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

NBER Summer Institute
July 2022

Motivation

- **India's Muslims and women have unequal access to social and economic opportunities** (Ito 2009, Bertrand et al. 2010, Hnatkovska et al. 2012, Hanna and Linden 2012, Jayachandran 2015, Borker 2017, Asher et al. 2020)

Motivation

- **India's Muslims and women have unequal access to social and economic opportunities** (Ito 2009, Bertrand et al. 2010, Hnatkovska et al. 2012, Hanna and Linden 2012, Jayachandran 2015, Borker 2017, Asher et al. 2020)
- Judicial system founded on premise that **individuals discriminated against in informal settings should receive equal treatment under the law** (Aldashev et al. 2010, Sandefur and Siddiqi 2015)

Motivation

- **India's Muslims and women have unequal access to social and economic opportunities** (Ito 2009, Bertrand et al. 2010, Hnatkovska et al. 2012, Hanna and Linden 2012, Jayachandran 2015, Borker 2017, Asher et al. 2020)
- Judicial system founded on premise that **individuals discriminated against in informal settings should receive equal treatment under the law** (Aldashev et al. 2010, Sandefur and Siddiqi 2015)
- This paper focuses on **in-group bias by religion and gender in the Indian judiciary**

Muslims and women are underrepresented in the Indian judiciary

India is home to **195 million Muslims**, and women represent **48%** of the population.

4% of high court judges in India are Muslim. *(Quint, 2016)*

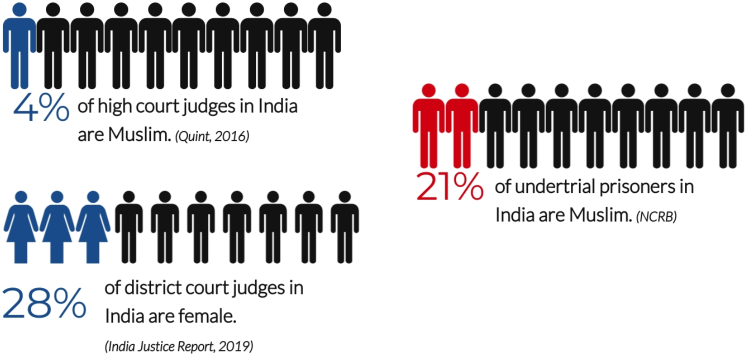
28% of district court judges in India are female.

(India Justice Report, 2019)

21% of undertrial prisoners in India are Muslim. *(NCRB)*

Muslims and women are underrepresented in the Indian judiciary

India is home to **195 million Muslims**, and women represent **48%** of the population.



Combination of in-group bias and under-representation could create **population-level discrimination**

This paper

What we do:

- Collect records from 80M cases

This paper

What we do:

- Collect records from 80M cases
- Use machine learning to classify judge and defendant names by religion/gender

This paper

What we do:

- Collect records from 80M cases
- Use machine learning to classify judge and defendant names by religion/gender
- Use (as if) random assignment of judges to defendants to look at in-group bias effects when judge and defendant identity match

What we find:

This paper

What we do:

- Collect records from 80M cases
- Use machine learning to classify judge and defendant names by religion/gender
- Use (as if) random assignment of judges to defendants to look at in-group bias effects when judge and defendant identity match

What we find:

- Tight zero estimates of in-group bias by gender or religion.
  - Can rule out effect sizes one-fifth those in other well-known studies.

This paper

What we do:

- Collect records from 80M cases
- Use machine learning to classify judge and defendant names by religion/gender
- Use (as if) random assignment of judges to defendants to look at in-group bias effects when judge and defendant identity match

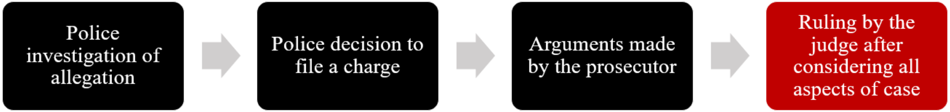
What we find:

- Tight zero estimates of in-group bias by gender or religion.
  - Can rule out effect sizes one-fifth those in other well-known studies.
- Limited evidence of in-group bias in some contexts where identity is salient
  - E.g. small effects during Ramadan and when judge and defendant share a rare last name
  - But null effects in other special contexts, e.g. crimes against women or when judge and victim identities match

Some caveats

We are focusing only on one form of bias in one part of the judicial system.

1. There could be in-group bias at each stage of the criminal justice process. We focus on the last stage, i.e: **judge decision-making**



Some caveats

2. There could be other kinds of bias—e.g. *all* judges could discriminate against Muslims and women
    - As we might expect, patterns of charges and convictions differ for Muslims and women.
    - But ID strategy based on exogenous judge assignment, so can only look at effects of judge identity (average vs in-group).

Data

Analysis

Conclusion

Case data

- **Source:** Indian eCourts platform (ecourts.gov.in):

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 Goverment

Respondent and Advocate

1) [redacted]

2) [redacted]

Acts

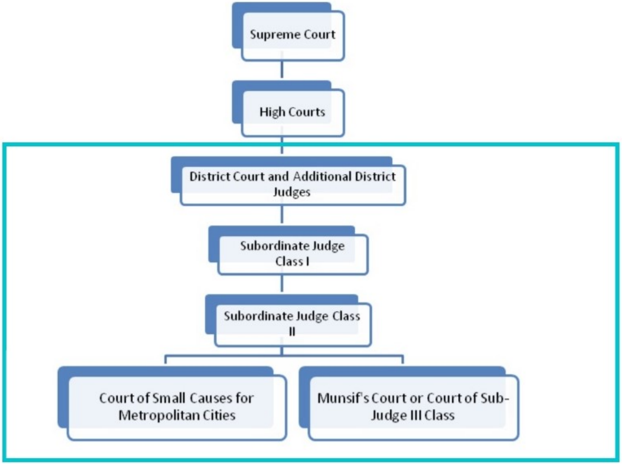
Under Act(s)	Under Section(s)
Indian Penal Code 1860	302,294,323,34

- 80 million records scraped, 2010-2018.

Case data

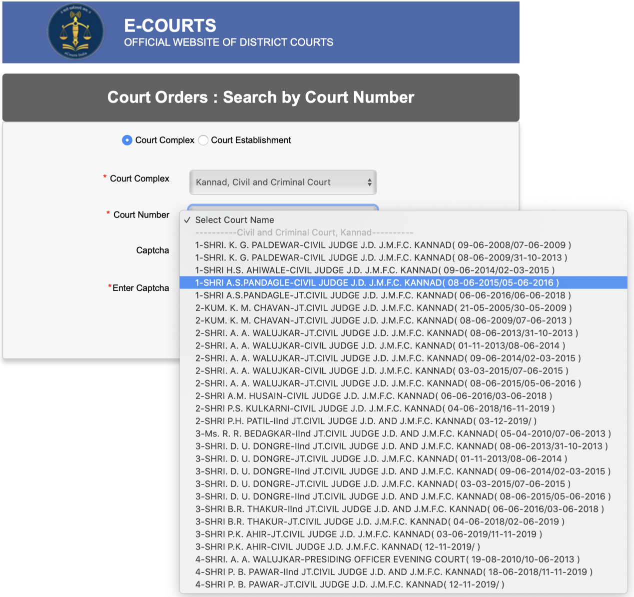
Scope:

- All courts that constitute the Indian **lower judiciary**: *District and session courts, and subordinate courts across all districts in India.*
- N=7,000 courts



Judge data:

- All judges in these district and session courts.
- N = 80,000 judges



Sample: All cases filed under the Indian Penal Code or the Code of Criminal procedure

8 mill. **Criminal case records**

The most cited relevant study in the literature draws from 1,758 case records in Israel (Shayo & Zussman, 2011)

~80,000 **Judge records**

Spanning all district courts, and subordinate courts across India.

7,640 **Trial courts**

Spanning all district courts, and subordinate courts across India.

Dataset

- Case data includes relevant dates (filing, registration, hearing, and decision), names of relevant actors (plaintiff, defendant, attorneys, victim), the acts and sections under which the case was filed, and the final decision or disposition.
- Judge data includes the judge's name, their position or designation, and the start and end date of the judge's appointment to each court.
- Joined case data with judge data based on judge's designation and case filing date.

Dataset

- Case data includes relevant dates (filing, registration, hearing, and decision), names of relevant actors (plaintiff, defendant, attorneys, victim), the acts and sections under which the case was filed, and the final decision or disposition.
- Judge data includes the judge's name, their position or designation, and the start and end date of the judge's appointment to each court.
- Joined case data with judge data based on judge's designation and case filing date.

- **Problem:** Records do not contain demographic data (gender/religion)

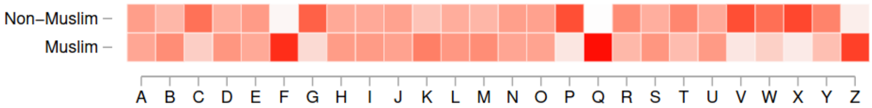
Dataset

- Case data includes relevant dates (filing, registration, hearing, and decision), names of relevant actors (plaintiff, defendant, attorneys, victim), the acts and sections under which the case was filed, and the final decision or disposition.
- Judge data includes the judge's name, their position or designation, and the start and end date of the judge's appointment to each court.
- Joined case data with judge data based on judge's designation and case filing date.

- **Problem:** Records do not contain demographic data (gender/religion)
- **Solution:** Apply a machine classifier to name string

Character-level Bidirectional Long Short-Term Memory (LSTM) network

- Specific type of Recurrent Neural Network – reads over name characters and interprets them based on a “memory” of the context of characters.



- Consider the last names *Khan* and *Khanna*.
- The fragment *”khan”* appears in both words; adding *”na”* changes the name from distinctly Muslim to distinctly non-Muslim.
- Standard fuzzy match would fail on this example because it ignores the context.

Training Dataset

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

Panel B: National Railway exam names		
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%

Pre-processing: Hindi characters transliterated to Latin. Normalize capitalization, punctuation, and spacing.

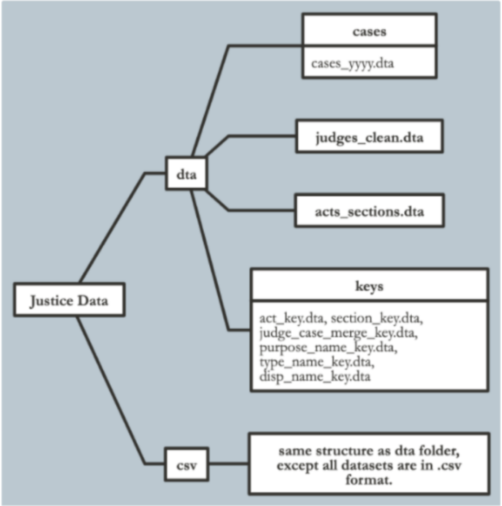
Classifier Performance on Unseen (Held-Out) Names

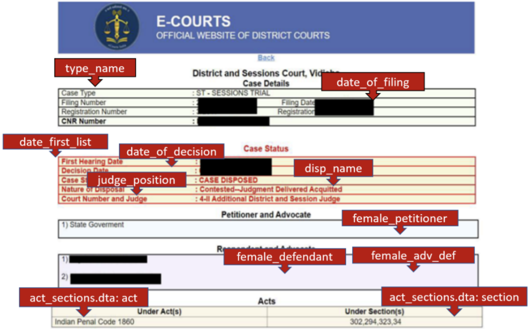
- Gender
  - Balanced accuracy = .975
  - F1 = .976.
- Religion
  - Balanced accuracy = .98
  - F1 = ..99

- Additional human annotation in the case dataset: accuracy is over 97%.

Code and trained gender classifier available at
`https://github.com/devdatalab/paper-justice/tree/main/classifier`.
Religion classifier available to researchers upon request.

Made anonymized data (with gender annotations) open to everyone





https://www.devdatalab.org/judicial-data

Data

Analysis

Conclusion

Constructing case outcomes

- $\sim$ 60% of case outcomes can be assigned without a problem, e.g:
    - $Y = 1$ when acquitted (good outcome for defendant)
    - $Y = 0$ when convicted (bad outcome for defendant)

- $\sim$ 40% of the cases have ambiguous outcomes in the metadata, such as: "decided", "judgement", "partly decreed", etc.

Constructing case outcomes

Outcome	Definition
Any decision	= 1 if case has a disposition at all, 0 if no decision
Acquitted	= 1 if disposition is clearly acquitted, 0 if disposition is something else
Not convicted	= 1 if the case has a disposition, 0 if the case has a disposition that is clearly convicted

Constructing case outcomes

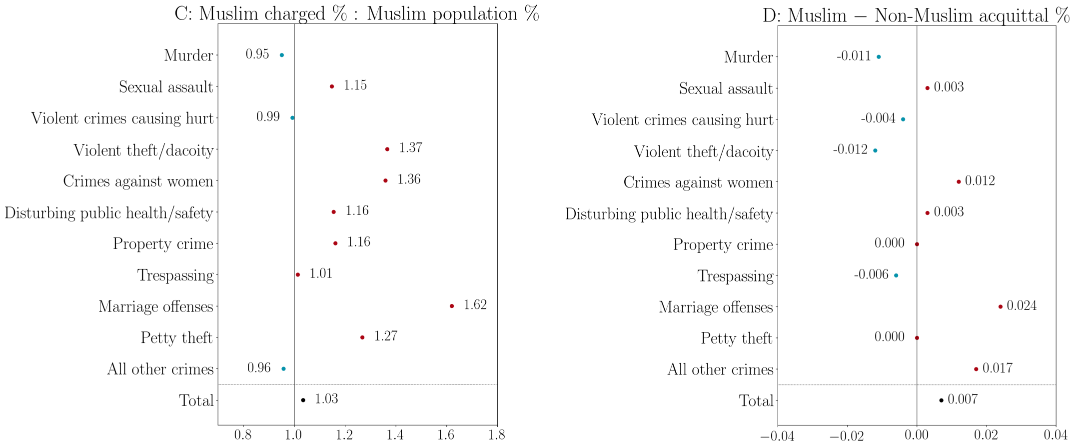
Outcome	Definition
Any decision	= 1 if case has a disposition at all, 0 if no decision
Acquitted	= 1 if disposition is clearly acquitted, 0 if disposition is something else
Not convicted	= 1 if the case has a disposition, 0 if the case has a disposition that is clearly convicted

Check robustness to dropping cases with ambiguous outcomes.

Women charged less often, acquitted more often

Notes: The left panel shows the ratio of share of accused female over the population share of females, for each crime category. The right panel shows the difference in mean acquittal rates between female and male defendants within crime categories

Muslims charged more often, acquitted more often; varies across offenses



Notes: The left panel shows the ratio of share of accused Muslim over the population share of Muslims, for each crime category. The right panel shows the difference in mean acquittal rates between muslim and non-muslim defendants within crime categories

Assignment of cases to judges in the lower judiciary

Cases are assigned to judges following a clear set of rules:

1. Police station location determines courthouse.

Assignment of cases to judges in the lower judiciary

Cases are assigned to judges following a clear set of rules:

1. Police station location determines courthouse.
2. Within court, courtroom is assigned based on station and charge.
   - More serious charges $\rightarrow$ more senior judge required.

Assignment of cases to judges in the lower judiciary

Cases are assigned to judges following a clear set of rules:

1. Police station location determines courthouse.
2. Within court, courtroom is assigned based on station and charge.
   - More serious charges $\rightarrow$ more senior judge required.
3. Judges rotate through rooms every few months, so same station/charge leads to different judge.

Assignment of cases to judges in the lower judiciary

Cases are assigned to judges following a clear set of rules:

1. Police station location determines courthouse.
2. Within court, courtroom is assigned based on station and charge.
   - More serious charges $\rightarrow$ more senior judge required.
3. Judges rotate through rooms every few months, so same station/charge leads to different judge.

Forum-shopping is prohibited and by all reports, not practiced.

Assignment of cases to judges in the lower judiciary

Cases are assigned to judges following a clear set of rules:

1. Police station location determines courthouse.
2. Within court, courtroom is assigned based on station and charge.
   - More serious charges $\rightarrow$ more senior judge required.
3. Judges rotate through rooms every few months, so same station/charge leads to different judge.

Forum-shopping is prohibited and by all reports, not practiced.

Conditional on court-time and charge fixed effects, cases are as good as randomly assigned.

Estimating equation

We model outcome Y_{ict} for case i in court c at time t as

$$\begin{aligned} Y_{i,c,t} = \alpha + \beta_1 \text{judge\_male}_{i,c,t} + \beta_2 \text{def\_male}_{i,c,t} + \\ \beta_3 \text{judge\_male}_{i,c,t} * \text{def\_male}_{i,c,t} + \phi_{c,t} + \delta \chi_{i,c,t} + \epsilon \end{aligned} \tag{1}$$

- **everything analogous for Muslim/non-Muslim**

Estimating equation

We model outcome Y_{ict} for case i in court c at time t as

$$\begin{gathered}Y_{i,c,t} = \alpha + \beta_1 \text{judge\_male}_{i,c,t} + \beta_2 \text{def\_male}_{i,c,t} + \\ \beta_3 \text{judge\_male}_{i,c,t} * \text{def\_male}_{i,c,t} + \phi_{c,t} + \delta \chi_{i,c,t} + \epsilon\end{gathered} \tag{1}$$

- **everything analogous for Muslim/non-Muslim**
- β_1 = effect of male judge on female defendant
- $\beta_1 + \beta_3$ = effect of male judge on male defendant

Estimating equation

We model outcome Y_{ict} for case i in court c at time t as

$$\begin{aligned} Y_{i,c,t} = \alpha + \beta_1 \text{judge\_male}_{i,c,t} + \beta_2 \text{def\_male}_{i,c,t} + \\ \beta_3 \text{judge\_male}_{i,c,t} * \text{def\_male}_{i,c,t} + \phi_{c,t} + \delta \chi_{i,c,t} + \epsilon \end{aligned} \tag{1}$$

- **everything analogous for Muslim/non-Muslim**
- β_1 = effect of male judge on female defendant
- $\beta_1 + \beta_3$ = effect of male judge on male defendant
- β_3 **= gender in-group bias**

Estimating equation

We model outcome Y_{ict} for case i in court c at time t as

$$
\begin{aligned}
Y_{i,c,t} = \alpha + \beta_1 \text{judge\_male}_{i,c,t} + \beta_2 \text{def\_male}_{i,c,t} + \\
\beta_3 \text{judge\_male}_{i,c,t} * \text{def\_male}_{i,c,t} + \phi_{c,t} + \delta\chi_{i,c,t} + \epsilon
\end{aligned}
\tag{1}
$$

- **everything analogous for Muslim/non-Muslim**
- β_1 = effect of male judge on female defendant
- $\beta_1 + \beta_3$ = effect of male judge on male defendant
- β_3 **= gender in-group bias**
- $\phi_{c,t}$: court-time fixed effect (month or year)
- $\delta\chi_{i,c,t}$: other covariates, including act-section fixed effects and other defendant characteristics
- standard err. clustered by judge (this does not matter much)

Testing exogenous judge assignment

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

Standard errors in parentheses

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

Ingroup gender bias is a tight zero

	Outcome variable: Acquittal rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Male judge on female defendant	-0.008***	-0.007**	—	-0.007***	-0.007**	—
	(0.003)	(0.003)		(0.003)	(0.003)	
Male judge on male defendant	-0.006***	-0.006**	—	-0.006***	-0.005**	—
	(0.002)	(0.003)		(0.002)	(0.003)	
Difference = Own gender bias	0.001	0.001	0.000	0.002	0.002	0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Reference group mean	0.176	0.177	0.177	0.176	0.177	0.177
Observations	5223433	5129780	5128269	5236865	5143294	5141492
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge 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.
Specification: $Y_{i,c,t} = \alpha + \beta_1 judge\_male_{i,c,t} + \beta_2 def\_male_{i,c,t} + \beta_3 judge\_male_{i,c,t} * def\_male_{i,c,t} + \phi_{c,t} + \delta\chi_{i,c,t} + \epsilon$

Ingroup religious bias is a tight zero

	Outcome variable: Acquittal rate					
	(1)	(2)	(3)	(4)	(5)	(6)
Non-Muslim judge on Muslim defendant	0.008	0.008	—	0.007	0.006	—
	(0.004)	(0.005)		(0.004)	(0.005)	
Non-Muslim judge on non-Muslim defendant	0.007**	0.007*	—	0.007**	0.006	—
	(0.003)	(0.004)		(0.003)	(0.004)	
Difference = Own religion bias	-0.001	0.000	0.002	-0.001	0.000	0.002
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
Reference group mean	0.18	0.184	0.184	0.181	0.184	0.184
Observations	5655320	5214531	5213019	5668388	5228040	5226225
Demographic controls	No	Yes	Yes	No	Yes	Yes
Judge 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.
Specification: $Y_{i,c,t} = \alpha + \beta_1 judge\_nonmus_{i,c,t} + \beta_2 def\_nonmus_{i,c,t} + \beta_3 judge\_nonmus_{i,c,t} * def\_nonmus_{i,c,t} + \phi_{c,t} + \delta \chi_{i,c,t} + \epsilon$

Contexts that Activate Bias

Identity is fluid, and different contexts can make some identities more salient than others.

- Some of the largest effects in the literature are from contexts that activate identity

Contexts that Activate Bias

Identity is fluid, and different contexts can make some identities more salient than others.

- Some of the largest effects in the literature are from contexts that activate identity

We explore three case subsets that could activate bias, all based on prior studies of judicial bias.

1. Victim identity is opposite to defendant

Contexts that Activate Bias

Identity is fluid, and different contexts can make some identities more salient than others.

- Some of the largest effects in the literature are from contexts that activate identity

We explore three case subsets that could activate bias, all based on prior studies of judicial bias.

1. Victim identity is opposite to defendant
2. Gender of judges ruling on crimes against women

Contexts that Activate Bias

Identity is fluid, and different contexts can make some identities more salient than others.

- Some of the largest effects in the literature are from contexts that activate identity

We explore three case subsets that could activate bias, all based on prior studies of judicial bias.

1. Victim identity is opposite to defendant
2. Gender of judges ruling on crimes against women
3. Muslim / non-Muslim judges during Ramadan

Contexts that Activate Bias

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

Standard errors in parentheses

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

- Ingroup gender bias is not activated, even when victim is from the ingroup.
- Some evidence that religious ingroup bias is activated during Ramadan. Effect size remains small vis-a-vis other studies.

What about Caste?

To proxy for caste similarity, we create a binary variable indicating judge and defendant share a family name

- Imperfect proxy:
    - Incorporates religion/family as well as caste
    - Some groups overly aggregated (Singh)
    - Some groups overly disaggregated (different last names, same caste)

Effect of Last Name Similarity on Judicial Decisions

	(1) Acquitted	(2) Acquitted	(3) Acquitted	(4) Acquitted	(5) Acquitted	(6) Acquitted
Same last name	-0.000	-0.001	0.014^{**}	0.012^{*}	0.001	-0.001
	(0.001)	(0.001)	(0.006)	(0.006)	(0.004)	(0.004)
Same name * Rare name					0.032^{**}	0.033^{**}
					(0.015)	(0.015)
Observations	2225312	2223403	2225312	2223403	2225312	2223403
Fixed Effect	Court-month	Court-month	Court-month	Court-month	Court-month	Court-month
Judge Fixed Effect	No	Yes	No	Yes	No	Yes
Inverse Group Weight	No	No	Yes	Yes	Yes	Yes
Last Name Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

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

- On average, no bias
- Inverse group weighting: Some groups are advantaged when they match their judge's name

Data

Analysis

Conclusion

Indian judges show little gender or religion in-group bias

Overall, we found little evidence of substantial judicial in-group bias:

- despite significant anecdotal evidence of bias toward women and Muslims in the broader Indian society,

Indian judges show little gender or religion in-group bias

Overall, we found little evidence of substantial judicial in-group bias:

- despite significant anecdotal evidence of bias toward women and Muslims in the broader Indian society,
- and despite such bias found in almost all other papers on the topic.

Indian judges show little gender or religion in-group bias

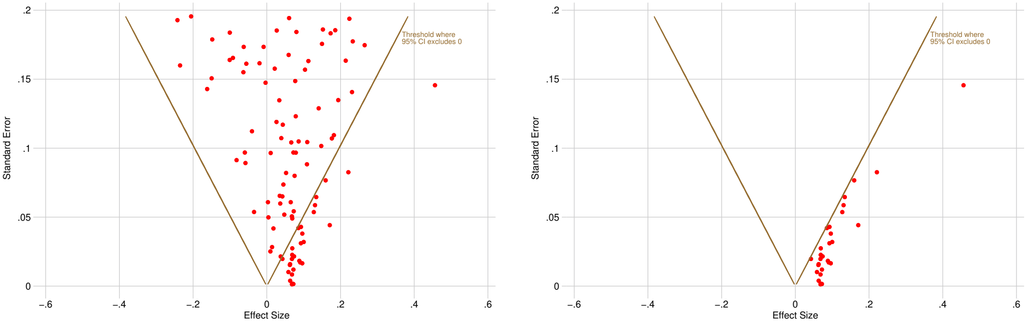
Overall, we found little evidence of substantial judicial in-group bias:

- despite significant anecdotal evidence of bias toward women and Muslims in the broader Indian society,
- and despite such bias found in almost all other papers on the topic.

We did find bias in some (but not all) areas where identity is particularly salient

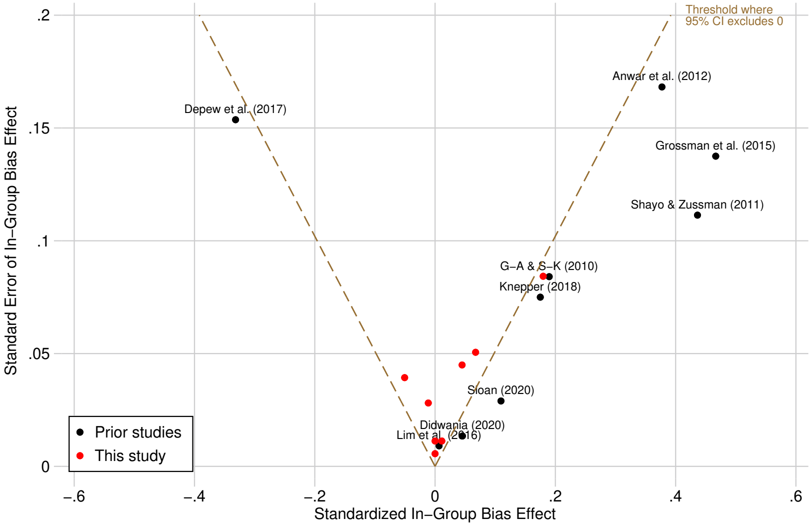
- Even here, it was sparse and small in magnitude.

Funnel / Pyramid Plot and Publication Bias



Notes: This figure plots reported effect magnitudes (Y axis) against effect standard errors. All effect sizes are standardized (outcome variables/standard deviation) to allow comparison.

Comparison with judicial in-group bias estimated in other settings



Notes: This figure plots reported effect magnitudes (Y axis) against effect standard errors. All effect sizes are standardized (outcome variables/standard deviation) to allow comparison.

Formal Test of Publication Bias (Andrews and Kasy 2021)

Table 1: Estimates of Publication Bias in Judicial In-Group Bias Studies

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

Notes: The table summarizes in-group bias in the judicial setting, measured across all papers we could find using randomized assignment of judges and juries, with adjustment for publication bias. Columns 1–4 respectively show the probability that a study gets published, given a t-statistic in the range of $(-\infty, -1.96]$, $(-1.96, 0]$, $(0, 1.96]$, and $(1.96, \infty)$ respectively. β^* in Column 5 gives the true predicted average in-group bias effect after taking publication bias into account and imputing unpublished studies.

- Studies with statistically insignificant positive estimates are only 3% as likely to be published as studies with statistically significant results.
- When adjusting for publication bias by imputing missing studies, the predicted true effect size is 0.046 (Column 5), a fraction of the average observed effect size of 0.24 from the published studies.

Conclusion

- We reject meaningful in-group bias at the judicial outcome stage, but we cannot rule out that the criminal justice system is biased as a whole.

Conclusion

- We reject meaningful in-group bias at the judicial outcome stage, but we cannot rule out that the criminal justice system is biased as a whole.
- **Future research can address:**
  1. Is bias at earlier stages of criminal process?
  2. Bias in higher courts where judges' discretion may be greater?
  3. Can we go deeper on the caste / income dimension?

Thanks!

elliottash.com
ashe@ethz.ch