political Science*, *60*(1), 44–69. https://doi.org/10.1111/ajps.12187

Hanna, R. N., & Linden, L. L. (2012). Discrimination in Grading. *American Economic Journal: Economic Policy*, *4*(4), 146–68. https://doi.org/10.1257/pol.4.4.146

Ito, T. (2009). Caste Discrimination and Transaction Costs in the Labor Market: Evidence from Rural North India. *Journal of Development Economics*, *88*(2), 292–300. https://doi.org/10.1016/j.jdeveco.2008.06.002

Jayachandran, S. (2015). The Roots of Gender Inequality in Developing Countries. *Annual Review of Economics*, *7*. https://doi.org/10.1146/annurev-economics-080614-115404

Kastellec, J. P. (2011). Panel Composition and Voting on the US Courts of Appeals over Time. *Political Research Quarterly*, *64*(2), 377–391. https://doi.org/10.1177/1065912909356889

Kastellec, J. P. (2013). Racial Diversity and Judicial Influence on Appellate Courts. *American Journal of Political Science*, *57*(1), 167–183. https://doi.org/10.1111/j.1540-5907.2012.00618.x

Knepper, M. (2018). When the Shadow is the Substance: Judge Gender and the Outcomes of Workplace Sex Discrimination Cases. *Journal of Labor Economics*, *36*(3), 623–664. https://doi.org/doi.org/10.1086/696150

Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation Between Effect Size and Sample Size. *PloS one*, *9*(9), e105825. https://doi.org/10.1371/journal.pone.0105825

Kumar, C., & Sharan, M. (2023). It's Complicated: The Distributional Consequences of Political Reservation. *Working Paper*.

La Porta, R., Lopez-de-Silanes, F., Pop-Eleches, C., & Shleifer, A. (2004). Judicial Checks and Balances. *Journal of Political Economy*, *112*(2). https://doi.org/10.1086/381480

La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2008). The Economic Consequences of Legal Origins. *Journal of Economics Literature*, *46*(2), 285–332. http://www.jstor.org.ezproxy.cul.columbia.edu/stable/27646991

Le, Q. V. (2004). Political and Economic Determinants of Private Investment. *Journal of International Development*, *16*(4), 589–604. https://doi.org/10.1002/jid.1109

Levine, T. R., Asada, K. J., & Carpenter, C. (2009). Sample Sizes and Effect Sizes are Negatively Correlated in Meta-Analyses: Evidence and Implications of a Publication Bias Against Non-significant Findings. *Communication Monographs*, *76*(3), 286–302. https://doi.org/10.1080/03637750903074685

Lichand, G., & Soares, R. R. (2014). Access to Justice and Entrepreneurship: Evidence from Brazil's Special Civil Tribunals. *The Journal of Law and Economics*, *57*(2). https://doi.org/10.1086/675087

Lim, C. S., Silveira, B. S., & Snyder, J. M. (2016). Do Judges' Characteristics Matter? Ethnicity, Gender, and Partisanship in Texas State Trial Courts. *American Law and Economics Review*, *18*(2), 302–357. https://doi.org/10.1093/aler/ahw006

Marx, B., Stoker, T. M., & Suri, T. (2019). There is no free house: Ethnic patronage in a kenyan slum. *American Economic Journal: Applied Economics*, *11*(4), 36–70. https://doi.org/10.1257/app.20160484

Mehmood, S., Seror, A., & Chen, D. L. (2023). Ramadan Fasting Increases Leniency in Judges from Pakistan and India. *Nature Human Behaviour*, *7*(6), 874–880. https://doi.org/10.1038/s41562-023-01547-3

Mullen, B., Brown, R., & Smith, C. (1992). Ingroup Bias as a Function of Salience, Relevance, and Status: An Integration. *European Journal of Social Psychology*, *22*(2), 103–122. https://doi.org/https://doi.org/10.1002/ejsp.2420220202

Mustard, D. B. (2001). Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts. *The Journal of Law and Economics*, *44*(1), 285–314. https://doi.org/10.1086/320276

National Crime Records Bureau. (2018). *Crime in India* (tech. rep.). Ministry of Home Affairs.

Neggers, Y. (2018). Enfranchising your own? experimental evidence on bureaucrat diversity and election bias in india. *American Economic Review*, *108*(6), 1288–1321. https://doi.org/10.1257/aer.20170404

Pande, R., & Udry, C. (2005). Institutions and Development: A View from Below. *Yale University Economic Growth Center Discussion Paper No. 928*. https://elischolar.library.yale.edu/egcenter-discussion-paper-series/936

Peresie, J. L. (2005). Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts. *The Yale Law Journal*, *114*(7), 1759–1790. http://www.jstor.org/stable/4135764

Ponticelli, J., & Alencar, L. S. (2016). Court Enforcement, Bank Loans, and Firm Investment: Evidence From a Bankruptcy Reform in Brazil. *The Quarterly Journal of Economics*, *131*(3), 1365–1413. https://doi.org/10.1093/qje/qjw015

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01569

Rao, M. (2020). Judges, Lenders, and the Bottom Line: Courting Firm Growth in India. *Working Paper*. http://manaswinirao.com/files/manaswini_jmp.pdf

Rehavi, M. M., & Starr, S. B. (2014). Racial Disparity in Federal Criminal Sentences. *Journal of Political Economy*, *122*(6), 1320–1354. https://doi.org/10.1086/677255

Rodrik, D. (2000). Institutions for High-Quality Growth: What They are and how to Acquire Them. *Studies in Comparative International Development*, *35*(3), 3–31. https://doi.org/10.1007/BF02699764

Rodrik, D. (2005). Growth Strategies. *Handbook of Economic Growth*, *1*, 967–1014.

Sachar Committee Report. (2006). *Social, Economic and Educational Status of the Muslim Community of India* (tech. rep.). Government of India.

Sandefur, J., & Siddiqi, B. (2015). Delivering Justice to the Poor: Theory and Experimental Evidence from Liberia. *Working Paper*.

Scherer, N. (2004). Blacks on the Bench. *Political Science Quarterly*, *119*(4), 655–675. https://doi.org/10.1002/j.1538-165X.2004.tb00534.x

Shayo, M., & Zussman, A. (2011). Judicial Ingroup Bias in the Shadow of Terrorism. *The Quarterly Journal of Economics*, *126*(3), 1447–1484. https://doi.org/10.1093/qje/qjr022

Shayo, M., & Zussman, A. (2017). Conflict and the persistence of ethnic bias. *American Economic Journal: Applied Economics*, *9*(4). https://doi.org/10.1257/app.20160220

Singh, K. S. (1992). *People of india* (Vol. 1). Anthropological Survey of India.

Slavin, R., & Smith, D. (2009). The Relationship Between Sample Sizes and Effect Sizes in Systematic Reviews in Education. *Educational Evaluation and Policy Analysis*, *31*(4), 500–506. https://doi.org/10.3102/0162373709352369

Sloane, C. (2019). Racial Bias by Prosecutors: Evidence from Random Assignment. *ICCJ 2019: International Conference on Criminal Justice*.

Tata Trusts. (2019). India Justice Report: Ranking States on Police, Judiciary, Prisons & Legal Aid.

The Economist. (2024). *How independent is India's Supreme Court?* Retrieved February 22, 2024, from https://www.economist.com/asia/2024/02/22/how-independent-is-indias-supreme-court

The World Bank Group. (2017). *World Development Report 2017: Governance and the Law*. The World Bank Group. https://doi.org/10.1596/978-1-4648-0950-7

University of Denver Institute for the Advancement of the American Legal System. (2021). *FAQs: Judges in the United States*. Retrieved December 10, 2022, from https://iaals.du.edu/sites/default/files/documents/publications/judge_faq.pdf

Vahini, M., Bantupalli, J., Chakraborty, S., & Mukherjee, A. (2022). Decoding Demographic Un-Fairness from Indian Names [Working Paper]. https://doi.org/10.48550/arXiv.2209.03089

Varshney, A., & Wilkinson, S. (2006). *Varshney-Wilkinson Dataset on Hindu-Muslim Violence in India, 1950–1995, version 2*. Inter-university Consortium for Political; Social Research.

Visaria, S. (2009). Legal Reform and Loan Repayment: The Microeconomic Impact of Debt Recovery Tribunals in India. *American Economic Journal: Applied Economics*, *1*(3), 59–81. https://doi.org/10.1257/app.1.3.59