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*

April 7, 2022

Abstract

We study judicial in-group bias in Indian criminal courts using a newly collected dataset on over 5 million criminal case records from 2010–2018. After detecting gender and religious identity using a neural-net classifier applied to judge and defendant names, we exploit quasi-random assignment of cases to judges to examine whether defendant outcomes are affected by assignment to a judge with a similar identity. In the aggregate, we estimate tight zero effects of in-group bias based on shared gender, religion, and last name (a proxy for caste). We do find limited in-group bias in some (but not all) settings where identity is salient – in particular, we find a small religious in-group bias during Ramadan, and we find shared-name in-group bias when judge and defendant match on a rare last name.
JEL codes: J15, J16, K4, O12

*Author Details: Ash, ETH Zurich: ashe@ethz.ch; Asher, John Hopkins: sasher2@jhu.edu; Bhowmick, Development Data Lab: bhowmick@devdatalab.org; Bhupatiraju: World Bank: sbhupatiraju@worldbank.org; Chen, Toulouse and World Bank: daniel.chen@iast.fr; Devi, Harvard: tdevi@g.harvard.edu; Goessmann, ETH Zurich: christoph.goessmann@gess.ethz.ch; Novosad, Dartmouth College: paul.novosad@dartmouth.edu; Siddiqi, UC Berkeley: bilal.siddiqi@berkeley.edu. 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, and the DFID Economic Development and Institutions program for financial support. For helpful feedback we thank participants of the Political Economy Seminar at ETH Zurich, Delhi School of Economics Winter School 2020, Texas Economics of Crime Workshop, Midwest International Economic Development Conference, Discrimination 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 Introduction

Structural inequalities across groups defined by gender, religion, and ethnicity are seen in almost all societies. Governments often try to remedy these inequalities through policies, such as anti-discrimination statutes or affirmative action, which must then be enforced by the legal system. A challenging problem is that the legal system itself may have unequal representation. It remains an open question whether legal systems in developing countries are effective at pushing back against structural inequality or whether they serve to entrench it.

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 nearly universal 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 and 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 and 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 and Linden, 2012; Jayachandran, 2015; Borker, 2017; Asher et al., 2020). In India's lower courts, unequal representation is a recognized issue. Women represent half the population but only 28% of district court judges. Similarly, India's 200 million Muslims represent 14% of the population but only 7% of district court judges.[2] 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 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

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

[2] Source: eCourts data, see Table 1. We did not find a statistic for overall representation of Scheduled Castes in the judiciary, but partial evidence suggests they are also underrepresented (Times of India, 2018).

for cases heard in India's trial courts.[3] These cases cover the universe of India's 7,000+ district and subordinate trial courts, staffed by over 80,000 judges. 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.[4]

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 new 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 Deli 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. This accuracy is significantly higher than the standard approach of fuzzy matching.[5] We apply the trained model to our case dataset to assign identity characteristics to judges, defendants, and victims.

Compared to gender and religion, caste identity is relatively complex and hierarchical, making it difficult to specify binary in-groups and out-groups. Because assigning categorical caste memberships based on names is not feasible, we instead define a caste identity match as a case where the defendant's last name matches the judge's last name. This is an imperfect measure because multiple family names may reflect the same caste and certain last names may be used by members of many castes. Nevertheless, for many names, individuals in the same region who share a last name are likely to belong to the same caste.[6]

The research question is whether judges treat defendants differently when they share the same gender, religion, or caste. 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 extreme delays in India's judicial system (Trusts, 2019; Rao, 2019), we additionally examine whether in-group judge identity

[3]The eCourts platform can be accessed at https://ecourts.gov.in/. That site hosts the case records only through a slow search engine that returns unstructured results. The data was not previously available as a structured dataset or API.

[4]The data can be accessed at https://www.devdatalab.org/judicial-data. The total dataset – civil and criminal, without filtering – contains 77 million case records. Users of the data are asked to cite this paper.

[5]We have made the name classifier code 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.

[6]Using the same last name to classify identity groups has predicted preferential outcomes in previous work, for instance in the banking setting (Fisman et al., 2017).

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.[7]

We find a precise and robust null estimate of in-group bias among Indian judges on all three dimensions. In the aggregate, sharing gender, religion, or last name with a defendant makes a judge no more likely to deliver a positive outcome. This null is seen both in decisions (i.e. acquittals and convictions) and in process (i.e. speed of decision). 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.[8] The upper end of our 95% confidence interval 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 (see Figure 2).

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 and Zussman, 2011; Anwar et al., 2012; Mehmood et al., 2021). First, we examine cases where the defendant and the victim 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. In both of these subset analyses, we continue to find a null bias.

Third, we examine whether in-group bias on the basis of religion is activated during the month of Ramadan, when religious identity may be more salient. We find suggestive

[7]Results are robust to adding judge fixed effects (which control for variation in the severity of specific judges), though these are not expected to make a difference under random assignment of cases to judges.

[8]The exception is Lim et al. (2016), who find zero effects of in-group gender bias and marginal effects of in-group racial bias among judges in Texas state district courts, notably the statistically highest-powered study in this class before ours.

evidence that being assigned to a judge with the same religious identity (i.e. Muslim or non-Muslim) raises the probability of acquittal *when the case is heard during Ramadan*. The estimate is only marginally statistically significant due to the smaller sample during Ramadan months.[9] This result confirms that district judges have discretion in their decisions and may apply that discretion in favor of an in-group if their identity is activated. But in most cases, and most of the time, the extent of religious and gender in-group bias in acquittal and conviction rates in Indian courts is effectively zero.

Fourth and finally, we examine the last name bias when defendants have uncommon last names. In this case, the shared identity with the judge is more narrowly defined, which may magnify the sense of shared identity. Here, we find statistically and economically important signs of pro-in-group bias. The effect remains small in aggregate because it applies to a narrow subset of defendants who both have uncommon names and are lucky enough to be assigned a judge with the same uncommon name. Still, we cannot rule out that judges show bias based 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 Arnold et al. (2018), 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 taken as seriously by the police. Our null estimates are nevertheless notable, given substantial evidence of this kind of bias in other countries, and in other settings in India.

In Section 6, we discuss several potential 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 Ramadan and rare-last-name results, the overall salience of the shared identity group.

[9]The point estimate, a two percentage point effect, is also small when compared with the prior literature. In particular, Mehmood et al. (2021) find that acquittal rates rise by 23 percentage points (or 40%) during Ramadan in Pakistan, and in India they rise by 7 percentage points for each additional hour of fasting. Mehmood et al 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.

The nding that in-group bias emerges only in cases where identity is salient is informative for our understanding of prior work, which consistently nds large in-group e ects 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 nd 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 contextswithout in-group bias are not prominently described in completed papers. To assess this possibility, we aggregate the e ect sizes and standard errors from earlier papers with highly similar empirical designs to ours. Following the approach from Andrews and Kasy (2019), we nd evidence consistent with a high degree of publication bias. Some studies with null results seem to be missing from the literature, perhaps due to projects being abandoned or failing to make it through the peer review process.

Our study makes two substantive 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 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. We also analyze the universe of criminal cases, heading o most concerns about external validity within the study context. Second, our ndings of di erential bias e ects in certain special cases when in-group size is small or when the external environment increases the salience of identity helps shed light on contexts where bias may be more or less likely to occur. In particular, the large and signi cant 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 de ned and recognizable, and the external environment is heightened.

More speci cally, our substantive 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.[10] The closest paper to ours is Shayo and

[10] These include racial disparities in the execution of stop-and-frisk programs (Goel et al., 2016), motor vehicle searches by police troopers (Anwar and Fang, 2006), bail decisions (Arnold et al., 2018),

Zussman (2011), who analyze the e ect of assigning a Jewish versus an Arab judge in Israeli small claims court. They nd robust evidence of in-group bias, where Jewish judges favor Jewish defendants (and Arab judges favor Arab defendants). Our nding that religious bias is magni ed during the month of Ramadan is consistent with their notion of endogenous social identi cation, though our point estimates on bias are an order of magnitude smaller even under these high-salience conditions.

A handful of other studies use quasi-random designs to estimate in-group biases in a similar fashion to our analysis. While most of these papers report large and statistically signi cant pro-in-group e ects, one paper nds anti-in-group bias.[11] Of the papers we could nd, only Lim et al. (2016) nd a null in-group e ect of judge ethnicity or gender, notably with the largest sample size in this set of papers (N=250,000).

In the Indian legal context, there is a growing body of evidence on the legal system, mostly focusing on judicial e cacy and economic performance (Chemin, 2009; Rao, 2019), and on corruption in the Indian Supreme Court (Aney et al., 2017). A recent working paper nds that judges are more prone to deny bail if they had been exposed to communal riots in early childhood (Bharti and Roy, 2020). We are aware of no prior large-scale empirical research on unequal legal treatment in India, a topic of substantial policy relevance.

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 e ective, 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

charge decisions (Rehavi and Starr, 2014), and judge sentence decisions (Mustard, 2001; Abrams et al., 2012; Alesina and La Ferrara, 2014; Kastellec, 2013). African-American judges have been found to vote di erently from Caucasian-American judges on issues where minorities are disproportionately a ected, such as a rmative action, racial harassment, unions, and search and seizure cases (Scherer, 2004; Chew and Kelley, 2008; Kastellec, 2011). In a similar manner, a number of papers have documented the e ect 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 nd 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 nd that the e ect of including even one Arab judge on the decision-making panel substantially in uences trial outcomes of Arab defendants. Didwania (2018) nd in-group bias in that prosecutors charge same-gender defendants with less severe o enses.

[11] Gazal-Ayal and Sulitzeanu-Kenan (2010) nd 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, nding signi cant in-group bias in trial outcomes. Depew et al. (2017) exploit random assignment of judges to juvenile crimes in Louisiana and nd negative in-group bias in sentence lengths and likelihood of being placed in custody.

and Udry, 2005; Visaria, 2009; Lichand and Soares, 2014; Ponticelli and Alencar, 2016; Bank, 2017). A handful of important cross-country studies have recovered some broad stylized facts on the causes and consequences of di erent broad features of legal systems (Djankov et al., 2003; La Porta et al., 2004, 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. Hence, a nal key contribution of this paper is the 77 million case dataset that we have posted, which may enable a wide range of future research projects in this domain.

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.

2 Background

2.1 Gender 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. 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 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.[12]

These courts are sta ed by over 80,000 judges. Due to common law institutions where court rulings serve as binding precedent in future cases, judges in India are e ectively 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 e cient 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 quali cation. 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 o ces in the judiciary after speci c 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) as well as a substantial backlog of cases (Trusts, 2019). In 2015, Prime Minister Modi attempted to implement a series of reforms giving his administration more control over judge selection by creating a National Judicial Appointments Commission. However, the e ort to move away from the collegium system of judicial appointment was reversed by the Supreme Court, citing breach of judicial independence.

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

[12] We de ne criminal cases as all cases led either under the Indian Penal Code Act or the Code of Criminal Procedure Act.

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 led. Each police station lies within the territorial jurisdiction of a speci c district courthouse, which receives the case. The case is then 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 speci c 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 di erent courtroom, as might a murder charge reported at a di erent police station. The police station charge lists leave little room for discretion over which charges are seen by which judges.

Judges typically spend two to three years in a given court, during which they rotate through several of the courtrooms.[13] The timing of the rst court appearance is unknown when charges are led (given judicial delays). Thus, even if a defendant or prosecutor had discretion over which police station led the charges, the rotation of judges between courtrooms would make it di cult to target a speci c 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.[14,15]

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 was introduced in India in the early 2000s, less than

[13] Severe cases (with severity de ned by the section or act under which the charge was led) 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.

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

[15] In Section 4, we present formal tests of the exogenous assignment of judges to cases in our dataset.

0.5% of all criminal cases pending in India are disposed through plea bargaining. It is thus unlikely to play a major factor in our analysis.

3 Data

3.1 Case Records

We obtained 77 million case records from the Indian eCourts platform a semi-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.[16] The publicly available information includes the ling, registration, hearing, and decision dates for each case, petitioner and respondent names, the position of the presiding judge, the acts and sections under which the case was led, and the nal decision or disposition.[17]

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. This paper focuses on cases led 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 de nition of identity match between judge and defendant. Second, it is relatively straightforward to identify good and bad outcomes for criminal defendants, which is more di cult in civil cases. This constraint lters out 70% of the dataset, leaving us with 23 million criminal case records (see Appendix Figure A2).

3.2 Judge Information

We also 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.[18]

We joined the case-level data with the judge-level data based on the judge's designation and the initial case ling 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

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

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

[18] See Appendix Figure A3 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.

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 (see Appendix Figure A2).

3.3 Assigning Religion and Gender Identity

The eCourts platform does not provide demographic metadata on judges and defendants. However, gender and religious identity can be determined quite accurately in India based on individuals' names. We train a machine classi er on a large database of labeled names and then use it to assign these characteristics in the legal data.[19]

We use two databases of names with associated demographic labels. To classify gender, we use a dataset of 13.7 million names with labeled gender from the Delhi voter rolls. To classify religion, we use a database of 1.4 million names with a religion label for individuals who sat for the National Railway Exam.

Summary tabulations on these datasets are provided in Appendix Table A2. For gender, we observe two categories: female or male. For religion, we observe ve categories: Hindu, Muslim, Christian, Buddhist, and Other. Our classi er takes a two-label speci cation: Muslim or non-Muslim. We do not distinguish between the non-Muslim religion categories because of their small number and because their names are not as distinctive as Muslim names. Each name record is therefore assigned two binary labels: male/female and Muslim/Non-Muslim.

The lists of labeled names from the Delhi voter rolls and National Railway Exam contain some inconsistent formatting and noise which we clean up with a set of pre-processing steps. First, Hindi characters are transliterated to Latin. Second, we normalize capitalization, punctuation, and spacing. Salutations are preserved as they indicate gender.

Taking these pre-processed name strings as inputs, we train a neural net classi er to predict the associated identity label. We use a bidirectional Long Short-Term Memory (LSTM) model applied directly to the sequence of name-string characters. LSTM uses a gated recurrent neural network architecture that takes as input a sequential data stream

[19] The existing available name classi ers for gender and religion in India are expensive proprietary solutions, e.g. Namsor (namsor.com), and trials with these yielded the same or lower accuracy than our own classi er.

and retains a memory of previous inputs while handling new items in the sequence. LSTMs are particularly useful in understanding text sequences because the meaning of an individual letter or word is often dependent on the context of other letters and words that both precede and follow it. Bidirectional means that the classi er reads the sequence backward and forward when trying to assign a label.[20]

The ability of the LSTM classi er to understand a text fragment within context greatly improves accuracy over standard fuzzy string matching methods. For instance, consider the last names Khan and Khanna. While the fragment KHAN appears in both words, the addition of two letters na following the fragment changes the meaning of the word where it is a distinctly Muslim last name without the letters na, and a non-Muslim last name once the letters na are added. A standard fuzzy match would fail on this example because it ignores the context (that is, the sequence of letters that appear before and after the fragment KHAN). A counter-example are the names Fatima and Fathimaa, where the addition of the letters h and a do not change the religious classi cation of the name. Given these nuances, the LSTM classi er is better suited to the objective than a simple fuzzy matching function.

We use hold-out test sets within the labeled databases to assess the out-of-sample performance of the LSTM classi ers for gender and religion. The classi ers perform well on standard metrics, including our preferred metrics that adjust for imbalance in the class shares. We report balanced accuracy, which is the average accuracy (recall) for each of the two identity categories, and F1, the harmonic mean of precision and recall.[21] For gender, the balanced accuracy is .975 with F1 = .976. For religion, the balanced accuracy is .98 and F1 = .99. The trained classi ers, as well as the code for training

[20] In more detail, the neural net architecture is as follows. The model takes as input a sequence of characters and outputs a probability distribution across name classes. The characters are input to an embedding layer, which was initialized randomly rather than using pre-trained weights. The embedded vectors are input to a bidirectional LSTM layer, then to a single dense hidden layer, and nally to the output layer, which uses sigmoid activation to output a probability across the binary classes. To avoid over tting, we used dropout between layers and used early stopping during training, which ceases network training when validation loss stops improving. To account for the imbalance in the sample, we used class weights during the training. See Chaturvedi and Chaturvedi (2020) for a similar approach to infer religion from names.

[21] Balanced accuracy and F1 are preferred as metrics to standard accuracy when the labels to be predicted are not balanced. While gender is roughly balanced in the voter rolls data, religion is heavily imbalanced with Muslims only comprising one-tenth of the sample. Therefore a model could achieve 90% accuracy in predicting religion by guessing non-Muslim. Balanced accuracy addresses this issue by rewarding good accuracy for both classes: we calculate the accuracy for each class and then average, rather than taking the accuracy measure across the whole sample. F1 addresses this issue by rewarding higher precision, which penalizes false positives, and higher recall, which penalizes false negatives.

them, are available as open-source software for use by the academic community. The code and the trained gender classi er are available at our GitHub repository.[22]

The next step is to apply the trained classi er to the eCourts case records. We have plain-text string variables for judge name and defendant name, to which we apply the same pre-processing steps as above (i.e., transliteration and normalization of punctuation/capitalization). We lter out names that are not possible to classify, for example due to emptiness. For defendants, in addition, we drop names that refer to governments or organizations (e.g. The State of Maharashtra).

For each pre-processed judge name and defendant name, we then apply the trained classi er and form a predicted probability for gender and religion. To improve precision, we further lter names which do not produce a con dent classi cation. Comparing the con dence scores to human annotations, we see that predicted probabilities near 50% include mostly ambiguous names. This happens for gender, for example, when the rst name is missing and no salutation is included. As a heuristic to drop these names, we set a con dence threshold that requires the model to be at least 65% con dent in a predicted gender or religion classi cation. For predicted probabilities between 0.35 and 0.65, the respective class is left empty.

For judges, names tend to be complete or else include salutations. Of the 81,232 judges (22,413 unique names) appearing in the case dataset, we are able to classify 96% according to gender (female/male) and 98% according to religion (Muslim/non-Muslim). The information on defendant names is of lower quality, mainly due to missing rst or last names. Still, we are able to classify 80% of defendants by religion and 74% by gender.[23] Cases with unclassi ed labels are dropped from analyses requiring those labels.

To verify the accuracy of the LSTM classi cation within the new domain of the court records, we manually checked a random sample of names classi ed by the above process. An annotator manually labeled 100 names by gender and 100 names by religion, with samples strati ed across states. This process con rms an accuracy of 97% for both the gender and religion classi cation in the new domain.[24]

[22] See https://github.com/devdatalab/paper-justice/tree/main/classifier .

[23] The proportion of defendants that can be assigned to gender and religion does not vary much by region of India (Appendix Table A1).

[24] As an additional automated validation, we compared the LSTM-classi ed Muslim defendant share by state to the state-level Muslim population shares from the 2011 Population Census. The correlation is 0.88.

3.4 De ning Case Outcomes

We de ne the defendant's outcome (represented by Y below) as a case-level indicator variable that takes the value one if the outcome is desirable for the defendant and zero otherwise. Our primary speci cation uses an indicator for defendant acquittal. A secondary speci cation uses an indicator for any outcome other than conviction. There are many cases where eCourts does not provide a clear indication of whether the outcome is desirable. For instance, a case outcome may be described in the metadata simply as disposed, with no additional judgment information uploaded for the case. For cases like these, we de ne the outcome as neither acquitted nor convicted that is, the positive outcome variable takes the value of 0 when Y = acquitted, and the value of 1 when Y = not convicted (Appendix Table A3). 27% of case dispositions can be clearly designated as good or bad, with the remainder ambiguous; we show that our results are robust when we restrict the sample to cases with unambiguous outcomes and that ambiguity is not in itself a ected by in-group bias.

Judicial delay is also a major policy issue in India, so getting a decision at all is therefore itself an outcome of interest. We de ne an outcome indicator for whether a decision is made on a case within six months of the case's ling date; about 30% of cases are decided in this time horizon.

3.5 Summary Statistics on Case Outcomes

Figure 1 presents descriptive statistics of charges and convictions by gender and religious identity of defendants, respectively.[25] 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 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 ve times more likely to be charged with crimes under any classi cation. Panel B shows that the acquittal rate varies by crime, but overall it is about 3 percentage point higher for women (the Total category, at the bottom).

Panel C shows that Muslims are over-represented by 3% in the universe of criminal charges. Representation changes substantially depending on the change: relative to their population share, Muslims are 36% more likely to be charged with crimes against

[25] The corresponding point estimates are reported in Appendix Tables A4 and A5.

Table 1: Summary statistics, by judge identity

		Judge gender		Judge religion	
	(1)	(2)	(3)	(4)	(5)
	Total	Female	Male	Muslim	Non-Muslim
Female judge	0.270		0.000	0.257	0.267
	(0.002)		(0.000)	(0.010)	(0.003)
Muslim judge	0.068	0.066	0.069		0.000
	(0.001)	(0.003)	(0.002)		(0.000)
Tenure length (Days)	520.765	532.378	524.671	528.661	524.180
	(2.501)	(5.128)	(2.995)	(10.226)	(2.607)
Decisions					
Decision (given rst ling)	0.308	0.302	0.304	0.306	0.309
	(0.002)	(0.004)	(0.002)	(0.007)	(0.002)
Acquitted	0.177	0.181	0.180	0.184	0.177
	(0.002)	(0.003)	(0.002)	(0.006)	(0.002)
Convicted	0.055	0.067	0.049	0.061	0.054
	(0.001)	(0.002)	(0.001)	(0.004)	(0.001)
N	33,332	8,085	22,802	2,024	30,252

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

women, 37% more likely to be charged with robbery, and 62% more likely to be charged with marriage o enses, but 5% less likely to face charges for murder. Panel D shows that aggregate di erences in acquittal rates between Muslims and non-Muslims are small.

Table 1 shows descriptive statistics of judges and case outcomes in the analysis sample. About 27% of judges are female and 6.8% of judges are Muslim. On average, Muslim and female judges have similar conviction and rapid decision rates to non-Muslim and male judges. Appendix Figure A4 maps the geographic distribution of our sample of courts, which covers the whole country.

4 Empirical Strategy

Our objective is to estimate whether defendants experience di erent outcomes depending on the identity of the judge presiding over their case. To estimate a causal e ect of judge identity, we need to e ectively control for any factors other than defendant identity that could a ect both judge identity and the case outcome.

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 di erence between female and male acquittal rates for each type of crime. Panel D shows the same di erence between Muslims and non-Muslims. Crimes are ordered by maximal punishment, from most to least severe.

We rely on the exogenous assignment of judges to cases, which produces as-good-as-random assignment of defendants to judges, conditional on charge and district. We formalize our empirical approach in the following subsection. For ease of exposition, we describe the empirical strategy investigating gender bias the speci cation and considerations for estimating religious identity bias are identical. Speci cations used in additional analysis on bias in contexts likely to activate identity are described with the results.[26]

4.1 Random Assignment of Judges to Cases

As with much of the prior empirical literature, judge assignment in district courts is as good as random, conditional on court-time and charge xed e ects, given rules that gives defendants and prosecutors virtually no control over which judge oversees the case (see Section 2). Random assignment of judges to cases addresses the concern that judges with di erent identities are assigned to di erent kinds of cases. 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 di erent crime distributions (with di erent acquittal rates), we might again mistakenly attribute those di erences to in-group bias.

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 di erent identities. In practice, the Indian court system runs this experiment whenever a defendant is charged in a jurisdiction with multiple judges of di erent identities on the bench. Even if there is bias at other stages of the criminal process (e.g. who gets charged), that would not undermine our identi cation strategy given the random assignment of judges.

We use a canonical regression approach to test for the e ect of judge identity on case outcomes, as used by Shayo and Zussman's (2011) analysis of judicial in-group bias in Israel. We model outcome Y_i (e.g. 1=acquitted) for case i with charge s, led

[26] We also explored an event study speci cation 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.

in court c at time t as:

$$Y_i = {}_1\text{judgeMale}_i + {}_2\text{defMale}_i + {}_3\text{judgeMale}_i \quad \text{defMale}_i + {}_{ct(i)} + {}_{s(i)} + X_i + {}_i \tag{1}$$

$$Y_i = {}_1\text{judgeNonMuslim}_i + {}_2\text{defNonMuslim}_i + {}_3\text{judgeNonMuslim}_i \quad \text{defNonMuslim}_i + {}_{ct(i)} + {}_{s(i)} + X_i + {}_i \tag{2}$$

where judgeMale and judgeNonMuslim are binary variables that indicate whether a judge is male or non-Muslim, respectively. Similarly, defMale and defNonMuslim indicate the defendant's identity. ${}_{ct(i)}$ is a court-month or court-year xed e ect, and ${}_{s(i)}$ is an act and section xed e ect. X_i includes controls for defendant religion, judge religion, and an interaction term of judge gender and defendant religion in the gender analysis. In the religion analysis, X_i represents controls for defendant gender, judge gender, and an interaction term of judge religion and defendant gender.

The charge section xed e ect ensures that we are comparing defendants charged with similar crimes. The court-time xed e ect ensures that we are comparing defendants who are being charged in the same court at the same time. Our primary speci cation uses a court-month xed e ect, while a secondary speci cation uses a court-year xed e ect. The court-year xed e ect allows a much 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 lings to match these absences, nor do we nd evidence of disproportionate identity matching in balance tests of either speci cation below. Court-time periods with no variation in judge identity are retained to increase the precision of xed e ects and controls, but they do not directly a ect the coe cients of interest. We also test a speci cation with judge xed e ects, which controls for the average acquittal behavior of each individual judge.[27] Standard errors are clustered at the judge level, since judge assignment is the level of randomization.

There are three causal e ects of interest. ${}_1$ describes the causal e ect on a female defendant of having a male judge assigned to her case rather than a female judge. ${}_1 + {}_3$ describes the causal e ect on a male defendant of having a male judge assigned to his case. The di erence between these e ects (${}_3$) is the own-gender bias it tells us whether individuals receive better outcomes when a judge matching their gender

[27] This speci cation is included for completeness, but is unnecessary for identi cation (as are the judge and defendant demographic controls) if judges are indeed e ectively assigned randomly.

identity is randomly assigned to their case. Since all three causal e ects are of interest, we report coe cients for each in the regression tables. The coe cient meanings are analogous in Equation 2.

About half the time, a case stays in the courts long enough such that the judge making the nal decision is di erent 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 speci cation because a rapid decision is itself an outcome. Further, even if the ling judge does not make the nal ruling on a case, they can make key decisions on the case process that in uence 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 identical e ects if we limit the sample to cases decided by the initially assigned judge.

A more subtle identi cation issue arises with our framing of these matching-gender and matching-religion e ects as capturing "in-group bias." This framing follows the prior empirical literature, where in-group bias describes the 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. (2019) and Canay et al. (2020), is that defendants from di erent identity groups share more characteristics than just their identity, most of which are unobserved. Further, judges from di erent identity groups might have correlated preferences or biases across those characteristics. For example, female defendants might tend to have children, and female judges might tend to be lenient for defendants with children. Our empirical approach would frame this as in-group bias. 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 rst step in addressing 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.

Table 2: Balance test for assignment of judge identity

	(1) Female judge	(2) Female judge	(3) Muslim judge	(4) Muslim judge
Female defendant	-0.000	-0.000	0.001	0.001
	(0.001)	(0.001)	(0.000)	(0.000)
Muslim defendant	0.001	0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	5155404	5168610	5240281	5253483
Fixed E ect	Court-month	Court-year	Court-month	Court-year

Standard errors in parentheses

$p < 0:10$, $p < 0:05$, $p < 0:01$

Notes: This table reports results from a formal test of random assignment of judges to cases in the study sample. For speci cation details, see Equations 3 and 4. Columns 1 2 report the likelihood of being assigned to a female judge relative to a male judge using court-month, and court-year xed e ects. Columns 3 4 report the likelihood of being assigned to a Muslim judge relative to a non-Muslim judge using court-month, and court-year xed e ects. Charge section xed e ects have been used across all columns reported. Heteroskedasticity robust standard errors are reported below point estimates.

4.2 Balance Tests

To test the validity of the random assignment of cases to judges, we run the following empirical balance test in the analysis sample:

$$\text{judgeFemale}_i = {}_1\text{defFemale}_i + {}_2\text{defMuslim}_i + {}_{ct(i)} + {}_{s(i)} + X_i + {}_i \tag{3}$$

$$\text{judgeMuslim}_i = + {}_1\text{defMuslim}_i + {}_2\text{defFemale}_i + {}_{ct(i)} + {}_{s(i)} + X_i + {}_i, \tag{4}$$

with variables de ned as above. The coe cients of interest are $_1$ and $_1$, which respectively tell us whether female judges are more likely to adjudicate cases with female defendants, and whether Muslim judges are more likely to adjudicate cases with Muslim defendants.

Balance estimates are shown in Table 2. Male and female defendants are equally likely to be assigned to female judges. Similarly, Muslim and non-Muslim defendants are equally likely to be assigned to Muslim judges. These balance tests provide support

for our identi cation assumption of exogenous judge assignment.

5 Results

5.1 E ect of assignment to judge types

The rst two rows of Table 3 Panel A present the impact, for female and male defendants respectively, of being randomly assigned to a male judge these are $_1$ and $_1 + _3$ in Equation 1. The third row shows the di erence between these two coe cients ($_3$), which is the own-gender bias. The outcome variable is an indicator for defendant acquittal. Columns 1 3 show results using court-month xed e ects, while Columns 4 6 use court-year xed e ects. Within each set of three columns, the second column adds additional demographic controls, while the third column adds judge xed e ects.

Male judges consistently deliver fewer acquittals than female judges. The point estimate on this e ect is nearly identical for both male and female defendants across all speci cations. The own-gender bias estimate is a tight zero; the e ect estimates rule out even a very small in-group bias e ect of 0.6 percentage points with 95% con dence.[28] The coe cients are stable across di erent xed e ect speci cations, as is expected given the as-good-as-random assignment of judges to defendants.

Table 3 Panel B shows the e ect of ling judge gender on an indicator for case resolution within six months of being led. Cases assigned to male judges are resolved slightly more quickly, but this di erence is una ected by defendant gender; the in-group bias e ect is again a precise zero. In short, we do not nd substantial gender bias on any dimension.

Table 4 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. The e ect of judge religion on the acquittal rate is again a precise zero. The point estimate on in-group bias is never higher than 0.2 percentage points and the estimates rule out an own-religion bias of 0.6 percentage points with 95% con dence.[29] Religious in-group bias is also absent in the speed of

[28] Appendix Table A6 shows bias e ects on conviction rates; the estimates again are a tight zero. Appendix Table A7 shows estimates when we exclude closed cases for which we are unable to determine the outcome. We prefer the speci cation in Table 3, because the inability to determine an outcome is itself an outcome. We also nd no e ect of gender or religious match on whether the outcome is clearly coded as acquittal or conviction (Appendix Table A8).

[29] Appendix Tables A9 and A10 show results on conviction rates, and on acquittals with ambiguous

Table 3: Impact of assignment to a male judge on defendant outcomes

Outcome variable: Acquittal rate						
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	-0.008***	-0.007**		-0.007***	-0.007**	
	(0.003)	(0.003)		(0.003)	(0.003)	
Male judge on male defendant	-0.006***	-0.006**		-0.006***	-0.005**	
	(0.002)	(0.003)		(0.002)	(0.003)	
Di erence = Own gender bias	0.001	0.001	0.000	0.002	0.002	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Reference group mean	0.176	0.177	0.177	0.176	0.177	0.177
Observations	5223433	5129780	5128269	5236865	5143294	5141492
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge xed e ect	No	No	Yes	No	No	Yes
Fixed E ect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Outcome variable: Decision within six months of ling						
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	0.023***	0.022***		0.022***	0.021***	
	(0.004)	(0.004)		(0.004)	(0.004)	
Male judge on male defendant	0.022***	0.021***		0.021***	0.020***	
	(0.003)	(0.004)		(0.003)	(0.004)	
Di erence = Own gender bias	-0.001	-0.001	-0.002	-0.001	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Reference group mean	0.283	0.283	0.283	0.283	0.282	0.282
Observations	4367368	4286787	4285368	4380105	4299591	4297894
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge xed e ect	No	No	Yes	No	No	Yes
Fixed E ect	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 xed e ects have been used across all columns reported.
Speci cation: $Y_i = {}_1 \text{judgeMale}_i + {}_2 \text{defMale}_i + {}_3 \text{judgeMale}_i \quad \text{defMale}_i + {}_{ct(i)} + {}_{s(i)} + X_i + {}_i$

Table 4: Impact of assignment to a non-Muslim judge on defendant outcomes

Outcome variable: Acquittal rate						
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	0.008	0.008		0.007	0.006	
	(0.004)	(0.005)		(0.004)	(0.005)	
Non-Muslim judge on non-Muslim defendant	0.007**	0.007*		0.007**	0.006	
	(0.003)	(0.004)		(0.003)	(0.004)	
Di erence = Own religion bias	-0.001	0.000	0.002	-0.001	0.000	0.002
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
Reference group mean	0.18	0.184	0.184	0.181	0.184	0.184
Observations	5655320	5214531	5213019	5668388	5228040	5226225
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge xed e ect	No	No	Yes	No	No	Yes
Fixed E ect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

Outcome variable: Decision within six months of ling						
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	0.010	0.008		0.009	0.005	
	(0.006)	(0.008)		(0.006)	(0.008)	
Non-Muslim judge on non-Muslim defendant	0.005	0.002		0.003	-0.001	
	(0.006)	(0.007)		(0.005)	(0.007)	
Di erence = Own religion bias	-0.005	-0.006	0.002	-0.006	-0.006	0.003
	(0.004)	(0.004)	(0.003)	(0.004)	(0.005)	(0.003)
Reference group mean	0.291	0.287	0.287	0.29	0.287	0.287
Observations	4732429	4360514	4359090	4744851	4373321	4371607
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge xed e ect	No	No	Yes	No	No	Yes
Fixed E ect	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 xed e ects have been used across all columns reported.
Speci cation: $Y_i = {}_1\text{judgeNonMuslim}_i + {}_2\text{defNonMuslim}_i + {}_3\text{judgeNonMuslim}_i \quad \text{defNonMuslim}_i + {}_{ct(i)} + {}_{s(i)} + X_i + {}_i$

judicial decisions, nor is there any evidence that Muslim and non-Muslim judges have di erent rates of resolving cases (Table 4).

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. Some of the prior literature suggests that various identities can be made more salient by speci c 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 de ning 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 signi cant.

We rst examine the subset of cases where the victim and defendant have di erent identities. In these cases, when the defendant and judge are mismatched, the judge and victim will share the same gender or religious identity.[30] 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 5 interacts an indicator for defendant-victim gender mismatch with the gender in-group bias indicator. Both the baseline bias e ect and the interacted e ect are null; judges do not show gender in-group bias even when the defendant and victim have di erent 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 di erent religions.[31] Standard errors are larger due to the smaller sample and interaction speci cation, but the in-group bias e ect is less than 1 percentage point in both cases.

We next look at whether male and female judges rule di erently on cases classi ed in

results dropped. While we nd marginally signi cant bias e ects (in the in-group direction) in a handful of speci cations, the majority are statistically insigni cant, and the point estimate on the bias term is never higher than 0.5 percentage points. Appendix Table A11 shows there is no e ect of in-group bias on an indicator for an ambiguous case outcome.

[30] 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.

[31] 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 xed e ects. Appendix Tables A13 and A14 show all of the coe cients from the regression with court-month xed e ects. 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.

Table 5: In-group bias in contexts that activate identity

	(1) Gender	(2) Religion	(3) Gender	(4) Religion
Ingroup Bias	0.004	0.001	0.000	-0.004
	(0.003)	(0.005)	(0.002)	(0.002)
Ingroup Bias * Victim Gender mismatch	-0.006			
	(0.005)			
Ingroup Bias * Victim Religion mismatch		0.007		
		(0.008)		
Ingroup Bias * Crime against women			-0.009	
			(0.007)	
Ingroup Bias * Ramadan				0.019
				(0.010)
Observations	1787144	2018018	5123288	5179792
Fixed E ect	Court-month	Court-month	Court-month	Court-month
Judge Fixed E ect	Yes	Yes	Yes	Yes
Sample	All	All	All	All

Standard errors in parentheses

p < 0:10, p < 0:05, p < 0:01

Notes: This tests whether in-group bias appears in a set of contexts that may make identity particularly salient. The context tested in each column is (1) the defendant and victim have di erent religions; (2) the defendant and victim have di erent genders; (3) the case includes one or more charges considered crimes against women; and (4) the judgment takes place during the month of Ramadan. The type of bias considered is based on religion in Columns 1 and 3, and on gender in Columns 2 and 4. Charge section xed e ects have been used across all reported columns.

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.[32] Column 3 of Table 5 shows that the interaction between an indicator for crimes against women and the in-group bias variable is small and statistically insigni cant. Male defendants do not receive di erential treatment from male and female judges, even in these cases.

Finally, we examine whether religious in-group bias emerges during the month of Ramadan, when Muslim religious identity may become particularly salient for both Muslims and non-Muslims.[33] Column 4 of Table 5 shows that the interaction between the Ramadan indicator and the in-group bias measure is positive and marginally statistically signi cant ($p = 0.09$).[34] Religious in-group bias seems to be activated when religious identity is particularly salient, yet the e ect size remains small relative to other studies.[35] Appendix Table A12 shows robustness of the estimates to using court-year instead of court-month xed e ects.

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 like to run an equivalent statistical test, where judge and defendant identity sometimes match on the caste dimension and sometimes do not. An equivalent caste analysis to what we have done for gender and religion is not feasible, however, for three reasons. First, unlike gender and religion, there is no classi cation for caste along which in- and out-groups can be con dently and universally de ned. 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 insu cient in characterizing the a nities that people may feel

[32] One reason kidnappings are so common in the data is that this may be the formal charge led against a man who elopes with a woman. Results are similar for both the assault and kidnapping subsets of the data.

[33] Unlike the sample in Mehmood et al. (2021), 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.

[34] 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 e ect of Ramadan a ects the outcome on the day the decision is reached, rather than on the day the case rst appeared before a judge. See Section 4 for more on how we treat cases seen by more than one judge.

[35] Mehmood et al. (2021) nd in Pakistan that conviction rates are 23 percentage points lower during the month of Ramadan. The nal section further discusses the size of our estimates compared with other studies.

within the caste system. For example, an upper caste person could identify with another upper caste person despite sharing neither *varna* or *jati*. Likewise, the term *bahujan* is often used to describe the shared identity of marginalized groups such as Scheduled Castes and Other Backwards Castes. Second, individual names do not identify caste as precisely as they identify Islamic religion or gender identity and the caste signi cance of names can vary across regions. Due to these limitations and to a lack of training data, we have not been able to develop a reliable correspondence between names and speci c castes. Third, there are few district judges in the most identi able caste categories: Scheduled Castes and Scheduled Tribes.

For these reasons, a direct analysis of caste bias in the Indian judiciary is not feasible at this time. Instead, we analyze caste indirectly. Speci cally, we follow Fisman et al. (2017) and de ne 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.

The measure is admittedly imperfect. Names are more numerous than castes, so members of the same caste usually have di erent last names. Further, sharing names can indicate greater a nity 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 di erent castes.

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

$$Y_i = {}_1 \text{sameLastName}_i + {}_{ct(i)} + {}_{s(i)} + X_i + {}_i. \tag{5}$$

where subscripts i, s, c and t are de ned as above. The court-time (${}_{ct(i)}$) and act/section (${}_{s}(i)$) xed e ects, and judge/defendant characteristics X_i are also as above. Further, we include additional xed e ects 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.[36]

The identi cation assumptions for consistent estimation of $\hat{}_1$ are the same as in the prior section. If judges are randomly assigned to cases (within the court-time

[36] Without this limitation we have substantially more last name xed e ects in the sample but there is no additional variation in terms of identity match, because the sameLastNamevariable always takes the value 0 for defendants whose last name never appears in the judge list.

Table 6: E ect 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	-0.000	-0.001	0.014	0.012	0.001	-0.001
	(0.001)	(0.001)	(0.006)	(0.006)	(0.004)	(0.004)
Same name * Rare name					0.032	0.033
					(0.015)	(0.015)
Observations	2225312	2223403	2225312	2223403	2225312	2223403
Fixed E ect	Court-month	Court-month	Court-month	Court-month	Court-month	Court-month
Judge Fixed E ect	No	Yes	No	Yes	No	Yes
Inverse Group Weight	No	No	Yes	Yes	Yes	Yes
Last Name Fixed E ect	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 e ect of assignment to a judge with the same last name as the defendant on likelihood of acquittal (Equation 5). Court-month xed e ects, charge section xed e ects, and judge and defendant last name xed e ects have been used across all columns reported. Standard errors are clustered by judge.

randomization block), then cases where defendant and judge name match will occur at random. The act/section xed e ects adjust for judge assignment rules based on the seriousness of the crime. Finally, the last-name xed e ects adjust for the possibility that individuals from some social groups are more or less likely to be acquitted, and that judges in di erent social groups may have di erent average acquittal rates.

The results for last name bias are reported in Table 6. Columns 1 and 2 report unweighted estimates from Equation 5, comparable to the speci cations 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. The results are thus driven by individuals with common last names, like Kumar and Singh. These are the names where a defendant-judge last-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 speci cation 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 1.2 1.4 percentage points (statistically signi cant). This result suggests that there is caste-based in-group bias, driven by groups with less common names. To con rm this more directly, 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 speci cation. The uninteracted coef- cient shows an absence of bias for common last names, and the interacted coe cient shows a 3.2 3.3 percentage point in-group bias for individuals with uncommon last names.

The e ect size among individuals with uncommon names is economically relevant and statistically signi cant, representing about a 15% increase in the probability of acquittal.[38] The social proximity signalled by sharing a rare last name, often indicating a shared caste (Fisman et al., 2017), is associated with judicial in-group bias. Yet this bias is only seen for the relatively narrow social groups demarcated by less common names. By de nition, then, the same-name e ect 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, may be large in magnitude for some individuals, but will be small in aggregate if it operates only at the level of narrow social groups. Of course, we cannot rule out that judges may be exhibiting in-group bias on the basis of cultural similarity measures that we are not able to observe.

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, insu cient 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

[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 e ects are robust to looser de nitions of last name similarity (for example, treating Patil and Patel as similar).

[38] Results are similar if we de ne rare names based on frequency among judges rather than among defendants.

et al., 2003; La Porta et al., 2008). Yet with a few exceptions (Ponticelli and 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, 2010 2018. We estimate robust, tight zero e ects of judicial in-group bias along the dimensions of gender, religion, and caste. We do not nd gender-based bias even when gender identity is more salient, but we do nd religion-based bias in one of two subsamples where religion is more salient (during Ramadan). We also nd some in-group bias among social groups with shared uncommon last names. The aggregate e ects of the measured biases are small, but there is evidence that bias can be magni ed in circumstances which make the dimension of shared (or unshared) identity more salient.

The aggregate null e ects are surprising, especially given well-documented gender and religious in-group bias in non-judicial contexts in India. Two relevant examples are Fisman et al. (2017), who nd that credit o ers and repayment rates rise when loan o cers and clients have the same last name, and Neggers (2018), who nds that random assignment of a minority election worker to a polling station has a large pro-minority e ect on vote counts at that station. The divergent ndings raise the question of how these contexts di er from the judicial setting.

One major di erence is the judge's incentive structure. Judges expect little direct economic bene t or cost from seeing members of the out-group punished. That "game" is quite di erent 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 di erences 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[39]. Consistent with this interpretation, our results on matching last names suggest that in-group bias is stronger under more narrow de nitions of the in-group.

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

[39] Similarly, Sharan (2020) nds that ethnic quotas in local government only improve public service delivery when lower-status groups occupy multiple positions in the political hierarchy.

of teachers grading student exams. Like judging, grading is a non-adversarial context, where teachers face at incentives for how students are assessed. Further, there are impactful class and authority di erences between teachers and students, which make di erences due to caste less salient. From a theoretical perspective, then, our results echo those from Hanna and Linden (2012).

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 di erent settings. Further empirical research drilling down on these theories will be valuable.

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

Another notable pattern in the graph is that the con dence intervals (and hence standard errors) grow with the e ect sizes. A positive relationship between e ect 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 nding. To show this more directly, Figure 2 Panel B plots (in open black circles) the e ect size of each of the previous studies against the standard error of the main estimated e ect. For comparison, the estimates from our study are plotted as closed red circles. In the absence of publication bias or a design-based mechanical relationship between e ect 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 asym-

[40] When papers report multiple speci cations for the main e ect, we use the e ect size described most prominently in the text or described by the authors as the main speci cation. 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.

[41] See Egger et al., 1997; Gerber et al., 2001; Levine et al., 2009; Slavin and Smith, 2009; Kühberger

Figure 2: Comparison with judicial bias estimates in other contexts

A. Coe cient Plot

B. Standardized Errors vs. E ect Sizes

Notes: This gure shows point estimates of in-group bias from other studies in the relevant literature. From the top, the coe cients 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 (2018), 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 e ects (Y axis) against e ect standard errors. All e ect sizes are standardized (dividing outcome variables by their standard deviation) to allow comparison across studies. From each table in this paper, we chose the speci cation with court-month and judge xed e ects. For contexts magnifying bias, we show the average e ect for the group facing magni ed bias. For example, for the Ramadan analysis, we show the sum of the bias coe cient and the bias * Ramadan coe cient, which describes religious in-group bias in the month of Ramadan. The only statistically signi cant estimate at the 95% level is the inverse group size weighted interaction between same name and rare last name (Table 6 Column 6); note the unweighted regression (which weights each case equally) found a zero estimate.

Table 7: Estimates of Publication Bias in Judicial In-Group Bias Studies

	(1)	(2)	(3)	(4)	(5)
		p(z)=Pr(Pub|t-stat)			
	$(-\infty; -1.96]$	$(-1.96; 0]$	$(0; 1.96]$	$(1.96; +\infty)$	
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. θ 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).

metric, 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 learned function then provides a structural estimate, based on the existing published papers, for the likelihood of publication for a given t-stat z. The method also provides an adjusted effect size based on imputing unpublished studies.

Table 7 reports the result of the test for publication bias. According to the learned publication function, studies with negative estimates (Columns 1 and 2) and statistically insignificant positive estimates (Column 3) 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 results. The estimates from the prior literature are thus consistent with severe publication bias. When adjusting for publication bias by imputing missing studies, the predicted true effect size is 0.046 (Column 5), a fraction of the average observed effect size of 0.24 from the published studies.

When interpreting these publication-bias results, it is important to remember that

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.

in-group bias di ers across contexts. Indeed, our own evidence shows substantial variation in in-group bias across social conditions and across social groups in India. In turn, several of the studies included in Table 7 focus on contexts where identity is salient and in-group bias is likely to be activated. Hence, our statistics do not imply that the published studies are wrong, but rather, that a collection of studies with smaller or null e ects have remained in the le drawer.[42]

The rest of the literature aside, our nding of a lack of in-group bias in India's lower courts should be celebrated, not least because it can inform policymakers allocating resources to address the clear and extant social disparities in Indian society. Yet our research 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 le cases, the severity of charges applied, the severity of penalties imposed, the appeals process, civil litigation, or via other factors. There could also be absolute bias, where both in- and out-group judges discriminate against out-groups. Based on our evidence, concerns about in-group bias might be better directed to other parts of the justice pipeline 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. Other papers that nd null or reverse e ects of in-group bias tend to focus on di erent aspects of their contexts and put little emphasis on the null in-group e ects (Arnold et al., 2018; Hanna and Linden, 2012).

References

Abrams, D. S., Bertrand, M., and Mullainathan, S. (2012). Do Judges Vary in Their Treatment of Race? The Journal of Legal Studies, 41(2):347 383.

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

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

Andrews, I. and Kasy, M. (2019). Identi cation of and correction for publication bias. American Economic Review, 109(8).

Aney, M. S., Dam, S., and Ko, G. (2017). Jobs for Justice(s): Corruption in the Supreme Court of India. Working Paper.

Anwar, S., Bayer, P., and Hjalmarsson, R. (2012). The Impact of Jury Race in Criminal Trials. The Quarterly Journal of Economics, 127(2):1 39.

Anwar, S., Bayer, P., and 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.

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

Arnold, D., Dobbie, W., and Yang, C. S. (2018). Racial Bias in Bail Decisions. The Quarterly Journal of Economics, 133(4):1885 1932.

Asher, S., Novosad, P., and Rafkin, C. (2020). Intergenerational Mobility in India: New Methods and Estimates Across Time, Space, and Communities. Working Paper.

Baldus, D. C., Woodworth, G., Zuckerman, D., and 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.

Bank, W. (2017). World Development Report 2017: Governance and the Law.

Baumgartner, F. R., Grigg, A. J., and Mastro, A. (2015). #BlackLivesDon'tMatter: Race-Of-Victim E ects in US Executions, 1976 2013. Politics, Groups, and Identities, 3(2):209 221.

Bertrand, M., Hanna, R., and Mullainathan, S. (2010). A rmative Action in Education: Evidence from Engineering College Admissions in India. Journal of Public Economics, 94(1-2):16 29.

Bharti, N. and Roy, S. (2020). The Early Origins of Judicial Bias in Bail Decisions: Evidence from Early Childhood Exposure to Hindu-Muslim Riots in India. Working Paper.

Borker, G. (2017). Safety First: Perceived Risk of Street Harassment and Educational Choices of Women. pages 12 45. Working Paper.

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

Canay, I. A., Mogstad, M., and Mountjoy, J. (2020). On the use of outcome tests for detecting bias in decision making.

Chaturvedi, R. and Chaturvedi, S. (2020). It's All in the Name: A Character Based Approach To Infer Religion. arXiv preprint arXiv:2010.14479.

Chemin, M. (2009). Do Judiciaries Matter for Development? Evidence from India. Journal of Comparative Economics, 37(2):230 250.

Chew, P. K. and Kelley, R. E. (2008). Myth of the Color-Blind Judge: An Empirical Analysis of Racial Harassment Cases. Wash. UL Rev, 86:1117.

Cousins, S. (2020). 2.5 Million More Child Marriages due to COVID-19 Pandemic. The Lancet, 396(10257):1059.

Depew, B., Eren, O., and Mocan, N. (2017). Judges, Juveniles, and In-Group Bias. The Journal of Law and Economics, 60(2):209 239.

Dev, A. (2019). Corruption Has India's Supreme Court Veering on the Edge. The Atlantic.

Didwania, S. H. (2018). Gender-Based Favoritism Among Criminal Prosecutors. Working Paper.

Djankov, S., La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2003). Courts. Quarterly Journal of Economics, 118(2):453 517.

Egger, M., Smith, G. D., Schneider, M., and 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., and et al. (2020). State of the World Population 2020. United Nations Population Fund.

Fisman, R., Paravisini, D., and Vig, V. (2017). Cultural Proximity and Loan Outcomes. American Economic Review, 107(2):457 92.

Fisman, R., Sarkar, A., Skrastins, J., and Vig, V. (2020). Experience of Communal Con icts and Intergroup Lending. Journal of Political Economy, 128(9):3346 3375.

ForsterLee, R., ForsterLee, L., Horowitz, I. A., and King, E. (2006). The E ects of Defendant Race, Victim Race, and Juror Gender on Evidence Processing in a Murder Trial. Behavioral Sciences & the Law, 24(2):179 198.

Frandsen, B. R., Lefgren, L. J., and Leslie, E. C. (2019). Judging judge xed e ects.

Gazal-Ayal, O. and 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.

Gerber, A. S., Green, D. P., and Nickerson, D. (2001). Testing for Publication Bias in Political Science. Political Analysis.

Goel, S., Rao, J. M., Shro , 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.

Grossman, G., Gazal-Ayal, O., Pimentel, S. D., and Weinstein, J. M. (2016). Descriptive Representation and Judicial Outcomes in Multiethnic Societies. American Journal of Political Science, 60(1):44 69.

Hanna, R. N. and Linden, L. L. (2012). Discrimination in Grading. American Economic Journal: Economic Policy, 4(4):146 68.

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.

Jayachandran, S. (2015). The Roots of Gender Inequality in Developing Countries. Annual Review of Economics, 7.

Kastellec, J. P. (2011). Panel Composition and Voting on the US Courts of Appeals over Time. Political Research Quarterly, 64(2):377 391.

Kastellec, J. P. (2013). Racial Diversity and Judicial In uence on Appellate Courts. American Journal of Political Science, 57(1):167 183.

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.

Kühberger, A., Fritz, A., and Scherndl, T. (2014). Publication Bias in Psychology: A Diagnosis Based on the Correlation Between E ect Size and Sample Size. PloS one, 9(9):e105825.

La Porta, R., Lopez-de Silanes, F., Pop-Eleches, C., and Shleifer, A. (2004). Judicial Checks and Balances. Journal of Political Economy, 112(2).

La Porta, R., Lopez-de Silanes, F., and Shleifer, A. (2008). The Economic Consequences of Legal Origins. Journal of Economics Literature, 46(2):285 332.

Le, Q. V. (2004). Political and Economic Determinants of Private Investment. Journal of International Development, 16(4):589 604.

Levine, T. R., Asada, K. J., and Carpenter, C. (2009). Sample Sizes and E ect Sizes are Negatively Correlated in Meta-Analyses: Evidence and Implications of a Publication Bias Against Nonsigni cant Findings. Communication Monographs, 76(3):286 302.

Lichand, G. and Soares, R. R. (2014). Access to Justice and Entrepreneurship: Evidence from Brazil's Special Civil Tribunals. The Journal of Law and Economics, 57(2).

Lim, C. S., Silveira, B. S., and 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.

Mehmood, S., Seror, A., and Daniel, C. (2021). Ramadan Spirit and Criminal Acquittals: Causal Evidence from Pakistan. Working Paper.

Mullen, B., Brown, R., and Smith, C. (1992). Ingroup Bias as a Function of Salience, Relevance, and Status: An Integration. European Journal of Social Psychology, 22(2):103 122.

Mustard, D. B. (2001). Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the US Federal Courts. The Journal of Law and Economics, 44(1):285 314.

Neggers, Y. (2018). Enfranchising your own? experimental evidence on bureaucrat diversity and election bias in india. American Economic Review, 108(6):1288 1321.

Pande, R. and Udry, C. (2005). Institutions and Development: A View from Below. Yale University Economic Growth Center Discussion Paper, (928).

Peresie, J. L. (2005). Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts. The Yale Law Journal, 114(7):1759 1790.

Ponticelli, J. and 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.

Rao, M. (2019). Judges, Lenders, and the Bottom Line: Courting Firm Growth in India. Working Paper.

Rehavi, M. M. and Starr, S. B. (2014). Racial Disparity in Federal Criminal Sentences. Journal of Political Economy, 122(6):1320 1354.

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.

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. Technical report, Government of India.

Sandefur, J. and 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.

Sharan, M. (2020). It's complicated: The distributional consequences of political reservation. Working paper.

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

Shayo, M. and Zussman, A. (2017). Con ict and the persistence of ethnic bias. American Economic Journal: Applied Economics, 9(4).

Slavin, R. and 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.

Sloane, C. (2019). Racial Bias by Prosecutors: Evidence from Random Assignment. Working paper.

Times of India (2018). Data: OBCs just 12% of lower court judges. January 29, 2018.

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

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.

A Appendix

Figure A1: India eCourts Case Record Sample

Notes: The gure displays an anonymized version of a sample court record from https://ecourts.gov.in/ for the District and Sessions Court of Vidisha. The `Petitioner and Advocate' and `Respondent and Advocate' sections contain the litigant names that we use for assigning gender and religion. The `Acts' section contains the data that allows us to discriminate between civil and criminal cases. We use the `Under Section(s)' column to infer the corresponding crime categories.