

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*

August 5, 2021

Abstract

We study judicial in-group bias in Indian criminal courts, collecting data on over 5 million criminal case records from 2010–2018. 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. We estimate tight zero effects of in-group bias along gender and religious identity. We do find limited in-group bias in some (but not all) settings where identity is particularly salient, but even here our confidence intervals reject effect sizes smaller than those in much of the prior literature.

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, University of 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, 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) and school-teachers (Hanna and Linden, 2012). The judicial setting is of particular interest, given the premise that individuals who are discriminated against in informal settings can find recourse from equal treatment under the law (Sandefur and Siddiqi, 2015).

We focus on the dimensions of gender and religion in India’s lower courts, where 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 lower court judges. There is growing evidence that India’s Muslims and women 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). We examine whether unequal representation in the courts has a direct effect on the judicial outcomes of Muslims and women, in the form of judges delivering better outcomes to criminal defendants who match their gender or religion.

Our analysis draws upon a new dataset of 77 million court records covering 2010–2018 from *eCourts*, an online platform documenting the complete set of cases heard in India’s district courts.² 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

¹See, 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.

²eCourts can be accessed at <https://ecourts.gov.in/>.

version of the dataset, opening the door to many new analyses of the judicial process in the world’s largest democracy and largest common-law legal system.³

We enrich the dataset by classifying judges and defendants to gender and religious (Muslim and non-Muslim) identity groups. The group assignment is achieved by a deep neural network applied to the sequence of characters in the names of each judge and defendant. The distinctive nature of female and Muslim names allows us to classify individuals with over 97% out-of-sample accuracy on both dimensions.⁴

The main research question is whether judges tend to treat defendants differently when they share the same gender or religion. We focus on the subset of cases filed under India’s criminal codes, where acquittal and conviction rates can be interpreted as positive and negative outcomes respectively. Given the extreme delays in India’s judicial system (Trusts, 2019; Rao, 2019), we additionally examine whether in-group judge identity affects the court’s speed in reaching a decision.

We exploit the arbitrary rules by which cases are assigned to judges, generating as-good-as-random variation in judge identity. Our preferred specification includes court, charge, and month-year fixed effects. Effectively, we compare 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.⁵

We find a robust null estimate of in-group bias among Indian judges. Judges of different genders do not treat defendants differently according to defendant gender, nor do judges display favoritism on the basis of religion. This null is seen both in terms of outcomes (i.e. acquittals and convictions) and in terms of process (i.e. speed of decision). Our 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.⁶ 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.

³The data can be accessed at <https://www.devdataalab.org/judicial-data>.

⁴We have limited ability to examine bias on the dimensions of income or caste because we do not yet have an algorithm that can classify these dimensions with high accuracy.

⁵Results are robust to adding judge fixed effects (which control for variation in the severity of specific judges), though these are not expected to have an effect under random assignment of cases to judges.

⁶The 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.

Our analysis largely excludes questions of caste, which remains a major cleavage in Indian society. Caste identity is more complex than religious or gender identity, making it more difficult to identify clear in-groups and out-groups. It is also more difficult to classify an individual’s caste on the basis of their name. To make some progress on caste bias, we define the in-group as the set of defendants whose last name matches the judge’s last name. This is admittedly an imperfect measure because multiple family names may reflect the same caste and certain last names (e.g. *Kumar*) 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. Using the same last name to classify identity groups has predicted affinity in previous work, for instance in the banking setting (Fisman et al., 2017). We find a small positive in-group bias from caste identity: defendants assigned to judges with the same last name are 2 percentage points more likely to be acquitted. The effect remains small in comparison with other studies of bias.

Our estimates do not rule out bias in the Indian legal system entirely. We observe only a subset of the legal process and we measure only in-group bias by gender and religion. For example, it is possible that both Muslim and non-Muslim judges discriminate against Muslims (as found for Black defendants in Arnold et al. (2018)). It is also possible that arrests and/or charges disproportionately target Muslims, or that judges exhibit bias based on defendant caste or income. In any case, this evidence could be useful for policymakers in India when deciding where to allocate resources in addressing discrimination and social disadvantage. There are likely other parts of the legal system, besides the acquittal choices of judges, where efforts would be more beneficial.

Notwithstanding a null effect of judge in-group bias on average, it is still possible that bias is activated in certain subsets of cases, for example where judge and defendant identity are more salient. We examine three special contexts where the literature suggests that in-group bias may be more likely to be activated (Mullen et al., 1992; Shayo and Zussman, 2011; Anwar et al., 2012; Mehmood and Seror, 2020). First, we examine cases where the defendant and the victim of the crime have different identities; in these cases, the judge has an identity matching either the victim or the defendant, but not both. Second, we examine gender bias in criminal cases categorized as crimes against women, which are mostly sexual assaults and kidnappings. In both of these subset analyses, we continue to find a null bias.

Third and finally, we examine whether in-group bias on the basis of religion is activated during the month of Ramadan, when religion may become more salient for

both Muslims and non-Muslims. In this analysis, we find suggestive evidence that religious in-group bias is activated: During the month of Ramadan, Muslim defendants assigned to Muslim judges have a 1.5–2.0 percentage point higher acquittal rate relative to being assigned to a non-Muslim judge. Still, the estimate is small in magnitude and only marginally statistically significant due to the smaller sample during the Ramadan months.⁷ This last 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.

In Section 6, we discuss several potential reasons that bias could be lower in the Indian setting than in other judicial settings. One explanation that stands out is that there could be a publication bias in the bias literature itself, such that contexts without in-group bias are not widely described in the literature. A funnel plot of point estimates from in-group bias papers relying on exogenous judge assignment (Figure 2 Panel B) is certainly consistent with publication bias, with many estimates lying just inside the zone of statistical significance, regardless of their power. This finding does not imply that these studies are wrong, but that some studies with null results may have remained in the file drawer or have not been published for other reasons.

Relative to the prior literature, we make several contributions. First, we demonstrate an absence of bias in an important context with substantial religious and gender cleavages. Second, the sample of our study is an order of magnitude larger than earlier studies, allowing us to measure bias much more precisely than prior work. Third, to our knowledge, this is the first large-scale study of judicial bias in a low- or middle-income country and it makes available a dataset which may have substantial utility to future scholars.

These results add to a 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.⁸ The closest paper to ours is [Shayo and Zussman \(2011\)](#), who analyze

⁷This point estimate remains small compared with much of the prior literature on in-group bias. In particular, the estimate is considerably smaller than the one in [Mehmood and Seror \(2020\)](#), who find in Pakistan that conviction rates fall by 14 percentage points during Ramadan, or 1 percentage point for each additional hour of fasting. In their study, nearly all judges and defendant are Muslim, so the effect of identity plays a different role than in ours, since we exploit difference in Muslim and non-Muslim judges during Ramadan. Note that we do not exploit differences in daylight hours because there is little variation in the timing of Ramadan across the 8 years in our study.

⁸These 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](#)),

the effect of assigning a Jewish versus an Arab judge in Israeli small claims court. They find robust evidence of in-group bias, where Jewish judges favor Jewish defendants (and Arab judges favor Arab defendants). Our finding that religious bias is magnified during the month of Ramadan is consistent with their notion of endogenous social identification, though our point estimates on bias are an order of magnitude lower than theirs even in the context where identity appears to be most salient.

Several more studies use identification strategies to generate point estimates that are directly comparable to ours. Of these, only [Lim et al. \(2016\)](#) find a null in-group effect of judge ethnicity or gender. The other directly comparable papers find large and statistically significant in-group bias.⁹

In the Indian context, there is a growing body of evidence on the legal system, mostly focusing on judicial efficacy and economic performance ([Chemin, 2009](#); [Rao, 2019](#)), or on corruption in the Indian Supreme Court ([Aney et al., 2017](#)). A recent working paper finds that judges are more prone to deny bail if they were exposed to communal riots in their early childhood ([Bharti and Roy, 2020](#)). However, we are aware of no prior large-scale empirical research on unequal legal treatment on either the gender or religion dimension 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 effective, inclusive institutions, with judicial equity and rule of law seen as key indicators of a country’s institutional quality ([Rodrik, 2000](#); [Le, 2004](#); [Rodrik, 2005](#); [Pande](#)

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 differently from Caucasian-American judges on issues where minorities are disproportionately affected, such as affirmative action, racial harassment, unions, and search and seizure cases ([Scherer, 2004](#); [Chew and Kelley, 2008](#); [Kastellec, 2011](#)). In a similar manner, a number of papers have documented the effect of judges’ gender in sexual harassment cases ([Boyd et al., 2010](#); [Peresie, 2005](#)). A smaller set of papers use information on both the identity of the defendant and the decision-maker. [Anwar et al. \(2012\)](#) look at random variation in the jury pool and find that having a black juror in the pool decreases conviction rates for black defendants. A similar result from Israel is documented by [Grossman et al. \(2016\)](#), who find that the effect of including even one Arab judge on the decision-making panel substantially influences trial outcomes of Arab defendants. [Didwania \(2018\)](#) find in-group bias in that prosecutors charge same-gender defendants with less severe offenses.

⁹[Gazal-Ayal and Sulitzeanu-Kenan \(2010\)](#) find positive in-group bias in bail decisions when Arab and Jewish defendants are randomly assigned to a judge of the same ethnicity. [Knepper \(2018\)](#) and [Sloane \(2019\)](#) leverage random assignment of cases in the U.S. to judges and prosecutors respectively, finding significant in-group bias in trial outcomes. [Depew et al. \(2017\)](#) exploit random assignment of judges to juvenile crimes in Louisiana and find *negative* in-group bias in sentence lengths and likelihood of being placed in custody. It is notable that of all these studies, [Lim et al. \(2016\)](#) has one of the largest sample sizes (N=250,000).

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 different 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.

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. 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.¹⁰

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

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

There is an active debate in India around reforming the court system. Problems under discussion include a reputation for corruption ([Dev, 2019](#)) 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 effort 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 practice in India’s district courts, senior research fellows at the Vidhi Center for Legal Policy, and several clerks in courts around the country.

Criminal cases are assigned to judges as follows. First, a crime is reported at a particular local police station, where a First Information Report (FIR) is filed. Each

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

police station lies within the territorial jurisdiction of a specific 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 specific courtroom in a court for several months at a time. A courtroom is assigned for every police station and every charge. For example, at a given police station, every murder charge will go to the same courtroom. A larceny charge might go to a different courtroom, as might a murder charge reported at a different police station. Judges typically spend two to three years in a given court, during which they rotate through several of the courtrooms.¹¹

The police station charge lists leave little room for discretion over which charges are seen by which judges. Since the timing of the first court appearance is unknown when charges are filed (given judicial delays), even if a defendant or prosecutor had discretion over which police station filed the charges, the rotation of judges between courtrooms would make it difficult to target a specific judge. Finally, the judiciary explicitly condemns the practice of “judge shopping” or “forum shopping”, where litigants select particular judges in search of a favorable outcome. 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.

Finally, it should be noted that since 2013, some courts have adopted a random assignment lottery mechanism implemented through the eCourts platform. The eCourts assignment mechanism was intended to be used throughout the country, but in practice, it has not been widely adopted to date. In Section 4, we present formal tests of the exogenous assignment of judges to cases in our dataset.

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 0.5% of all criminal cases pending in India are disposed through plea bargaining. It is

¹¹Severe cases (with severity defined by the section or act under which the charge was filed) require judges with higher levels of seniority. Thus, a case in a given district may be eligible to be seen only by a subset of judges in that district.

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 as a “national data warehouse for case data including the orders/judgments for courts across the country.”¹² The publicly available information includes the filing, 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 filed, and the final decision or disposition.¹³

The database covers India’s lower judiciary, consisting of all courts including and under the jurisdiction of District and Sessions courts. This paper focuses on cases filed either under the Indian Penal Code or the Code of Criminal Procedure, for two reasons. First, there is only a single litigant, rather than two, providing a clear definition of identity match between judge and defendant. Second, it is relatively straightforward to identify good and bad outcomes for criminal defendants, and it is much more difficult to do so for litigants in civil cases. This constraint filters 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.¹⁴

We joined the case-level data with the judge-level data based on the judge’s designation and the initial case filing date. In this process, another 17% of the initial observations are dropped. The remaining dataset where cases are linked to a unique judge consists of 10 million cases. From this subset, we drop all bail decisions since they are a narrow share of the data (see Appendix Figure [A2](#)).

¹²https://ecourts.gov.in/ecourts_home/static/about-us.php, accessed Oct 14, 2020

¹³We illustrate such a record in Appendix Figure [A1](#).

¹⁴See 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.

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 classifier on a large database of labeled names and then use it to assign these characteristics in the legal data.¹⁵

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 A1. For gender, we observe two categories: female or male. For religion, we observe five categories: Hindu, Muslim, Christian, Buddhist, and Other. Our classifier takes a two-label specification: 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.

Before applying the classifier, we pre-process the name strings by transliterating characters from Hindi to Latin and normalizing capitalization, punctuation, and spacing. We then apply a neural net classifier to predict the identity label based on the name string, similar to the approach in Chaturvedi and Chaturvedi (2020). We use a bidirectional Long Short-Term Memory (LSTM) model applied directly to the sequence of name-string characters. LSTMs are a gated recurrent neural network architecture that takes as input a sequential data stream 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 classifier reads the sequence backward and forward when trying to assign a label.¹⁶

¹⁵The existing available name classifiers 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 classifier.

¹⁶In 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 finally to the output layer, which uses sigmoid activation to output a probability across the binary classes. To avoid overfitting, we used dropout between layers and used early stopping during training,

The ability of the LSTM classifier 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 meaning of the name in terms of religion. Given these nuances, the LSTM classifier 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 classifiers for gender and religion. The classifiers perform well on the standard metrics, including our preferred metrics that adjust for imbalance in the class shares. We report balanced accuracy, the average accuracy (recall) for each of the two identity categories, and F1, the harmonic mean of precision and recall.¹⁷ For gender, the balanced accuracy is .975 with $F1 = .976$. For religion, the balanced accuracy is .98 and the $F1 = .99$.

We then apply the trained classifier to the eCourts case records. The judge names tend to be complete (first and last name) and often include salutations indicating gender. Our algorithm can classify the names of 96% of the 81,232 judges (22,413 unique names) appearing in the case dataset according to gender (female/male) and 98% according to religion (Muslim/non-Muslim). The information on litigant names is of lower quality, often missing either the first name or last name. We are able to classify 80% of litigants by religion and 74% by gender. Cases with unclassified labels are dropped from analyses requiring those labels. For the analysis sample, we have 6 million cases for analyzing judge/defendant gender, and 6.6 million cases for analyzing judge/defendant religion (see Appendix Figure A2).

which ceases network training when validation loss stops improving. To account for the imbalance in the sample, we used class weights during the training.

¹⁷Balanced 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.

To verify the accuracy of the LSTM classification within the new domain of the court records, we conduct a manual verification of random subsets of names classified by gender and religion, stratified across all states. We confirm an accuracy rate of 97% for both the gender and religion classification based on manual verification. As an additional validation step, we compare the LSTM-classified Muslim defendant share by state to the state-wise Muslim population shares reported in the 2011 Population Census, and find a correlation of 0.88.

3.4 Defining Case Outcomes

We define the defendant’s outcome (represented by Y below) as a case-level indicator variable that takes the value one if the outcome is desirable for the defendant and zero otherwise. Our primary specification uses an indicator for defendant acquittal. A secondary specification 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 a case like this, we define 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 $Y=not\ convicted$ (Table A2). About 40% of case dispositions can be clearly designated as good or bad, while 60% are ambiguous; we show that our results are robust when we restrict the sample to cases with unambiguous outcomes.

In about 40% of cases, the judge presiding over the initial case filing does not reach a decision – a decision is reached by a future judge or else the case remains undecided. Our analysis is focused entirely on the first judge to see the case. Because decision deferral may be endogenous, we cannot treat the assignment of the second judge as random. Judicial delay is also a major policy issue in India, so getting a decision at all is therefore itself an outcome of interest. We define an indicator for whether a decision is made on a case within six months of the case’s filing date. We discuss our treatment of cases that are not resolved before they pass to other judges in Section 4.

Table 1: Coding of outcome variables

Outcome in eCourts Data	1(Decision)	1(Acquitted)	1(Not (Convicted))
No decision within 6 months	0	0	0
Acquitted	1	1	1
Neither acquitted nor convicted	1	0	1
Convicted	1	0	0

Notes: The outcome variables were coded based on the trial outcome recorded in eCourts. Columns 2 and 3 are defined so that a value of 1 represents a positive outcome.

3.5 Summary Statistics on Case Outcomes

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

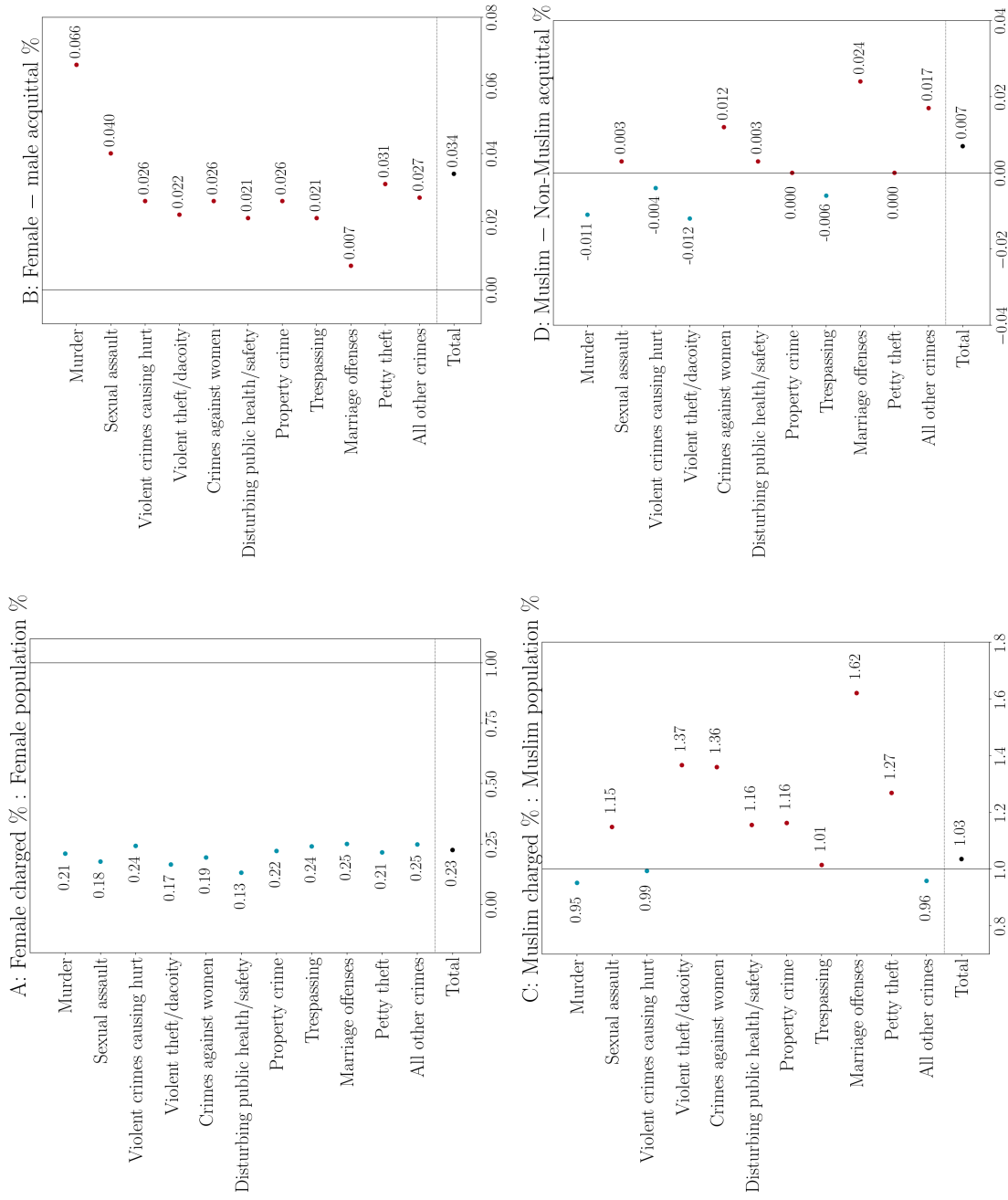
Figure 1 Panel A shows that the share of females charged under all crime categories is substantially lower than their population share. Men are three to five times more likely to be charged with crimes under any classification. Panel B shows that the conviction rate varies by crime, but overall it is about 1 percentage point lower 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 charge: relative to their population share, Muslims are 36% more likely to be charged with crimes against women, 37% more likely to be charged with robbery, and 62% more likely to be charged with marriage offenses, but 5% less likely to face charges for murder. Panel D shows that aggregate differences in conviction rates between Muslims and non-Muslims are small.

Table 2 shows descriptive statistics of judges 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 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.

¹⁸The corresponding point estimates are reported in Appendix Tables A3 and A4.

Figure 1: Summary statistics by crime category and defendant identity



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

Table 2: Summary statistics, by judge identity

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

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

4 Empirical Strategy

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

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 specification and considerations for estimating religious identity bias are identical. Specifications used in additional analysis on bias in contexts likely to activate identity are described with the results.¹⁹

¹⁹We also explored an event study specification exploiting case timing and changes in the cohort of judges sitting in each court, but we found that recently changed courts are more likely to see younger cases, violating the assumptions required for the event study analysis.

4.1 Random Assignment of Judges to Cases

As with much of the prior empirical literature, judge assignment in district courts is not strictly random but follows a set of rules that gives defendants and prosecutors virtually no control over which judge oversees the case. As described in Section 2, a case is assigned to a room in a court (and thus the judge in that room) when charges are filed, based on the police station and charge type, giving prosecutors and defendants little control over the judge’s identity. From a defendant’s perspective, the judge assignment is as good as random; for simplicity and consistency with the prior literature, we describe the approach as random assignment below.

Random assignment of judges to cases is empirically important because of the concern that judges could treat defendants differently not because of their identity but because of other case characteristics that are correlated with judge identity. For example, if Muslim judges could systematically choose to sit in cases with Muslim defendants who had committed less serious crimes, we might see in-group differences, but they would be due to differences in the underlying cases of Muslim defendants matched to Muslim judges rather than due to bias. Alternately, Muslim defendants and judges may be more likely to appear in some parts of the country than others. If those regions are characterized by different crime distributions, we might again mistakenly attribute those differences 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 different identities. In practice, the Indian court system runs this experiment whenever a defendant is charged in a jurisdiction with multiple judges of different identities on the bench.

We use a canonical regression approach to test for the effect of judge identity on case outcomes, as used by Shayo and Zussman’s (2011) analysis of judicial in-group bias in Israel. We model outcome $Y_{i,s,c,t}$ (e.g. 1=acquitted) for case i with charge s , filed in court c at time t as:

$$Y_{i,s,c,t} = \beta_1 \text{judgeMale}_{i,s,c,t} + \beta_2 \text{defMale}_{i,s,c,t} + \beta_3 \text{judgeMale}_{i,s,c,t} * \text{defMale}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t} \quad (1)$$

$$Y_{i,s,c,t} = \beta_1 \text{judgeNonMuslim}_{i,s,c,t} + \beta_2 \text{defNonMuslim}_{i,s,c,t} + \beta_3 \text{judgeNonMuslim}_{i,s,c,t} * \text{defNonMuslim}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t} \quad (2)$$

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. $\phi_{c,t}$ is a court-month or court-year fixed effect, and ζ_s is an act and section fixed effect. $\chi_{i,s,c,t}$ 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, $\chi_{i,s,c,t}$ represents controls for defendant gender, judge gender, and an interaction term of judge religion and defendant gender.

The charge section fixed effect ensures that we are comparing defendants charged with similar crimes. The court-time fixed effect ensures that we are comparing defendants who are being charged in the same court at the same time. Our primary specification uses a court-month fixed effect, while a secondary specification uses a court-year fixed effect. The court-year fixed effect allows a 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 filings to match these absences, nor do we find evidence of disproportionate identity matching in balance tests of either specification below. Court-time periods with no variation in judge identity are retained to increase the precision of fixed effects and controls, but they do not affect the coefficients of interest. We drop court-time periods where only one judge appears. We also test a specification with judge fixed effects, which controls for the average acquittal behavior of each individual judge.²⁰ Standard errors are clustered at the judge level since judge assignment is the level of randomization.

There are three causal effects of interest. β_1 describes the causal effect on a female defendant of having a male judge assigned to her case rather than a female judge. $\beta_1 + \beta_3$ describes the causal effect on a *male* defendant of having a male judge assigned to his case. The difference between these effects (or β_3) is the own-gender bias — it tells us whether individuals receive better outcomes when a judge matching their gender identity is randomly assigned to their case. Since all three causal effects are of interest, we report all three coefficients in the regression tables. The coefficient meanings are analogous in Equation 2.

In 44% of cases, a case stays in the courts long enough such that the judge making a decision on the case is different from the judge to whom the case was randomly assigned. In these cases, we continue to use the randomly assigned initial judge as the relevant

²⁰This specification is included for completeness, but is unnecessary for identification (as are the judge and defendant demographic controls) if judges are indeed effectively assigned randomly.

actor. We find virtually identical effects if we limit the sample to cases decided by the initially assigned judge, but this is not our primary specification because a rapid decision is itself an outcome. Further, even if the filing judge does not make the final ruling on a case, they can make key decisions on the case process that influence the decision, such as allowing witnesses, admitting evidence, and determining the schedule on which the case is resolved.

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:

$$\begin{aligned} \text{judgeFemale}_{i,s,c,t} = & \beta_1 \text{defFemale}_{i,s,c,t} + \beta_2 \text{defMuslim}_{i,s,c,t} + \gamma \phi_{c,t} + \zeta_s \\ & + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t} \end{aligned} \quad (3)$$

$$\begin{aligned} \text{judgeMuslim}_{i,s,c,t} = & \eta + \gamma_1 \text{defMuslim}_{i,s,c,t} + \gamma_2 \text{defFemale}_{i,s,c,t} + \gamma \phi_{c,t} + \zeta_s \\ & + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}, \end{aligned} \quad (4)$$

with variables defined as above. Our identification strategy relies on β_1 and γ_1 being equal to zero.

Balance estimates are shown in Table 3.²¹ Male and female defendants are effectively equally likely to be assigned to female judges, and similarly, Muslim and non-Muslim defendants are equally likely to be assigned to Muslim judges. Nevertheless, to ensure that any small differences in assignment to judges do not influence our results, we control for judge and defendant gender in the religion regressions (and for judge and defendant religion in the gender regressions).

Overall, the balance test indicates that defendants of a given identity do not face different odds of encountering judges with the same identity. This null supports the essential assumption underlying causal identification of judge identity on case outcomes.

²¹Note that here, as below the sample sizes are closer to 5 million than 6 million (the numbers from Appendix Figure A2). This reduction in the sample is due to court-time subsets that lack treatment variation (active judges of both identity groups).

Table 3: Balance test for assignment of judge identity

	(1)	(2)	(3)	(4)
	Female judge	Female judge	Muslim judge	Muslim judge
Female defendant	-0.000 (0.001)	-0.000 (0.001)	0.001 (0.000)	0.001 (0.000)
Muslim defendant	0.001 (0.001)	0.001 (0.001)	0.000 (0.001)	0.000 (0.001)
Observations	5155404	5168610	5240281	5253483
Fixed Effect	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 specification 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 fixed effects. 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 fixed effects. Charge section fixed effects have been used across all columns reported. Heteroskedasticity robust standard errors are reported below point estimates.

5 Results

5.1 Effect of assignment to judge types

The first two rows of Table 4 Panel A present the impact, for female and male defendants respectively, of being randomly assigned to a male judge – these are β_1 and $\beta_1 + \beta_3$ in Equation 1. The third row shows the difference between these two coefficients (β_3), which is the own-gender bias. The outcome variable is an indicator for defendant acquittal. Columns 1–3 show results using court-month fixed effects, while Columns 4–6 use court-year fixed effects. Within each set of three columns, the second column adds additional demographic controls, while the third column adds judge fixed effects.²²

Male judges consistently deliver fewer acquittals than female judges. The point estimate on this effect is nearly identical for both male and female defendants across all specifications. The coefficient stability is what we would expect if judge assignment is

²²We note again that the reported sample size of about 5.2 million (reduced from 6 million or 6.6 million) is due to dropping court-time subsets that lack treatment variation (active judges of both identity groups).

indeed as good as random. The own-gender bias is a tight zero; the effect estimates rule out even a very small in-group bias effect of 0.6 percentage points with 95% confidence.²³

Table 4 Panel B shows the effect of filing judge gender on an indicator for case resolution within six months of being filed. Cases assigned to male judges are resolved slightly more quickly, but this difference is unaffected by defendant gender; the in-group bias effect is again a precise zero. In short, we do not find substantial gender bias on any dimension.

Table 5 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 effect of judge religion on the acquittal rate is again a precise zero. The point estimates on any form of bias for acquittals are never higher than 0.2 percentage points and the estimates rule out an own-religion bias of 0.6 percentage points with 95% confidence.²⁴ Religious in-group bias is also absent in the speed of judicial decisions, nor is there any evidence that Muslim and non-Muslim judges have different rates of resolving cases (Table 5).

5.2 In-group Bias on the Basis of Caste

The analysis thus far ignores one of the most important social cleavages in India: caste. Ideally, we would like to run similar tests, where judge and defendant identity match on the caste dimension. Such an analysis is more difficult for three reasons. First, individuals in the same broad caste category (*varna*) may not be in the same subcaste (*jati*). Identity similarity based on caste is arguably continuous rather than discrete. Second, individual names do not identify caste as precisely as they identify Islamic religion or gender identity. Caste-name identification can vary by location, for example. We have not managed to develop a reliable correspondence between names and caste, mainly because training data for this correspondence are scarce and have limitations. Third, there are very few judges in the most identifiable caste categories: Scheduled

²³Appendix Table A5 shows bias effects on conviction rates; the estimates again are a tight zero. Appendix Table A6 shows estimates when we exclude closed cases for which we are unable to determine the outcome. We prefer the specification in Table 4, because the inability to determine an outcome is itself an outcome. We also find no effect of gender or religious match on whether the outcome is determinate (Appendix Table A7).

²⁴Appendix Tables A8 and A9 show results on conviction rates, and on acquittals with ambiguous results dropped. While we find marginally significant bias effects (in the expected direction) in a handful of specifications, the majority are statistically insignificant, and the point estimate on the bias term is never higher than 0.5 percentage points. Appendix Table A10 shows there is no effect of in-group bias on an indicator for an ambiguous case outcome.

Table 4: Impact of assignment to a male judge on defendant outcomes

<i>Panel A: Outcome variable: Acquittal rate</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	-0.008*** (0.003)	-0.007** (0.003)	—	-0.007*** (0.003)	-0.007** (0.003)	—
Male judge on male defendant	-0.006*** (0.002)	-0.006** (0.003)	—	-0.006*** (0.002)	-0.005** (0.003)	—
Difference = Own gender bias	0.001 (0.002)	0.001 (0.002)	0.000 (0.002)	0.002 (0.002)	0.001 (0.002)	0.000 (0.002)
Reference group mean	0.175	0.176	0.176	0.175	0.176	0.176
Observations	5250907	5156887	5155378	5264320	5170380	5168583
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year
<i>Panel B: Outcome variable: Decision within six months of filing</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	0.023*** (0.004)	0.022*** (0.004)	—	0.022*** (0.004)	0.020*** (0.004)	—
Male judge on male defendant	0.022*** (0.003)	0.021*** (0.004)	—	0.021*** (0.003)	0.020*** (0.004)	—
Difference = Own gender bias	-0.001 (0.002)	-0.001 (0.002)	-0.002 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.002 (0.002)
Reference group mean	0.285	0.284	0.284	0.284	0.284	0.284
Observations	4382525	4301724	4300307	4395262	4314529	4312834
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

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

Reference group: Female judges, female defendants.

Charge section fixed effects have been used across all columns reported.

Specification: $Y_{i,s,c,t} = \beta_1 \text{judgeMale}_{i,s,c,t} + \beta_2 \text{defMale}_{i,s,c,t} + \beta_3 \text{judgeMale}_{i,s,c,t} * \text{defMale}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}$

Table 5: Impact of assignment to a non-Muslim judge on defendant outcomes

<i>Panel A: Outcome variable: Acquittal rate</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	0.008 (0.004)	0.007 (0.005)	—	0.007 (0.004)	0.006 (0.005)	—
Non-Muslim judge on non-Muslim defendant	0.007** (0.003)	0.007* (0.004)	—	0.007** (0.003)	0.006 (0.004)	—
Difference = Own religion bias	-0.001 (0.003)	0.000 (0.003)	0.002 (0.002)	-0.001 (0.003)	0.000 (0.003)	0.002 (0.002)
Reference group mean	0.18	0.183	0.183	0.18	0.183	0.183
Observations	5684426	5241649	5240140	5697480	5255137	5253328
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year
<i>Panel B: Outcome variable: Decision within six months of filing</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	0.010 (0.006)	0.008 (0.008)	—	0.009 (0.006)	0.005 (0.007)	—
Non-Muslim judge on non-Muslim defendant	0.005 (0.006)	0.002 (0.007)	—	0.003 (0.005)	-0.001 (0.007)	—
Difference = Own religion bias	-0.006 (0.004)	-0.006 (0.004)	0.002 (0.003)	-0.006 (0.004)	-0.006 (0.005)	0.003 (0.003)
Reference group mean	0.292	0.289	0.289	0.292	0.288	0.288
Observations	4748758	4375562	4374140	4761183	4388369	4386657
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

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

Reference group: Muslim judges, Muslim defendants.

Charge section fixed effects have been used across all columns reported.

Specification: $Y_{i,s,c,t} = \beta_1 \text{judgeNonMuslim}_{i,s,c,t} + \beta_2 \text{defNonMuslim}_{i,s,c,t} + \beta_3 \text{judgeNonMuslim}_{i,s,c,t} * \text{defNonMuslim}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}$

Castes and Scheduled Tribes.

To measure bias by caste, we follow [Fisman et al. \(2017\)](#) and define individuals as being in the same cultural group if they share a last name. The measure is admittedly imperfect. Last names are effective identifiers of caste for many social groups, but they are more numerous than castes. Thus, two individuals with the same last name are likely to be in the same social group, but two individuals in the same social group will often have different last names. The measure is also a combination of caste and religious group similarity – for instance, the names *Kaur* and *Singh* (two of the most common names in the data), are likely indicators of Sikhism but are not very informative of socioeconomic status. Our measure thus underestimates the degree of affinity. As such, it will underestimate the true extent of in-group bias.

Exploiting exogenous assignment of judges to cases, we use Equation 5 to determine whether judges deliver more favorable outcomes to defendants who share their last name:

$$Y_{i,s,c,t} = \beta_1 \text{sameLastName}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \nu_i + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t} \quad (5)$$

where subscripts i, s, c, t , court-time ($\phi_{c,t}$) and act/section (ζ_s) fixed effects are defined as above. We include an additional fixed effect for the defendant’s last name (ν_i) to control for the possibility that individuals from some social groups are more or less likely to be acquitted or to appear as judges. The vector $\delta \chi_{i,s,c,t}$ includes defendant and judge gender and religion. We limit the sample to individuals with last names that match at least one judge in their district at any time, and we weight the regression by the inverse size of the defendant name group, such that each last name receives equal weight.

Table 6 shows the results. Across specifications, we find a 1.3–2.5 percentage point increase in the acquittal rate when judges are given a case where the defendant’s last name matches their own. The effect is robust to looser definitions of last name similarity (for example, treating *Patil* and *Patel* as similar). There are several factors that differentiate this result from the null effects of gender and religious similarity above. First, the social group defined by last names is much more narrow than the social group defined by all people of the same gender or religion. The family name similarity could also capture similarity in economic status.

The effect size remains small compared with the majority of studies on judicial bias. The effect size is comparable in magnitude to [Fisman et al. \(2017\)](#), however, who find

Table 6: Impact of assignment to a judge with the same last name on defendant outcomes

	(1)	(2)	(3)	(4)
	Acquitted	Acquitted	Acquitted	Acquitted
Same last name	0.013** (0.006)	0.014** (0.006)	0.019* (0.011)	0.025*** (0.009)
Observations	2239516	2237502	2258437	2256242
Fixed Effect	Court-month	Court-month	Court-year	Court-year
Judge Fixed Effect	No	Yes	No	Yes

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports results from a test of the impact of random assignment to a judge with the same last name as the defendant on likelihood of acquittal, see Equation 5. Charge section and last name fixed effects have been used across all columns reported.

that loan officers in India provide 6.5% more loans to borrowers with matching last names.

5.3 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 specific contexts or primes. This section examines several circumstances where gender or religious identity may become particularly salient to judges.

We first examine the subset of cases where the victim and defendant have different identities. In this case, when the defendant and judge are mismatched, it implies that the judge and victim share the same gender or religious identity.²⁵ The identity match or mismatch between judge and defendant may be particularly salient in this case (Baldus et al., 1997; Baumgartner et al., 2015; ForsterLee et al., 2006). Examining a subset of cases where the victim in the case is identified and can be matched to an identity group, we interact all of the terms in the standard in-group bias estimation (Equation 3) with an indicator for whether the victim and defendant have an identity mismatch.

Columns 1 and 2 of Table 7 show the results for gender and religious bias respectively.

²⁵In 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.

Table 7: Differential judge bias effect based on victim of crime

	(1)	(2)	(3)
	Acquitted	Acquitted	Acquitted
In-group Bias	0.004 (0.003)	0.001 (0.005)	0.000 (0.002)
In-group Bias * Victim Gender Mismatch	-0.006 (0.005)		
In-group Bias * Victim Religion Mismatch		0.007 (0.008)	
In-group Bias * Crime Against Women			-0.009 (0.007)
Observations	1790929	2022473	5149667
Fixed Effect	Court-month	Court-month	Court-month
Judge Fixed Effect	Yes	Yes	Yes
Bias	Gender	Religion	Gender
Sample	All	All	All

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports results from a test of in-group bias in cases where gender or religious identity are salient, i.e. when the defendant and victim belong to opposing identities (Columns 1 and 2), or when the charge is classified as a crime against women (Column 3). Charge section fixed effects have been used across all reported columns.

For legibility, we show only the coefficients on the in-group bias term (e.g. judgeMale * defMale for gender), and its interaction with the victim/defendant identity mismatch indicator. There is no evidence that in-group bias is larger when victim and defendant have different identities. Standard errors are larger due to the smaller sample and interaction specification, but the in-group bias effect is under 1 percentage point in both cases.²⁶

We next examine whether male and female judges rule differently on cases classified in the criminal code as crimes against women, where the judge and defendant gender identities may be particularly salient. These are largely evenly split between sexual assaults and kidnappings.²⁷ As above, we interact a “crime against women” indicator

²⁶The sample is smaller because cases in eCourts often do not code information about the crime victim.

²⁷One reason “kidnappings” are so common in the data is that this may be the formal charge filed against a man who elopes with a woman. Results are very similar for both the assault and kidnapping

Table 8: Impact of Ramadan on own religion bias in acquittal rates

	(1)	(2)	(3)	(4)
	Acquitted	Acquitted	Acquitted	Acquitted
Non-muslim judge	0.012** (0.006)		0.012** (0.006)	
Non-muslim defendant	0.002 (0.004)	0.002 (0.002)	0.003 (0.004)	0.003 (0.002)
Ramadan	0.125*** (0.012)	0.091*** (0.012)	0.123*** (0.013)	0.091*** (0.012)
Own religion bias	-0.003 (0.004)	-0.003 (0.002)	-0.002 (0.004)	-0.002 (0.002)
Own religion bias * Ramadan	0.016 (0.010)	0.015 (0.009)	0.020** (0.010)	0.018* (0.010)
Observations	5211432	6050334	5224676	6062453
Fixed Effect	Court-month	Court-month	Court-year	Court-year
Judge fixed effect	No	Yes	No	Yes
Sample	Full sample	Full sample	Full sample	Full sample

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table reports results from a test of in-group religious bias conditional on trial during the months of Ramadan (following [Mehmood and Seror \(2020\)](#)). Charge section fixed effects have been included across all reported columns.

with all of the variables in the standard bias specification, and show the bias and the interaction term. We find no evidence that male judges are particularly more or less strict against male defendants in these cases.

Appendix Table [A11](#) shows estimates from the standard bias specification for crimes against women only. Male judges are less likely to acquit for these cases than female judges, but show no disproportional harshness for male defendants.

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.²⁸ We use a simple binary variable indicating days in the subsets of the data.

²⁸Unlike the sample in [Mehmood and Seror \(2020\)](#), 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.

month of Ramadan and interact it with all the variables in the standard bias specification.²⁹

Table 8 reports the results, with and without judge fixed effects, and with either court-month or court-year fixed effects. The coefficient of interest is the interaction of the own-religion bias with the Ramadan indicator. There is a consistent 1.5–2.0 percentage point in-group bias effect, though it is of marginal statistical significance, with a p-value under 0.05 in only one of the four specifications. The result suggests that religious in-group bias may be activated when religious identity is particularly salient, but the effect size remains small relative to other studies.³⁰

6 Conclusion

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

In this paper, we analyze a new, large-scale dataset for the analysis of court proceedings in India, consisting of over 5 million decisions in criminal cases. We estimate robust, tight zero effects of judicial in-group bias along the dimensions of religion and gender. While we do find evidence of in-group bias along the caste dimension, and evidence that religious in-group bias is activated during the month of Ramadan, even these point estimates are small.

Our null estimates of in-group bias contrast with findings in the previous literature, which has tended to find large effects. Figure 2 Panel A compares our point estimates with estimates of bias from the studies most similar to ours.³¹ Effect sizes are stan-

²⁹Note that for this table only, we use the identity of the judge *deciding* on the case, rather than the judge to whom it was assigned initially. Our implicit assumption is that the effect of Ramadan will affect the outcome on the day the decision is reached, rather than on the day the case first appeared before a judge (see Section 4 for more on how we treated cases seen by more than one judge).

³⁰Mehmood and Seror (2020) find in Pakistan that conviction rates are 16 percentage points lower during the month of Ramadan. The following section discusses the size of all of our estimates compared with other studies.

³¹We included every study we could find that focused on measuring in-group bias among judges on

standardized by dividing each in-group bias effect by the sample standard deviation of the outcome variable. The high end of our confidence interval is an order of magnitude smaller than nearly all prior studies.

One reason our results stand out from the literature could be publication bias in studies of judicial in-group bias. Figure 2 Panel B plots the effect size of each study above against the standard error of the main estimated effect.³² With the standard error axis reversed, this is the standard “funnel plot” used to examine publication bias. In the absence of publication bias or a design-based mechanical correlation (such as adaptive sampling), study estimates should form a cone that is centered around the true estimate (Egger et al., 1997; Gerber et al., 2001; Levine et al., 2009; Slavin and Smith, 2009; Kühberger et al., 2014).³³ The set of studies examined here show a highly non-conic and asymmetric shape, where effect magnitude is highly correlated with effect size, such that many of the studies fall just outside the cone boundary defining statistical significance at the 5% level. While other explanations may be considered, the funnel plot is at least consistent with a substantial degree of publication bias.

The most straightforward interpretation of our findings is that, unlike judges analyzed in the other papers, India’s district court judges do not exhibit in-group bias along the identity margins that we study here. Our research is consistent with judges taking their role seriously and working hard to provide justice on fair terms to all litigants, or with judicial institutions that constrain discretion and protect defendants from biased decision-making. It is also possible that the social distance between (normally) upper class judges and (normally) lower class criminal litigants may mitigate a sense of shared identity between judges and litigants. Yet it is also consistent with corruption that is blind to religious and gender identity. Distinguishing among these explanations for the absence of bias is an important area for future work.

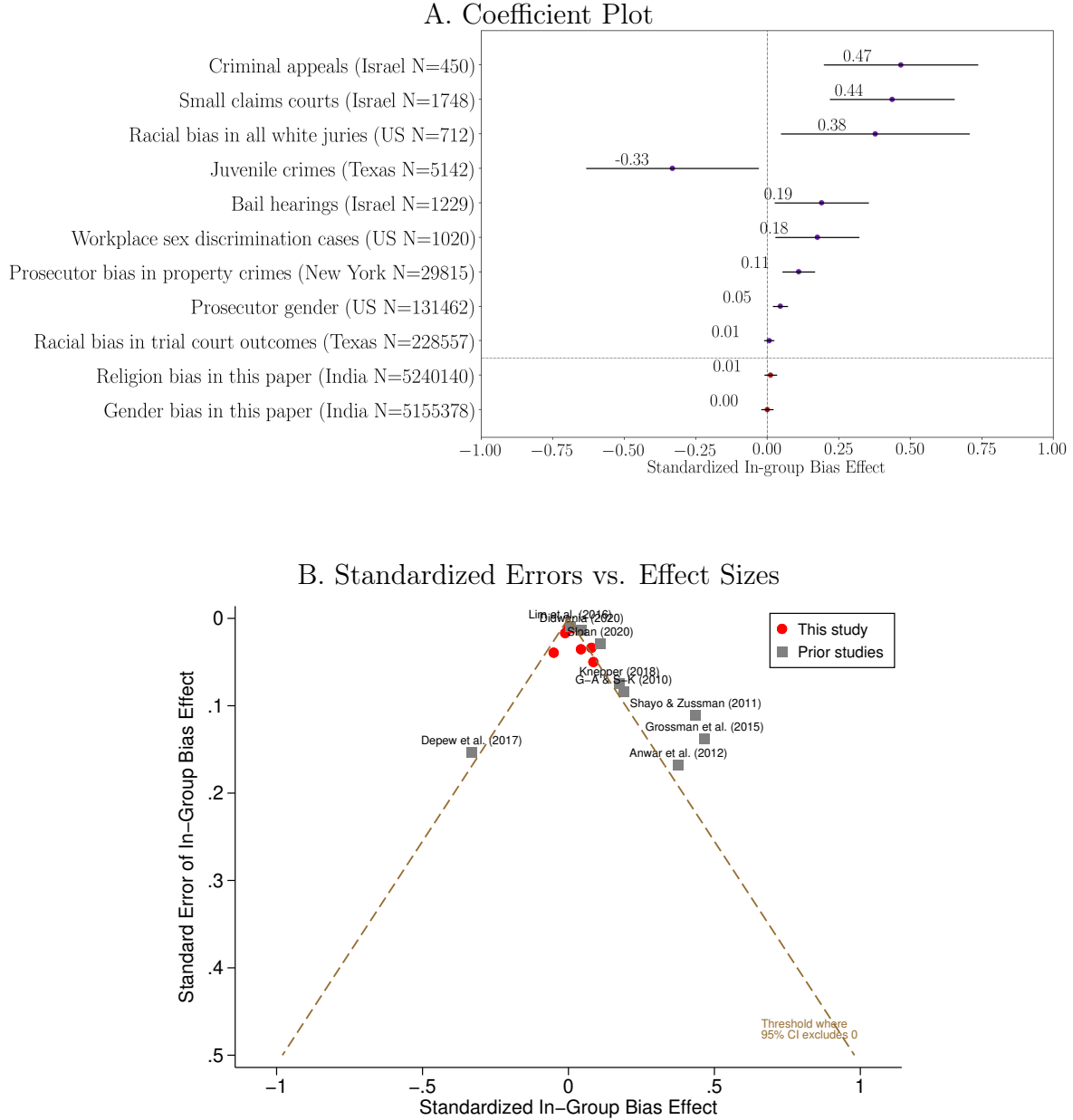
Whatever the explanation, our evidence suggests that despite India’s deep cleavages and much evidence of discrimination, the state has successfully built a judicial institution where existing disparities are not entrenched by in-group bias among judges.

a race/ethnicity, gender, or religious dimension, that exploited an as-good-as-random judge or jury assignment mechanism for causal identification.

³²When papers report multiple specifications for the main effect, we used the effect size described most prominently in the text or described by the authors as the “main specification.” When papers had multiple outcomes, we used the outcome most similar to the acquittal or conviction rate, as in this study. If these were unavailable, we used the outcome most prominently described in the paper’s abstract and introduction.

³³A cone shape is expected because studies with larger standard errors should produce a wider range of estimates around the true value.

Figure 2: Comparison with judicial bias estimates in other contexts



Notes: This figure shows point estimates of in-group bias from other studies in the relevant literature. From top to bottom, the coefficients of in-group bias (Panel A) correspond to [Shayo and Zussman \(2011\)](#), [Anwar et al. \(2012\)](#), [Depew et al. \(2017\)](#), [Knepper \(2018\)](#), [Sloane \(2019\)](#), [Didwania \(2018\)](#), [Lim et al. \(2016\)](#), [Gazal-Ayal and Sulitzeanu-Kenan \(2010\)](#), and the main estimates from the present study respectively. Panel B plots reported bias effects (Y axis) against effect standard errors. All effect sizes are standardized (dividing outcome variables by their standard deviation) to allow comparison across studies.

This institutional success should be celebrated, not least because it provides guidance to policymakers in how to allocate resources to address the clear and extant social disparities in Indian society. As a starting point, we have not yet ruled out bias in the broader criminal justice system. Notwithstanding our results on acquittals, the legal system could still be biased against Muslims and women due to unequal geographic distribution of policing, discrimination in investigations, police/prosecutor decisions to file cases, the severity of charges applied, the severity of penalties imposed, the appeals process, civil litigation, or via other factors. Based on our evidence, these other parts of the justice pipeline might deserve more immediate attention from policymakers than judge acquittal decisions.

These points also provide some guidance and motivation for social scientists. More research is needed to study the entire justice process in India and other developing countries. That process could start with expansion of the publicly available datasets.

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.
- 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 Effects in US Executions, 1976–2013. *Politics, Groups, and Identities*, 3(2):209–221.
- Bertrand, M., Hanna, R., and Mullainathan, S. (2010). Affirmative 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 Effects of Sex on Judging. *American Journal of Political Science*, 54(2):389–411.
- 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., Wen, 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 Conflicts and Intergroup Lending. *Journal of Political Economy*, 128(9):3346–3375.
- ForsterLee, R., ForsterLee, L., Horowitz, I. A., and King, E. (2006). The Effects of Defendant Race, Victim Race, and Juror Gender on Evidence Processing in a Murder Trial. *Behavioral Sciences & the Law*, 24(2):179–198.
- 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., Shroff, R., et al. (2016). Precinct or Prejudice? Understanding Racial Disparities in New York City’s Stop-And-Frisk Policy. *The Annals of Applied Statistics*, 10(1):365–394.
- 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 Influence 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 Effect 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 Effect Sizes are Negatively Correlated in Meta-Analyses: Evidence and Implications of a Publication Bias Against Nonsignificant 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. and Seror, A. (2020). 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.
- 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.
- 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.
- Shayo, M. and Zussman, A. (2011). Judicial Ingroup Bias in the Shadow of Terrorism. *The Quarterly Journal of Economics*, 126(3):1447–1484.
- 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.
- 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


E-COURTS
 OFFICIAL WEBSITE OF DISTRICT COURTS

[Back](#)

District and Sessions Court, Vidisha

Case Details

Case Type	: ST - SESSIONS TRIAL		
Filing Number	: [REDACTED]	Filing Date	: [REDACTED]
Registration Number	: [REDACTED]	Registration	: [REDACTED]
CNR Number	: [REDACTED]		

Case Status

First Hearing Date	: [REDACTED]		
Decision Date	: [REDACTED]		
Case Status	: CASE DISPOSED		
Nature of Disposal	: Contested-Judgment Delivered Acquitted		
Court Number and Judge	: 4-II Additional District and Session Judge		

Petitioner and Advocate

1) State Government

Respondent and Advocate

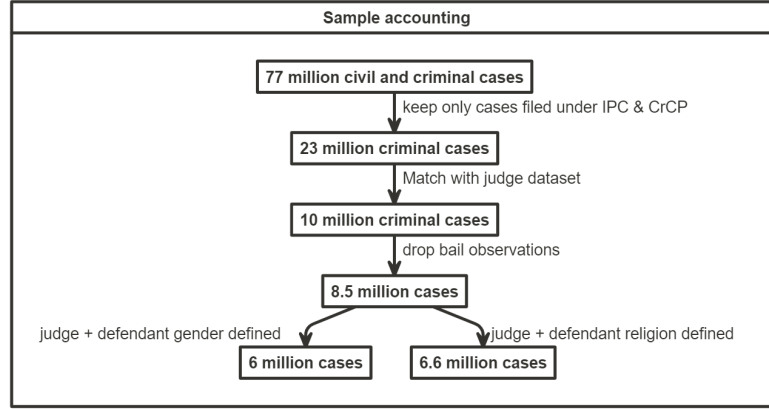
1) [REDACTED]
 2) [REDACTED]

Acts

Under Act(s)	Under Section(s)
Indian Penal Code 1860	302,294,323,34

Notes: The figure 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.

Figure A2: Sample accounting



Notes: The figure displays the process through which we arrive at the analysis dataset from the parent dataset of 77 million legal case records. After restricting the sample to criminal cases, matching these criminal cases with our judge dataset, and dropping bail observations, 8.5 million case records remain. For 6 million (6.6 million) records we can then assign the gender (religion) of the judge and defendant using our machine classifier. Further reductions of the sample size, for instance in results Tables 4 and 5, are due to the inclusion of fixed effects. When during the same court-month/court-year no female **and** male (or Muslim **and** non-Muslim) judges were present, observations are dropped.

Figure A3: India eCourts Sample Judge Information inside the Search Engine

E-COURTS
OFFICIAL WEBSITE OF DISTRICT COURTS

Court Orders : Search by Court Number

☒ Court Complex ☐ Court Establishment

* Court Complex: Kannad, Civil and Criminal Court

* Court Number: **Select Court Name**

Captcha

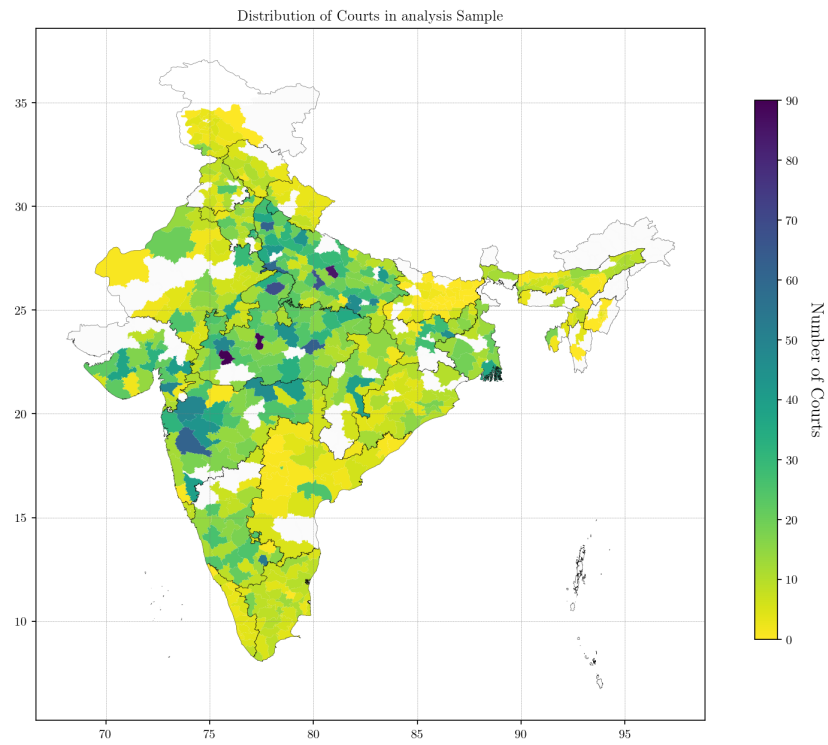
* Enter Captcha

-----Civil and Criminal Court, Kannad-----

- 1-SHRI. K. G. PALDEWAR-CIVIL JUDGE J.D. J.M.F.C. KANNAD(09-06-2008/07-06-2009)
- 1-SHRI. K. G. PALDEWAR-CIVIL JUDGE J.D. J.M.F.C. KANNAD(08-06-2009/31-10-2013)
- 1-SHRI H.S. AHIWALE-CIVIL JUDGE J.D. J.M.F.C. KANNAD(09-06-2014/02-03-2015)
- 1-SHRI A.S.PANDAGLE-CIVIL JUDGE J.D. J.M.F.C. KANNAD(08-06-2015/05-06-2016)
- 1-SHRI A.S.PANDAGLE-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(06-06-2016/06-06-2018)
- 2-KUM. K. M. CHAVAN-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(21-05-2005/30-05-2009)
- 2-KUM. K. M. CHAVAN-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(08-06-2009/07-06-2013)
- 2-SHRI. A. A. WALUJKAR-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(08-06-2013/31-10-2013)
- 2-SHRI. A. A. WALUJKAR-CIVIL JUDGE J.D. J.M.F.C. KANNAD(01-11-2013/08-06-2014)
- 2-SHRI. A. A. WALUJKAR-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(09-06-2014/02-03-2015)
- 2-SHRI. A. A. WALUJKAR-CIVIL JUDGE J.D. J.M.F.C. KANNAD(03-03-2015/07-06-2015)
- 2-SHRI. A. A. WALUJKAR-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(08-06-2015/05-06-2016)
- 2-SHRI A.M. HUSAIN-CIVIL JUDGE J.D. J.M.F.C. KANNAD(06-06-2016/03-06-2018)
- 2-SHRI P.S. KULKARNI-CIVIL JUDGE J.D. J.M.F.C. KANNAD(04-06-2018/16-11-2019)
- 2-SHRI P.H. PATIL-lind JT.CIVIL JUDGE J.D. AND J.M.F.C. KANNAD(03-12-2019/)
- 3-Ms. R. R. BEDAGKAR-lind JT.CIVIL JUDGE J.D. AND J.M.F.C. KANNAD(05-04-2010/07-06-2013)
- 3-SHRI. D. U. DONGRE-lind JT.CIVIL JUDGE J.D. AND J.M.F.C. KANNAD(08-06-2013/31-10-2013)
- 3-SHRI. D. U. DONGRE-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(01-11-2013/08-06-2014)
- 3-SHRI. D. U. DONGRE-lind JT.CIVIL JUDGE J.D. AND J.M.F.C. KANNAD(09-06-2014/02-03-2015)
- 3-SHRI. D. U. DONGRE-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(03-03-2015/07-06-2015)
- 3-SHRI. D. U. DONGRE-lind JT.CIVIL JUDGE J.D. AND J.M.F.C. KANNAD(08-06-2015/05-06-2016)
- 3-SHRI B.R. THAKUR-lind JT.CIVIL JUDGE J.D. AND J.M.F.C. KANNAD(06-06-2016/03-06-2018)
- 3-SHRI B.R. THAKUR-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(04-06-2018/02-06-2019)
- 3-SHRI P.K. AHIR-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(03-06-2019/11-11-2019)
- 3-SHRI P.K. AHIR-CIVIL JUDGE J.D. J.M.F.C. KANNAD(12-11-2019/)
- 4-SHRI A. A. WALUJKAR-PRESIDING OFFICER EVENING COURT(19-08-2010/10-06-2013)
- 4-SHRI P. B. PAWAR-lind JT.CIVIL JUDGE J.D. AND J.M.F.C. KANNAD(18-06-2018/11-11-2019)
- 4-SHRI P. B. PAWAR-JT.CIVIL JUDGE J.D. J.M.F.C. KANNAD(12-11-2019/)

Notes: Sample view of the eCourts court order search engine. We scraped the judge information implicitly given in the 'Court Number' drop-down list of the search mask on – in this case – https://services.ecourts.gov.in/ecourtindia_v4_bilingual/cases/s_order.php?state=D&state_cd=1&dist_cd=19 to obtain judge names and tenures.

Figure A4: Distribution of courts across districts in the analysis sample



Notes: This figure shows the geographical distribution of the trial courts in our sample. Black lines delineate states, and within those the unit of observation for this graphical illustration are districts. Districts marked in white are not included in our analysis.

Table A1: Summary of Name Classifier Training Datasets

<i>Panel A: Delhi voter rolls names</i>		
Gender	Instances	Percentage
Female	6,138,337	44.8%
Male	7,556,138	55.2%
Total	13,694,475	100.0%

<i>Panel B: National Railway exam names</i>		
Religion	Instances	Percentage
Buddhist	1,910	0.1%
Christian	11,194	0.8%
Hindu	1,174,076	84.8%
Muslim	163,861	11.8%
NA	33,882	2.4%
Total	1,384,923	100.0%

Notes: Panels A & B of this table show the distribution of identities in the underlying training datasets of the gender and religion LSTM name classification models respectively.

Table A2: Outcome variables mapped to dispositions

Disposition Name	Mapped Outcome(s)		
	Acquitted	Convicted	Decision
258 crpc [acquitted]	X		X
Acquitted	X		X
Allowed	X		X
Committed			X
Compromise			X
Convicted		X	X
Decided			X
Dismissed			X
Disposed			X
Fine			X
Judgement			X
Other			X
Plead guilty		X	X
Prison		X	X
Referred to lok adalat			X
Reject			X
Remanded			X
Transferred			X
Withdrawn			X
Missing			

Notes: This table illustrates the classification of the raw dispositions into our three outcome variables. In the table, no entry corresponds to the default value 0, and X denotes that the corresponding outcome value is set to 1. If a case has a disposition at all, the indicator variable *Decision* equals 1, and 0 otherwise. Conditional on having a disposition, if the disposition is clearly acquitted, the outcome variable *Acquitted* takes the value 1, and 0 otherwise. The outcome variable for *Conviction* has been coded analogously.

Table A3: Summary of charges, by gender of defendant

	(1) Female share	(2) Female share/ population share	(3) Female acquittal rate	(4) Male acquittal rate	(5) Difference (3) - (4)	(6) Number of cases
Murder	0.101	0.210	0.249	0.183	0.066	1,129,000
Sexual assault	0.085	0.177	0.275	0.235	0.040	254,928
Violent crimes causing hurt	0.116	0.242	0.213	0.187	0.026	1,846,000
Violent theft/dacoity	0.079	0.165	0.170	0.148	0.022	252,046
Crimes against women	0.093	0.194	0.274	0.248	0.026	725,388
Disturbed pub. health/tranquility	0.063	0.131	0.096	0.075	0.021	1,852,000
Property Crime	0.106	0.221	0.184	0.158	0.026	2,558,000
Trespass	0.115	0.240	0.223	0.202	0.021	339,045
Marriage offenses	0.120	0.250	0.271	0.264	0.007	326,214
Petty theft	0.103	0.215	0.180	0.149	0.031	946,890
Other crimes	0.119	0.248	0.204	0.177	0.027	9,008,000
Total	0.108	0.225	0.201	0.167	0.034	17,170,000

Notes: Column 1 of this table reports the share of female defendants for each crime category. Column 2 reports the ratio of the female share for each crime to the female population share in India. Column 3 reports the acquittal rate for females accused of each crime category. Column 4 reports the analogous acquittal rates for males. Column 5 reports the difference in female and male acquittal rates for each crime category. Column 6 reports the total number of case records in each crime category. The total number of cases in this table is larger than the 6 million cases mentioned in [A1](#) as we also include cases records in the statistics where only the defendant gender is defined, even if the judge gender is unknown.

Table A4: Summary of charges, by religion of defendant

	(1) Muslim share	(2) Muslim share/ population share	(3) Muslim acquittal rate	(4) Non-Muslim acquittal rate	(5) Difference (3) - (4)	(6) Number of cases
Murder	0.135	0.951	0.182	0.193	-0.011	1,204,000
Sexual assault	0.163	1.148	0.241	0.238	0.003	271,622
Violent crimes causing hurt	0.141	0.993	0.187	0.191	-0.004	1,980,000
Violent theft/dacoity	0.194	1.366	0.140	0.152	-0.012	271,901
Crimes against women	0.193	1.359	0.260	0.248	0.012	771,555
Disturbed pub. health/tranquility	0.164	1.155	0.078	0.075	0.003	2,002,000
Property Crime	0.165	1.162	0.161	0.161	0.000	2,711,000
Trespass	0.144	1.014	0.200	0.206	-0.006	362,459
Marriage offenses	0.230	1.620	0.285	0.261	0.024	344,708
Petty theft	0.180	1.268	0.153	0.153	0.000	1,003,000
Other crimes	0.136	0.958	0.195	0.178	0.017	9,556,000
Total	0.147	1.035	0.177	0.170	0.007	18,280,000

Notes: Column 1 of this table reports the share of Muslim defendants for each crime category. Column 2 reports the ratio of the Muslim share for each crime to the Muslim population share in India. Column 3 reports the acquittal rate for Muslims accused of each crime category. Column 4 reports the analogous acquittal rates for non-Muslims. Column 5 reports the difference in Muslim and non-Muslim acquittal rates for each crime category. Column 6 reports the total number of case records in each crime category. The total number of cases in this table is larger than the 6.6 million cases mentioned in [A1](#) as we also include cases records in the statistics where only the defendant religion is defined, even if the judge religion is unknown.

Table A5: Impact of assignment to a male judge on non-conviction

	<i>Outcome variable: Not convicted</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	0.003** (0.002)	0.002 (0.002)	—	0.003* (0.002)	0.002 (0.002)	—
Male judge on male defendant	0.003** (0.002)	0.002 (0.002)	—	0.003* (0.002)	0.002 (0.002)	—
Difference = Own gender bias	0.000 (0.001)	0.000 (0.001)	0.001 (0.001)	0.000 (0.001)	0.000 (0.001)	0.001 (0.001)
Reference group mean	0.947	0.947	0.947	0.947	0.947	0.947
Observations	5250907	5156887	5155378	5264320	5170380	5168583
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

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

Reference group: Female judges, female defendants.

Charge section fixed effects have been used across all columns reported.

Specification: $Y_{i,s,c,t} = \beta_1 \text{judgeMale}_{i,s,c,t} + \beta_2 \text{defMale}_{i,s,c,t} + \beta_3 \text{judgeMale}_{i,s,c,t} * \text{defMale}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}$

Table A6: Impact of assignment to a male judge on acquittal rates, dropping ambiguous outcomes

	<i>Outcome variable: Acquittal rate</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	-0.003 (0.005)	-0.008 (0.005)	—	-0.005 (0.004)	-0.009* (0.005)	—
Male judge on male defendant	0.000 (0.004)	-0.005 (0.005)	—	-0.001 (0.004)	-0.006 (0.004)	—
Difference = Own gender bias	0.002 (0.003)	0.003 (0.003)	0.004 (0.003)	0.003 (0.003)	0.003 (0.003)	0.004 (0.003)
Reference group mean	0.663	0.664	0.664	0.666	0.666	0.666
Observations	1182938	1162073	1159551	1203939	1183155	1180587
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

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

Reference group: Female judges, female defendants.

Charge section fixed effects have been used across all columns reported.

Specification: $Y_{i,s,c,t} = \beta_1 \text{judgeMale}_{i,s,c,t} + \beta_2 \text{defMale}_{i,s,c,t} + \beta_3 \text{judgeMale}_{i,s,c,t} * \text{defMale}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}$

Table A7: Impact of assignment to a male judge on whether the disposition is ambiguous

<i>Outcome variable: Ambiguous outcome</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	0.011*** (0.003)	0.008** (0.004)	—	0.010*** (0.003)	0.007** (0.004)	—
Male judge on male defendant	0.011*** (0.003)	0.008** (0.003)	—	0.010*** (0.003)	0.007** (0.003)	—
Difference = Own gender bias	0.000 (0.002)	0.000 (0.002)	0.002 (0.002)	0.000 (0.002)	0.000 (0.002)	0.003 (0.002)
Reference group mean	0.737	0.736	0.736	0.737	0.735	0.735
Observations	5250907	5156887	5155378	5264320	5170380	5168583
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

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

Reference group: Female judges, female defendants.

Charge section fixed effects have been used across all columns reported.

Specification: $Y_{i,s,c,t} = \beta_1 \text{judgeMale}_{i,s,c,t} + \beta_2 \text{defMale}_{i,s,c,t} + \beta_3 \text{judgeMale}_{i,s,c,t} * \text{defMale}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}$

Table A8: Impact of assignment to a non-Muslim judge on non-conviction

	<i>Outcome variable: Not convicted</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	0.002 (0.003)	-0.003 (0.003)	—	0.001 (0.003)	-0.004 (0.003)	—
Non-Muslim judge on non-Muslim defendant	0.005 (0.003)	0.001 (0.003)	—	0.005 (0.004)	0.001 (0.003)	—
Difference = Own religion bias	0.003* (0.002)	0.004** (0.002)	0.003* (0.002)	0.004* (0.002)	0.005** (0.003)	0.003** (0.002)
Reference group mean	0.936	0.937	0.937	0.937	0.938	0.938
Observations	5684426	5241649	5240140	5697480	5255137	5253328
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

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

Reference group: Muslim judges, Muslim defendants.

Charge section fixed effects have been used across all columns reported.

Specification: $Y_{i,s,c,t} = \beta_1 \text{judgeNonMuslim}_{i,s,c,t} + \beta_2 \text{defNonMuslim}_{i,s,c,t} + \beta_3 \text{judgeNonMuslim}_{i,s,c,t} * \text{defNonMuslim}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}$

Table A9: Impact of assignment to a non-Muslim judge on acquittal rates, dropping ambiguous outcomes

	<i>Outcome variable: Acquittal rate</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	0.010 (0.006)	0.000 (0.008)	—	0.007 (0.006)	-0.006 (0.007)	—
Non-Muslim judge on non-Muslim defendant	0.010* (0.006)	0.002 (0.007)	—	0.009 (0.006)	-0.002 (0.007)	—
Difference = Own religion bias	0.000 (0.005)	0.002 (0.005)	-0.002 (0.004)	0.002 (0.005)	0.004 (0.005)	0.000 (0.004)
Reference group mean	0.676	0.682	0.682	0.677	0.683	0.683
Observations	1285585	1187003	1184449	1306413	1208242	1205650
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

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

Reference group: Muslim judges, Muslim defendants.

Charge section fixed effects have been used across all columns reported.

Specification: $Y_{i,s,c,t} = \beta_1 \text{judgeNonMuslim}_{i,s,c,t} + \beta_2 \text{defNonMuslim}_{i,s,c,t} + \beta_3 \text{judgeNonMuslim}_{i,s,c,t} * \text{defNonMuslim}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}$

Table A10: Impact of assignment to a non-Muslim judge on whether the disposition is ambiguous

<i>Outcome variable: Ambiguous outcome</i>						
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	-0.006 (0.005)	-0.013 (0.006)	—	-0.006 (0.005)	-0.013 (0.006)	—
Non-Muslim judge on non-Muslim defendant	-0.002 (0.005)	-0.008 (0.006)	—	-0.002 (0.005)	-0.008 (0.006)	—
Difference = Own religion bias	0.004 (0.004)	0.005 (0.004)	0.001 (0.003)	0.005 (0.004)	0.005 (0.004)	0.001 (0.003)
Reference group mean	0.735	0.732	0.732	0.734	0.732	0.732
Observations	5684426	5241649	5240140	5697480	5255137	5253328
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

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

Reference group: Muslim judges, Muslim defendants.

Charge section fixed effects have been used across all columns reported.

Specification: $Y_{i,s,c,t} = \beta_1 \text{judgeNonMuslim}_{i,s,c,t} + \beta_2 \text{defNonMuslim}_{i,s,c,t} + \beta_3 \text{judgeNonMuslim}_{i,s,c,t} * \text{defNonMuslim}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}$

Table A11: Impact of assignment to a male judge when the offence was a crime against women

	<i>Outcome variable: Acquittal rate</i>					
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	-0.027*** (0.009)	-0.018* (0.010)	—	-0.029*** (0.008)	-0.020** (0.009)	—
Male judge on male defendant	-0.028*** (0.005)	-0.021*** (0.007)	—	-0.025*** (0.005)	-0.018*** (0.007)	—
Difference = Own gender bias	-0.001 (0.008)	-0.003 (0.008)	-0.004 (0.008)	0.004 (0.007)	0.002 (0.007)	0.003 (0.007)
Reference group mean	0.27	0.27	0.27	0.276	0.276	0.277
Observations	261459	258216	255314	280236	276998	273964
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge fixed effect	No	No	Yes	No	No	Yes
Fixed Effect	Court-month	Court-month	Court-month	Court-year	Court-year	Court-year

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

Reference group: Female judges, female defendants.

Charge section fixed effects have been used across all columns reported.

Specification: $Y_{i,s,c,t} = \beta_1 \text{judgeMale}_{i,s,c,t} + \beta_2 \text{defMale}_{i,s,c,t} + \beta_3 \text{judgeMale}_{i,s,c,t} * \text{defMale}_{i,s,c,t} + \phi_{c,t} + \zeta_s + \delta \chi_{i,s,c,t} + \epsilon_{i,s,c,t}$