Incremental AI

Judicial Corpora

U.S. Circuit Courts

- All 380K cases, 1M judge votes, from 1870-
- 2B 8-grams, 5M citation edges across cases

U.S. District Courts

- 1M criminal sentencing decisions
- 2.5M opinions from 1923-

U.S. Supreme Court

- Speech patterns in oral arguments from 1955-
- Identical introductory sentences

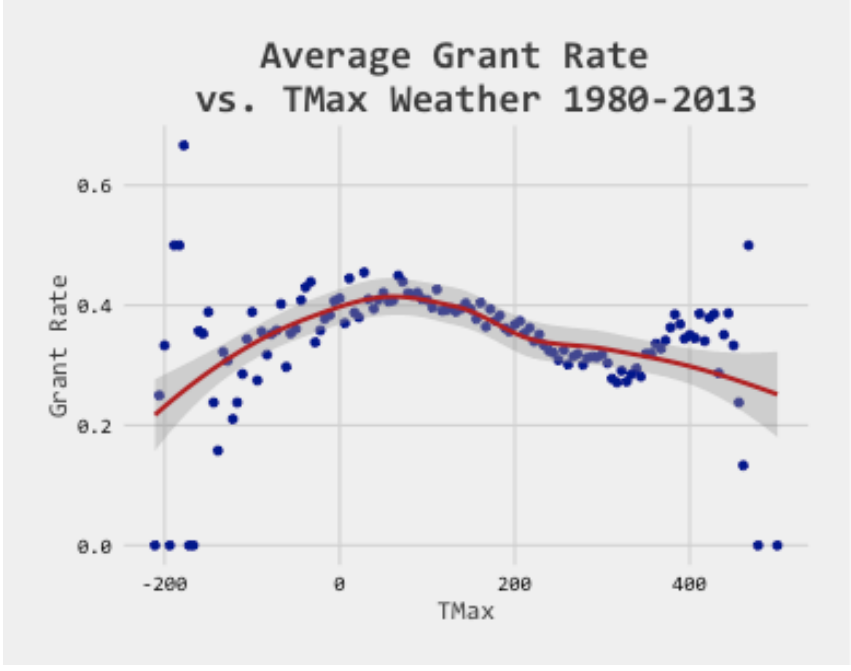
U.S. Immigration Courts

Prosecutors

WW1 Courts martials

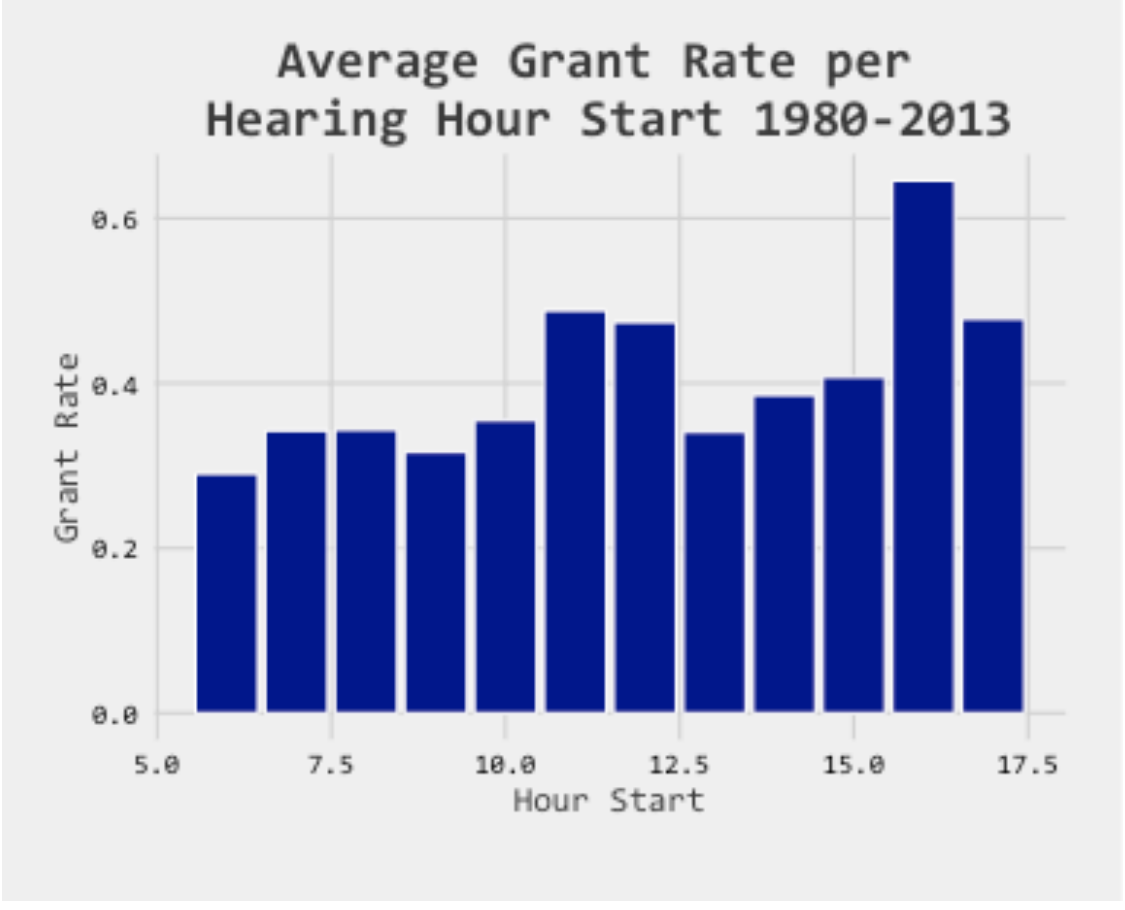
The weather

Judges deny refugees asylum when the weather is too hot or too cold



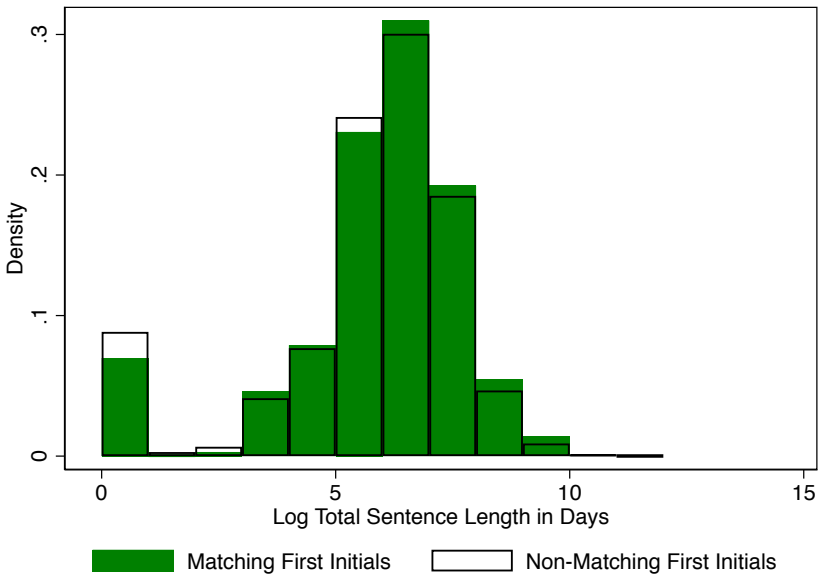
Time of Day

They grant asylum more before lunch and less after.

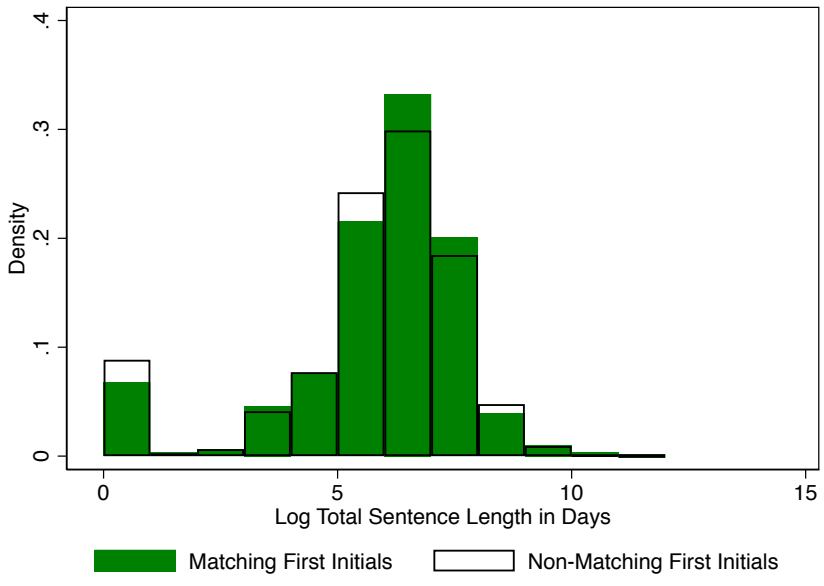


The defendant's name

They assign longer sentence lengths to defendants whose first initial matches their own.



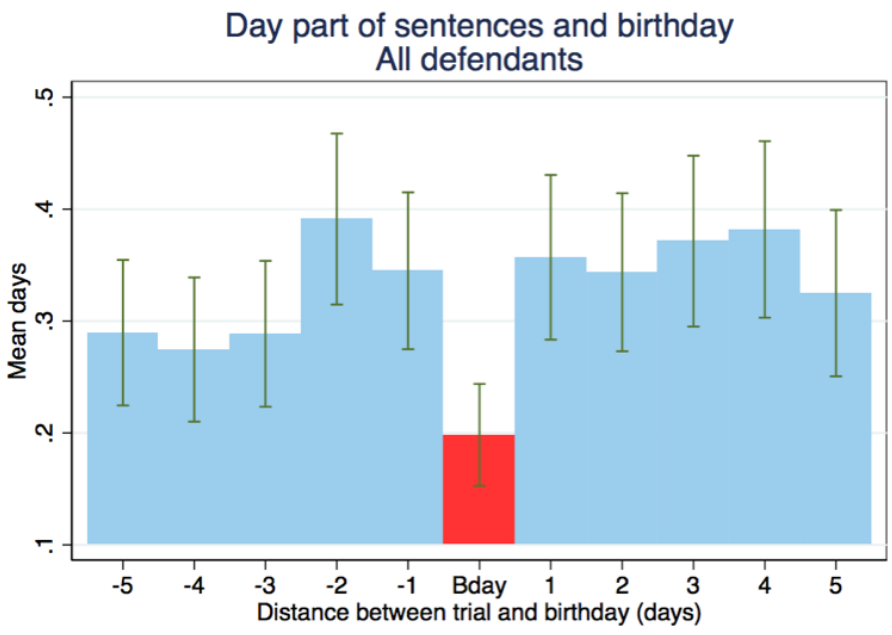
First Letter of First Name



First Letter of Last Name

The defendant's birthday

When they do the opposite and give the gift leniency



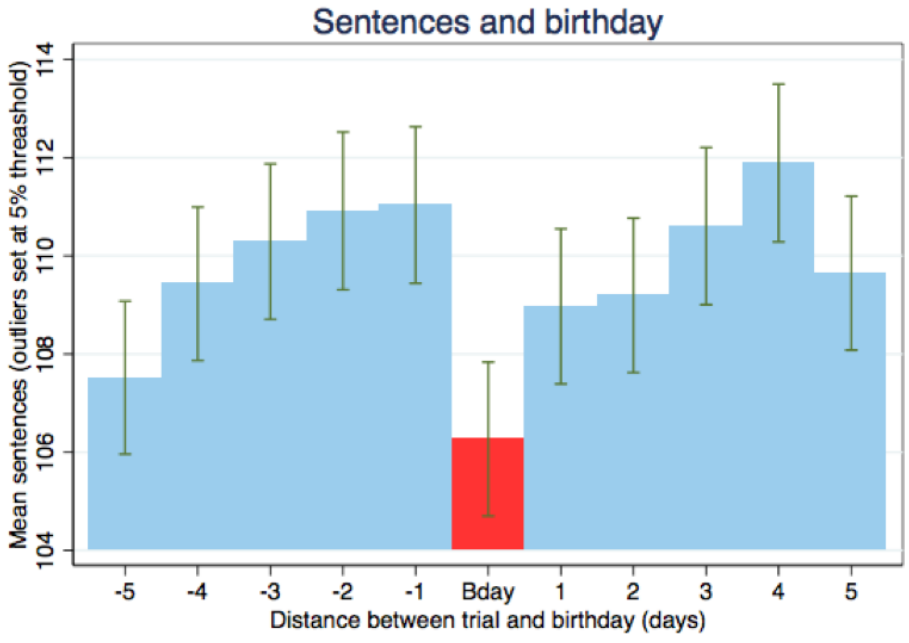
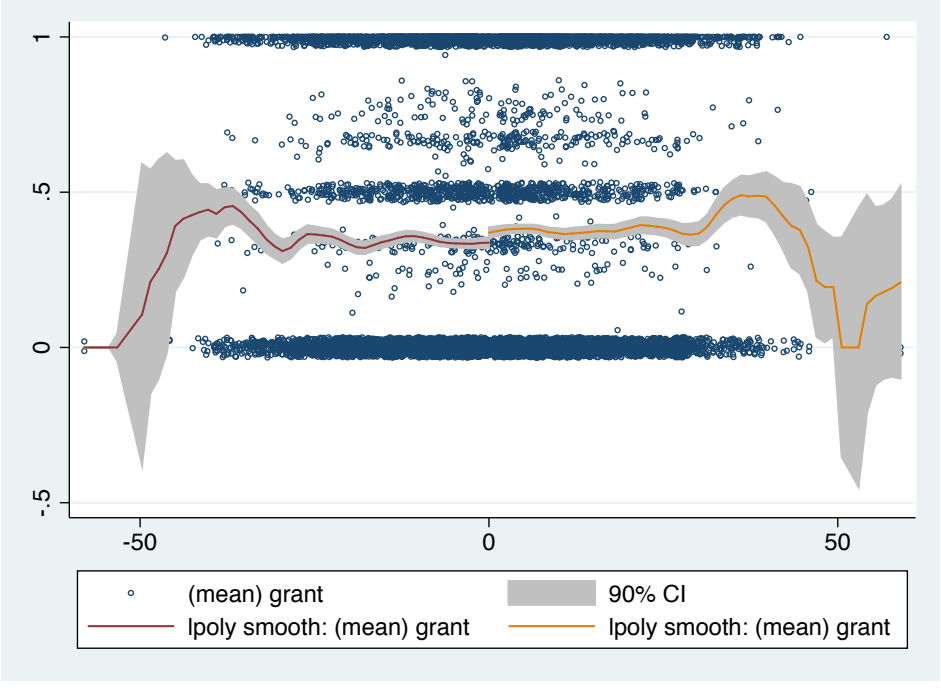


Figure: US and French judicial leniency on defendant birthdays

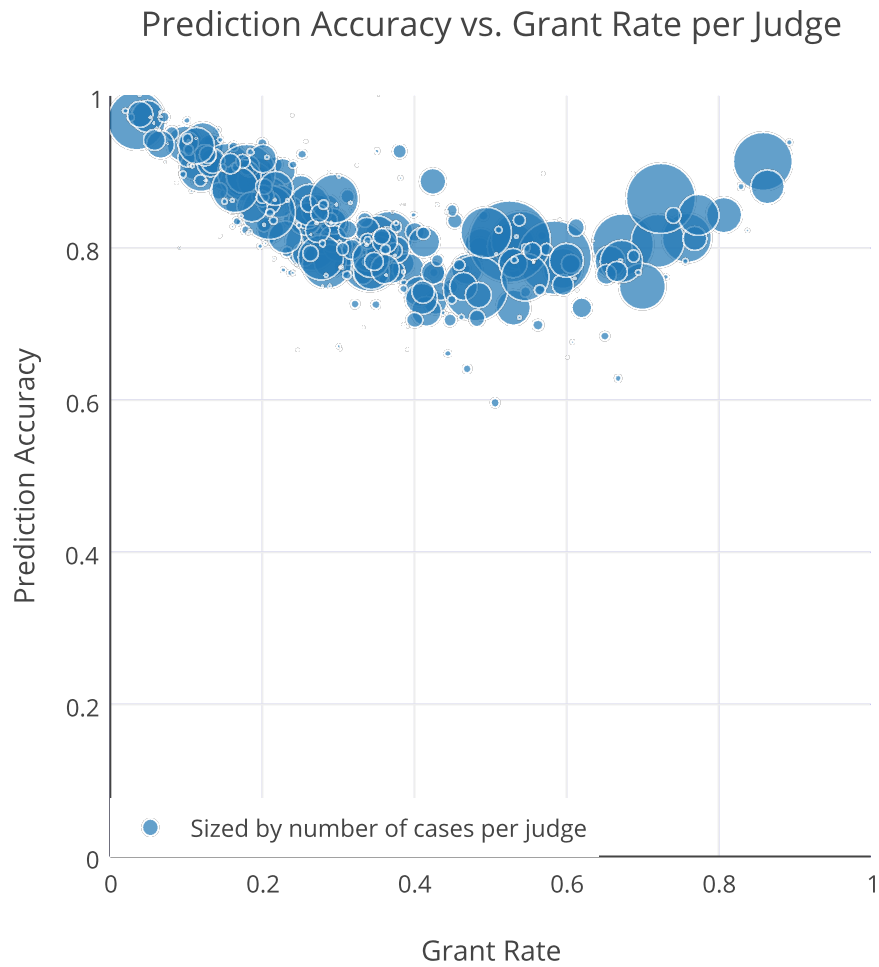
NFL Football

Judges are more lenient the day after their team wins, rather than loses.

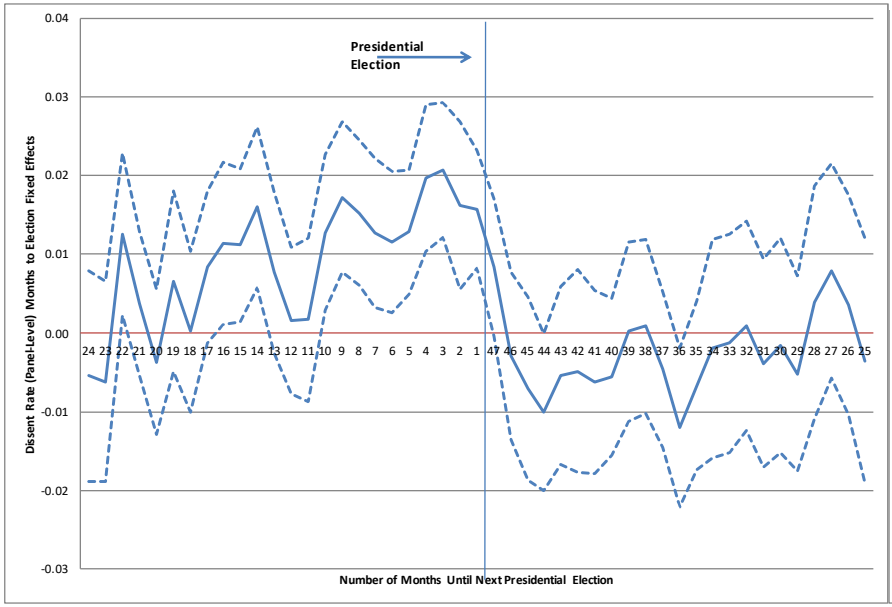


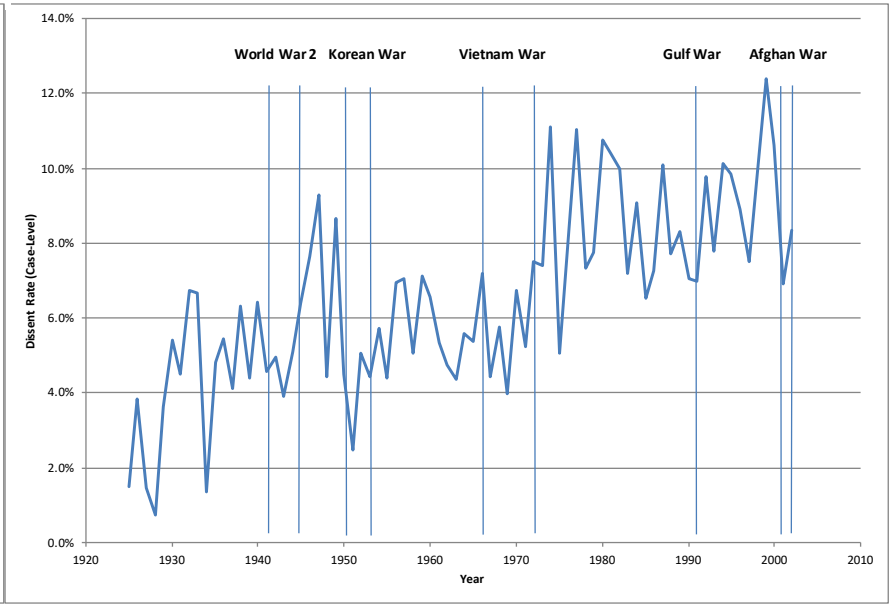
Snap judgments

We can use machine learning to predict asylum decisions with 80% accuracy the date the case opens.. and when it closes.



Elections and wartime also affect decisions





Gambler's Fallacy

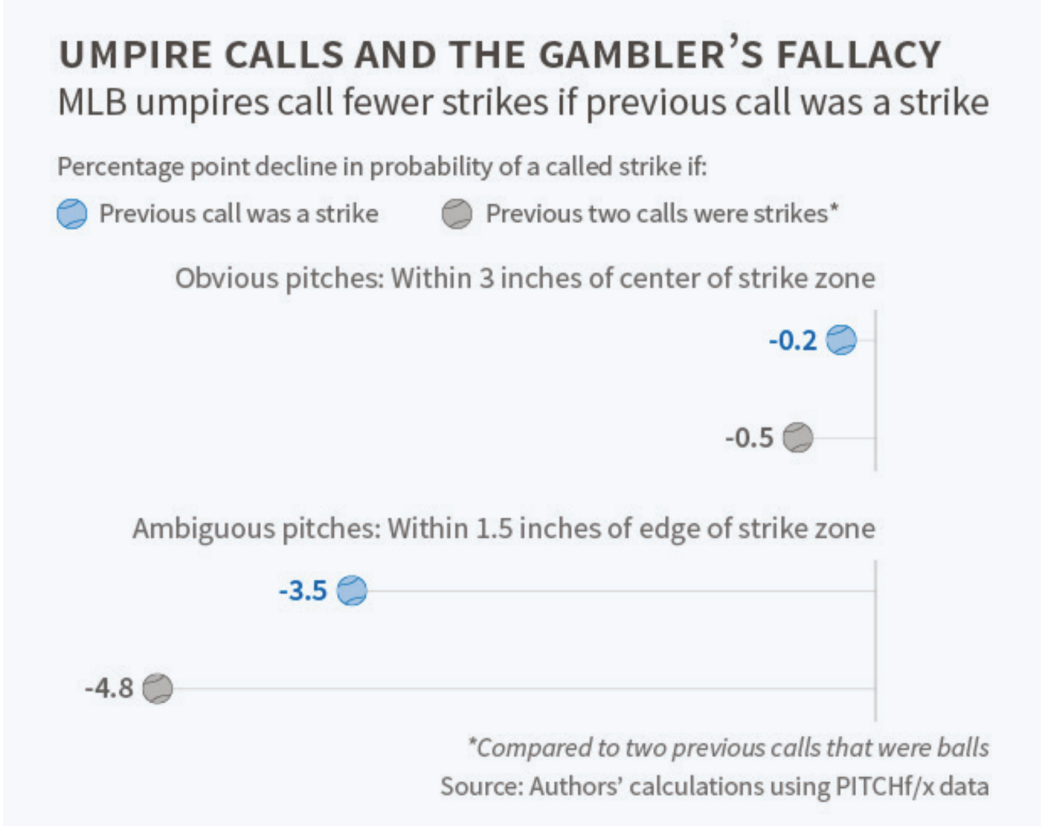
How people often imagine a sequence of coin flips:

0101001011001010100110100

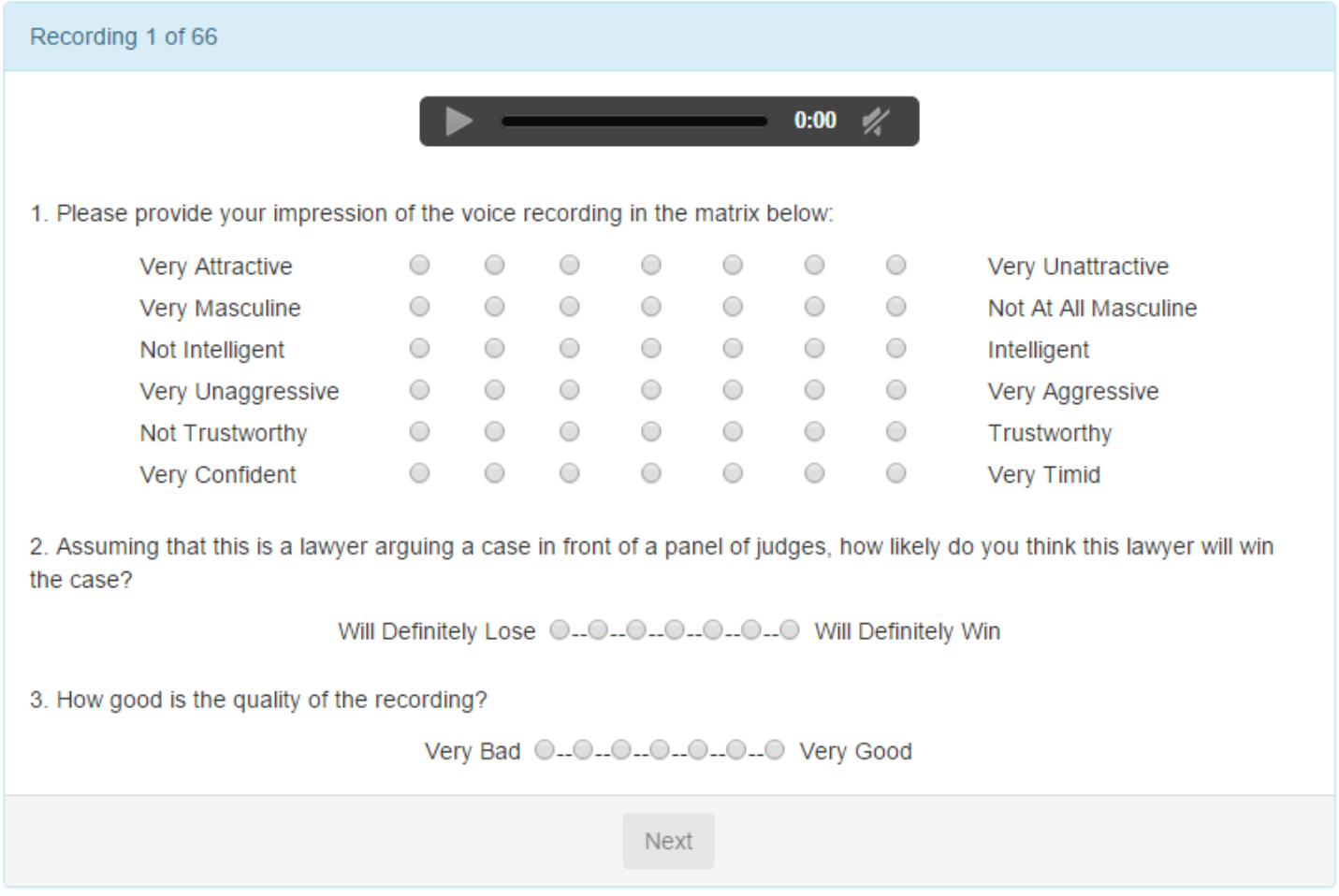
A real sequence of coin flips:

0101011111011000001001101

Up to 5% of decisions reversed due to the gambler's fallacy



In the US Supreme Court, the first sentence of the lawyers oral arguments are identical



Recording 1 of 66

0:00

1. Please provide your impression of the voice recording in the matrix below:

Very Attractive	○	○	○	○	○	○	○	Very Unattractive
Very Masculine	○	○	○	○	○	○	○	Not At All Masculine
Not Intelligent	○	○	○	○	○	○	○	Intelligent
Very Unaggressive	○	○	○	○	○	○	○	Very Aggressive
Not Trustworthy	○	○	○	○	○	○	○	Trustworthy
Very Confident	○	○	○	○	○	○	○	Very Timid

2. Assuming that this is a lawyer arguing a case in front of a panel of judges, how likely do you think this lawyer will win the case?

Will Definitely Lose ○--○--○--○--○--○--○ Will Definitely Win

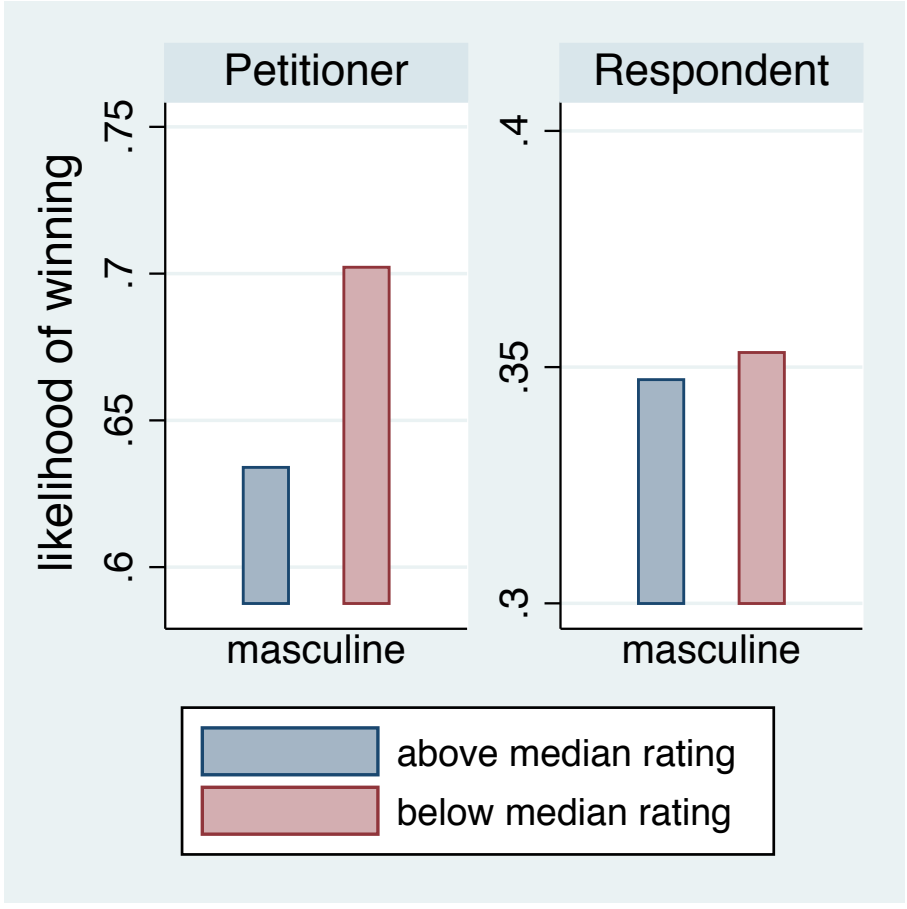
3. How good is the quality of the recording?

Very Bad ○--○--○--○--○--○--○ Very Good

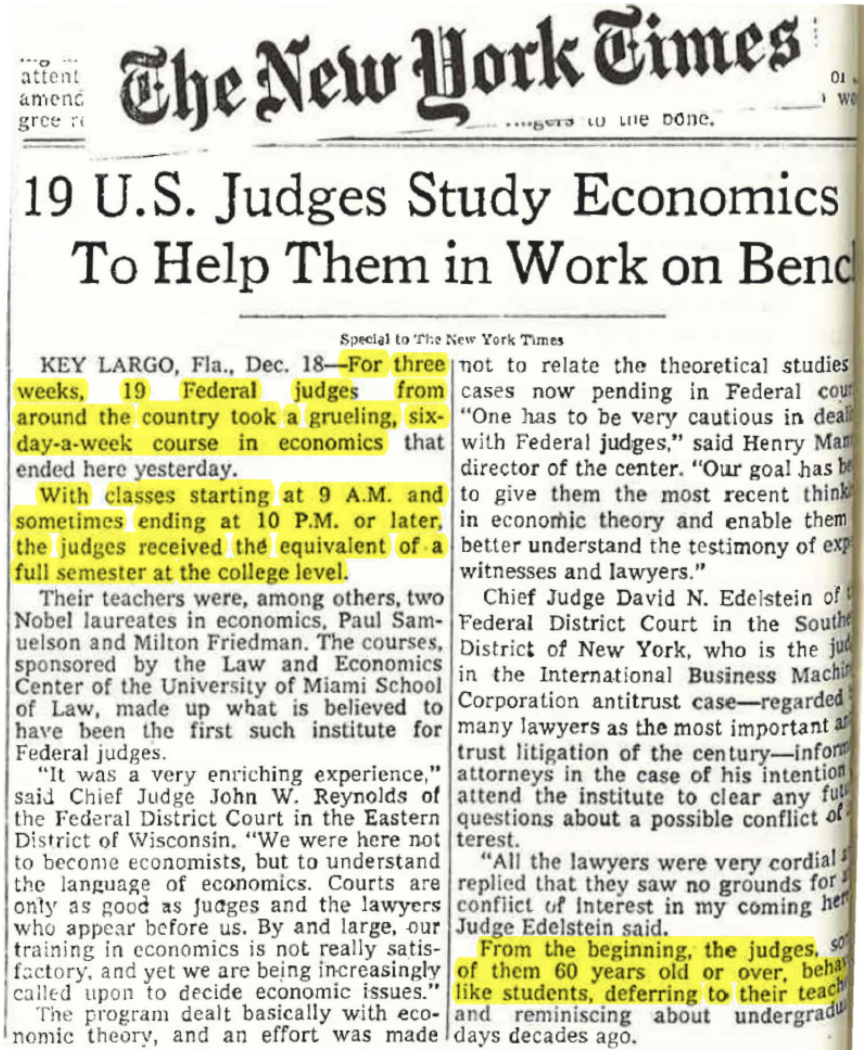
Next

"Mr. Chief Justice, (and) may it please the Court?"

Male petitioners below median in masculinity rating are 7 percentage points more likely to win



By 1990, 40% of federal judges had attended an economics-training program.



The New York Times

19 U.S. Judges Study Economics To Help Them in Work on Benc

Special to The New York Times

KEY LARGO, Fla., Dec. 18—For three weeks, 19 Federal judges from around the country took a grueling, six-day-a-week course in economics that ended here yesterday.

With classes starting at 9 A.M. and sometimes ending at 10 P.M. or later, the judges received the equivalent of a full semester at the college level.

Their teachers were, among others, two Nobel laureates in economics, Paul Samuelson and Milton Friedman. The courses, sponsored by the Law and Economics Center of the University of Miami School of Law, made up what is believed to have been the first such institute for Federal judges.

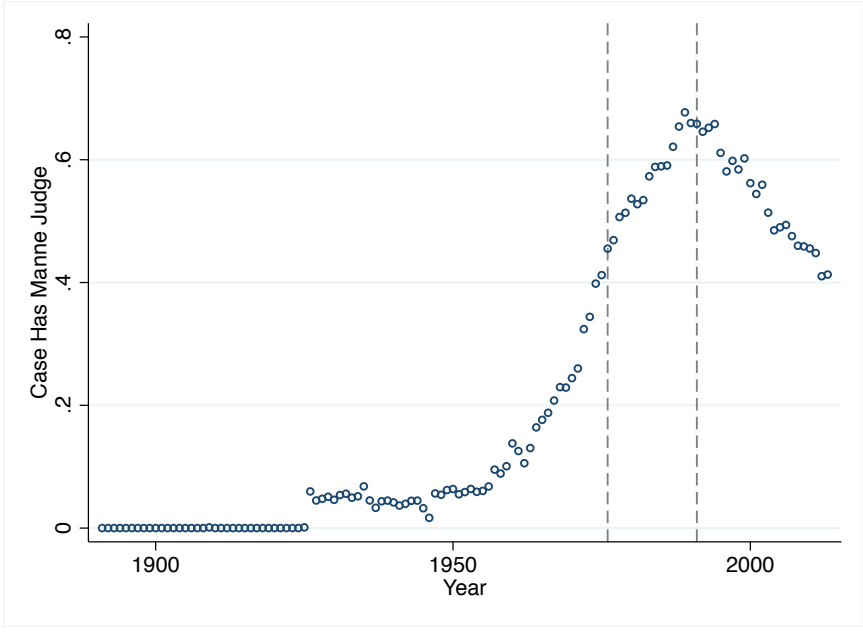
"It was a very enriching experience," said Chief Judge John W. Reynolds of the Federal District Court in the Eastern District of Wisconsin. "We were here not to become economists, but to understand the language of economics. Courts are only as good as judges and the lawyers who appear before us. By and large, our training in economics is not really satisfactory, and yet we are being increasingly called upon to decide economic issues."

The program dealt basically with economic theory, and an effort was made not to relate the theoretical studies cases now pending in Federal cou "One has to be very cautious in deali with Federal judges," said Henry Man director of the center. "Our goal has be to give them the most recent thinki in economic theory and enable them better understand the testimony of exp witnesses and lawyers."

Chief Judge David N. Edelstein of t Federal District Court in the Southe District of New York, who is the jud in the International Business Machi Corporation antitrust case—regarded many lawyers as the most important a trust litigation of the century—inform attorneys in the case of his intention attend the institute to clear any fut questions about a possible conflict of terest.

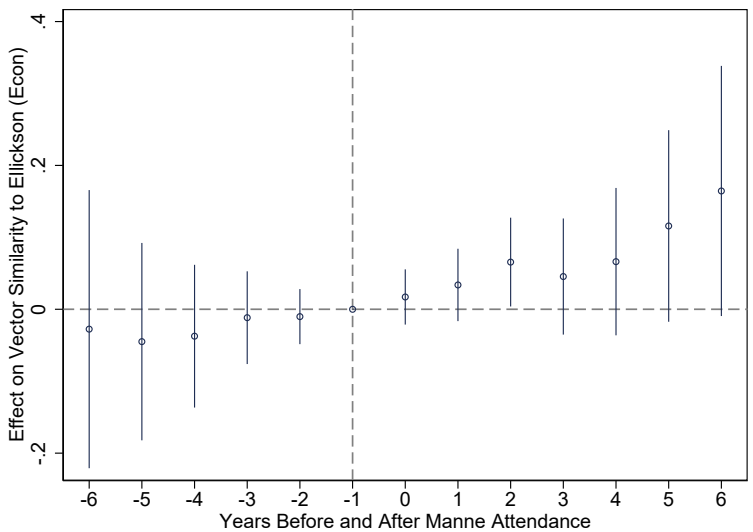
"All the lawyers were very cordial a replied that they saw no grounds for conflict of interest in my coming her Judge Edelstein said.

From the beginning, the judges, so of them 60 years old or over, behav like students, deferring to their teach and reminiscing about undergradu days decades ago.



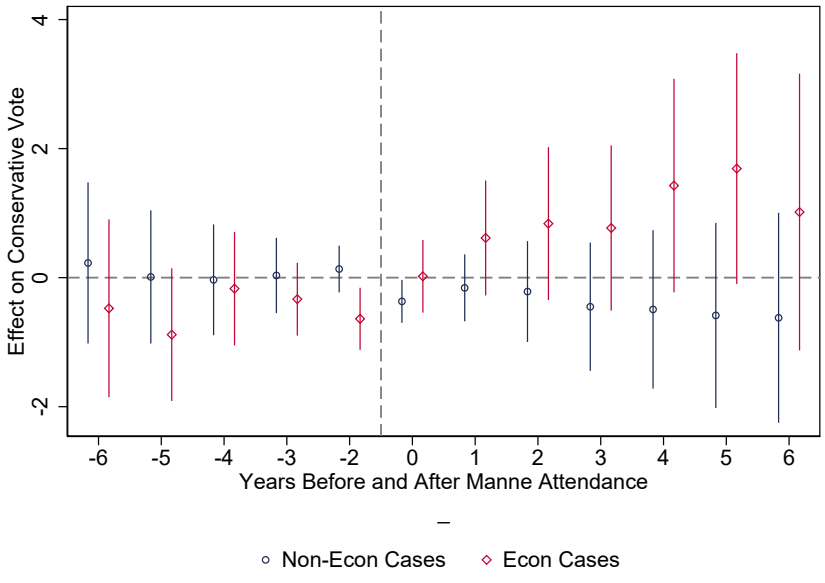
The results of these seminars were dramatic

We can see economics language used in academic articles became prevalent in opinions.

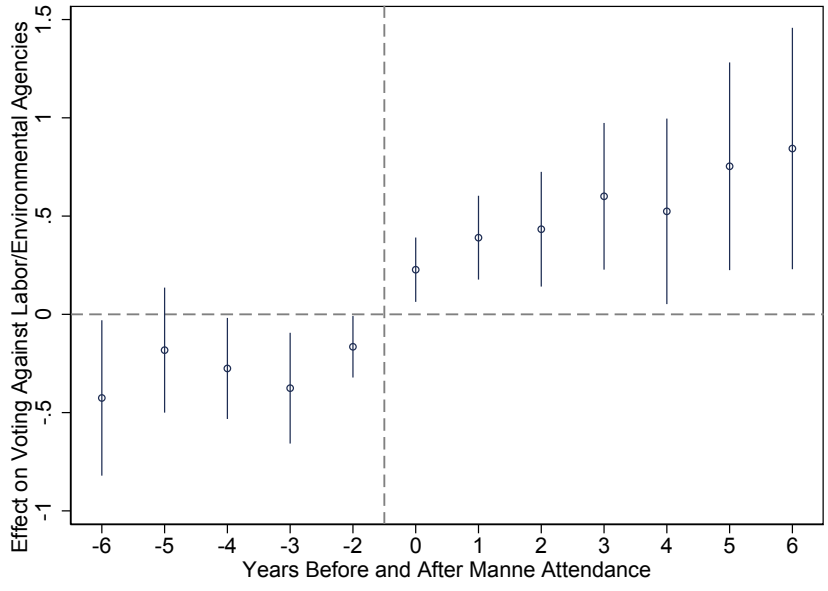


The results of these seminars were dramatic

We can see economics trained judges changing how they decided



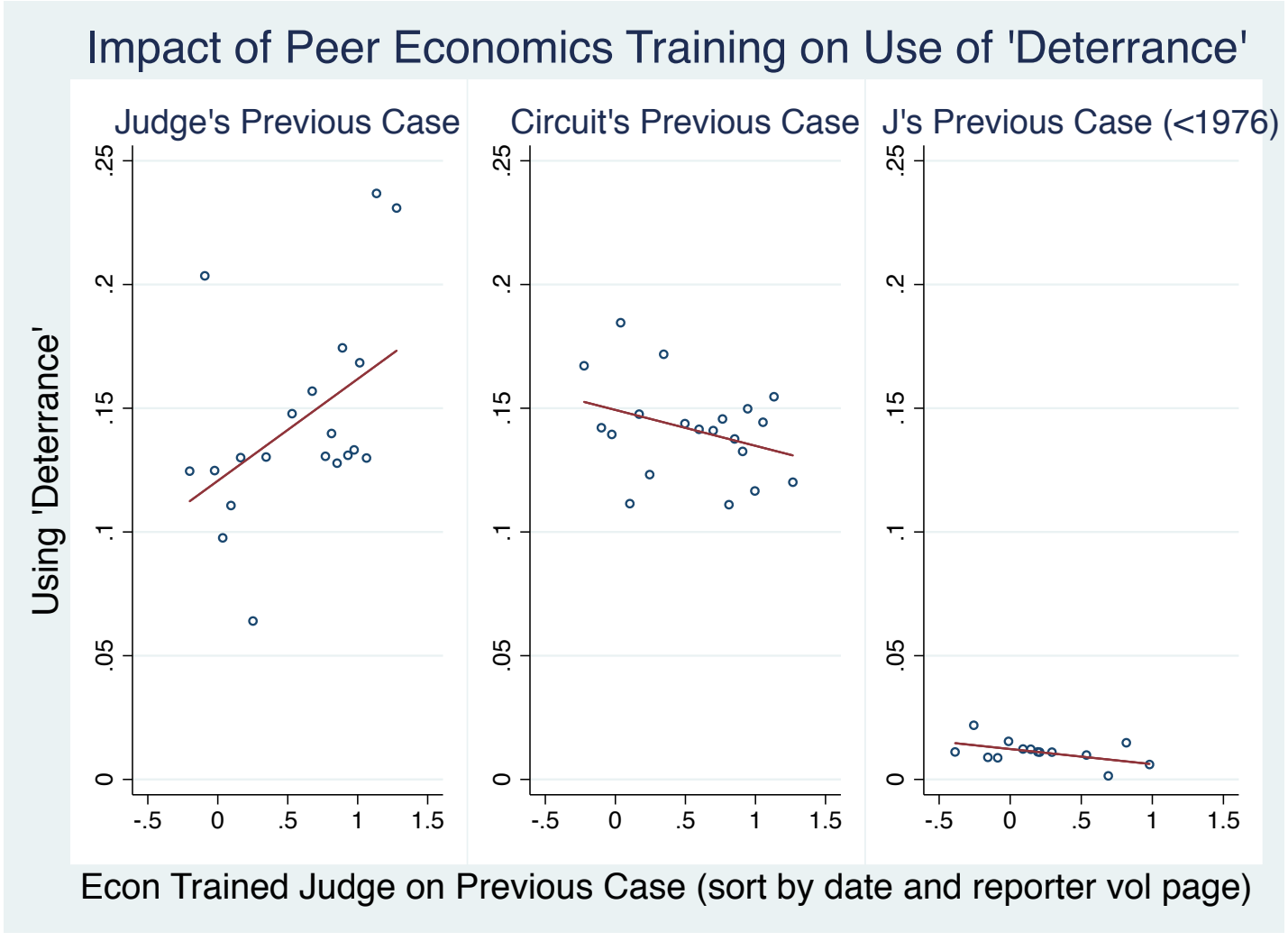
Econ vs Non-Economics Cases



on Labor/Environmental Cases

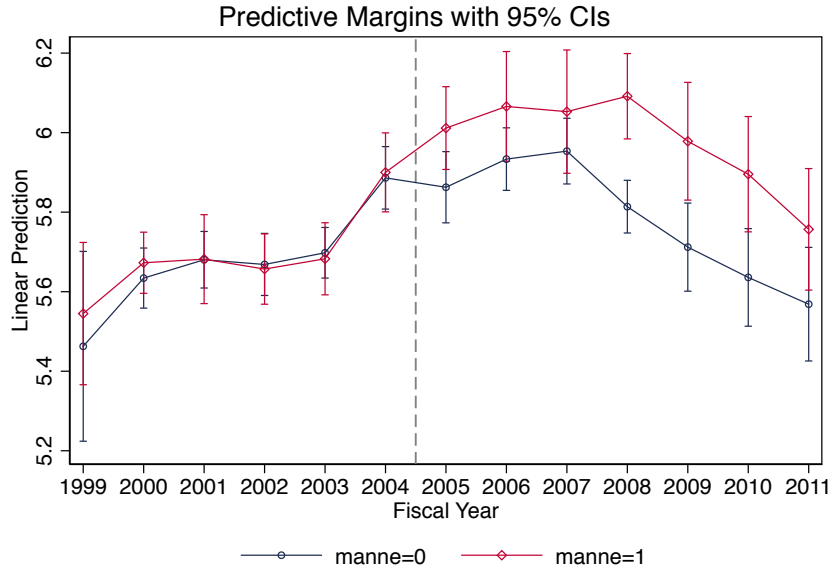
Impacting their peers

We can see economic language traveling from one judge to another and across legal areas.



When judges were given discretion in sentencing

economics trained judges immediately rendered 20% longer sentences relative to the non-economics counterparts.



Incremental AI

- Backlash to AI vs. Incremental AI
- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
- In Stage 1, people use AI as a support tool, speeding up existing processes (for example, by prefilling forms)
- Once they're used to this, they can more easily accept an added functionality (Stage 2) in which AI becomes a choice monitor, pointing out choice inconsistencies (pay more attention / be less indifferent)
- Stage 3 elevates the AI to the role of a more general coach, providing outcome feedback on choices and highlighting decision patterns.
  - Transparent + explanable | explain why deviate
- Then, in Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
- Only in Stage 5, recommend the 'optimal decision'

Incremental AI

- In Stage 0, assess judges vs. a bootstrapped judge (predicted decision-maker)
  - using retrospective historical data
  - virtual testing (background) vs. humans
- Stage 4, the AI brings in other people's decision histories and patterns, serving as a platform for a community of experts.
  - distribution of others' predictions
  - appeal prediction
  - expert coding
- Stage 6, WebMD for litigants, increasing access to justice
- Stage 7, experts advised it helps train novices (who tend to make more mistakes)
- Stage 8, use feedback in stage 3B as recommender system with A|B testing to generate tailored causal inference

"Ferguson and the Violence of Indifference" (Cuenca 2017)

The New York Times

Before an Arrest, Officers Tossed a (Virtual) Coin

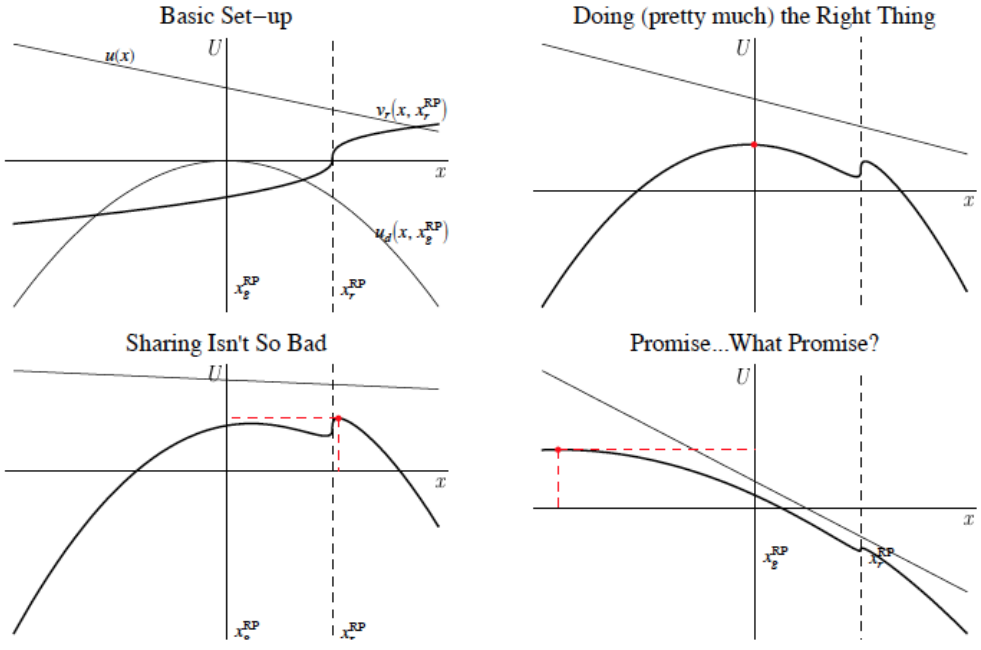
July 14, 2018

Justice: **equal treatment before the law** $(y = f(X) + \varepsilon, a \to X)$
equality based on recognition of difference
$(y \perp W, var(\varepsilon) \perp W, a \nrightarrow W)$

control principle and **merit** principle: individuals liable only for events that are under their control
W: race, gender, masculinity, name, football, weather, judge's lunchtime, preceding case, ...

Judicial Inattention

- Behavioral anomalies offer intuitive understanding of feature relevance
- "settings where people are closer to *indifference* among options are more likely to lead to detectable effects [of behavioral biases] outside of it." (Simonsohn, JPSP 2011)



A model of recognition-respect and

revealed preference indifference

Using ML to Diagnose Judicial Inattention

1. Early predictability
2. Behavioral anomalies
3. If systematic indifference, judge identity might predict appeal
4. Inattentiveness to appellate reversals
5. Implicit risk rankings of asylees closer to random
6. Is indifference greater for some refugees (e.g., from Global South)?
7. Can we use judicial analytics to increase recognition & dignity?

Early Predictability

Early Predictability of Asylum Decisions *Chen, Dunn, Sagun, Sirin, JCAIL, 2017*

- Gambler's fallacy, mood, time of day, order, age ...
  - highlight fragility of asylum courts
    - ★ "In a crowded immigration court, 7 minutes to decide a family's future" (Wash Post 2/2/14)

- High stakes: Denial of asylum usually results in deportation
  - "Applicant for asylum reasonably fears imprisonment, torture, or death if forced to return to her home country" (Stanford Law Review 2007)

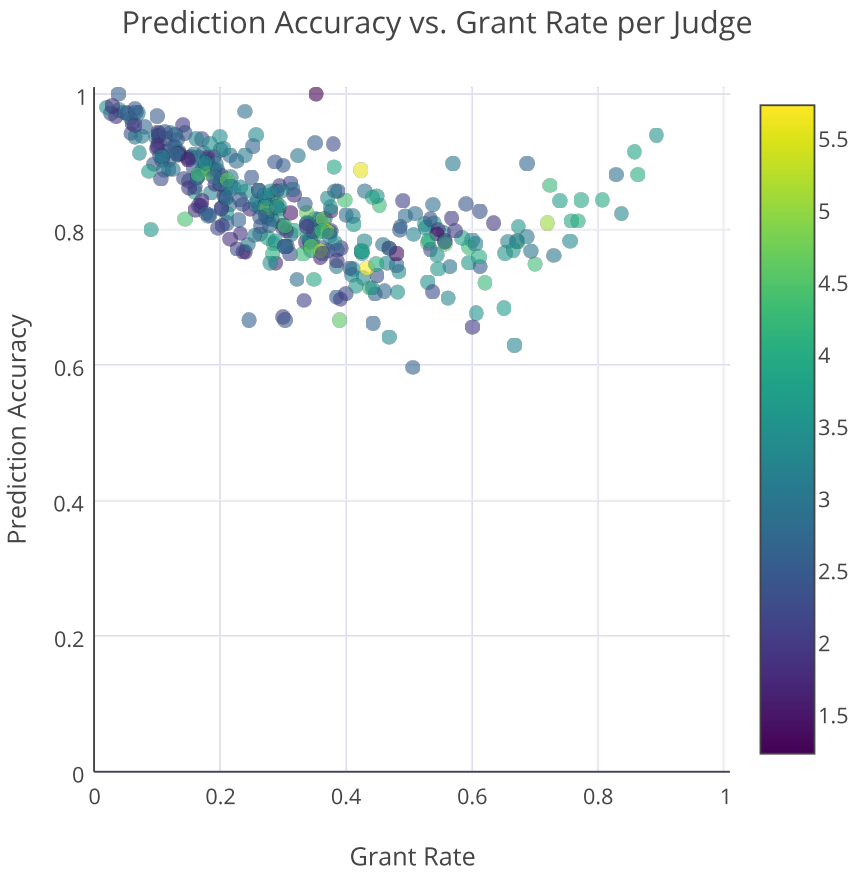
WHAT IS AN AGGREGATE MEASURE OF "REVEALED PREFERENCE INDIFFERENCE"?

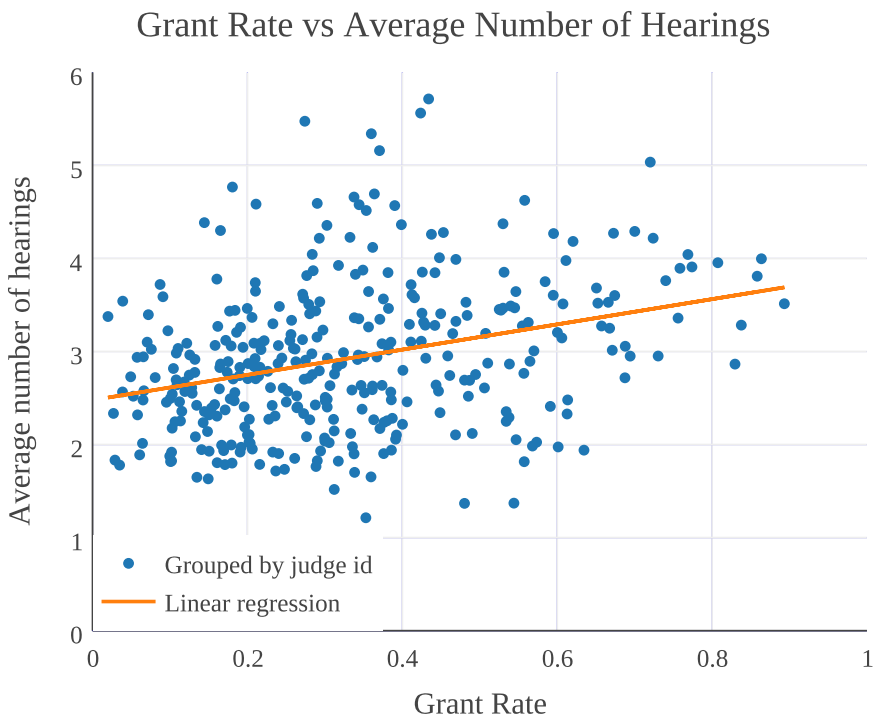
- Using only data available up to the **decision date**, 82% accuracy
  - base rate of 64.5% asylum requests denied
  - predominantly trend features and judicial characteristics - unfair?
  - one third-driven by case, news events, and court information
- Using only data available up to the **case opening**, 78% accuracy

Revealed Preference Indifference

- If case outcomes could be **completely predicted**
  - **prior to judicial inquiry** into the case,
  - then judges **did not take into account** differences between cases
  - (did not recognize-respect defendant's individuality/dignity)
- There may be cases for which country and date of application *should* completely determine outcomes (e.g., during violent conflict)
  - But significant inter-judge disparities in predictability suggest that this understanding of the country circumstances does not apply to all
- Some judges are highly predictable, always granting or rejecting
  - **Snap judgments** and **predetermined** judgments (Ambady and Rosenthal 1993)
  - Stereotypes pronounced with time pressure & distraction (Bless et al 1996)

Early Predictability of Asylum Decisions





Less predictable judges are not simply flipping a coin: hearing sessions are greater for less predictable judges

and for judges with higher grant rates

Early Predictability of Asylum Decisions

Model	Accuracy	ROC AUC
Judge ID	0.71	0.74
Judge ID & Nationality	0.76	0.82
Judge ID & Opening Date	0.73	0.77
Judge ID & Nationality & Opening Date	0.78	0.84
Full model at case completion	0.82	0.88

Variation over time has little additional impact on the outcome of adjuciations.

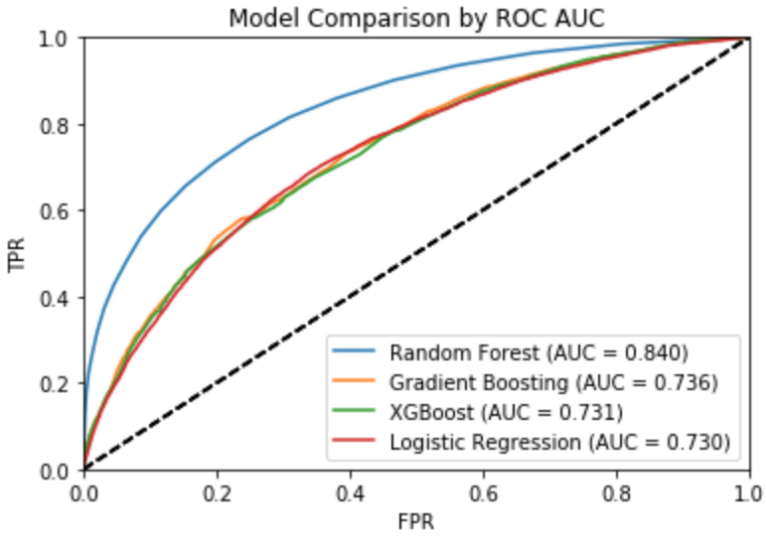
Dataset includes 70 additional features about the hearings

If systematic indifference, judge identity might predict appeal

Predictability of Asylum Appeals?

- We have shown evidence of early predictability that varies by judge
  - We see evidence of behavioral anomalies
- If systematic mistakes, judge identity might predict appeal

Machine Prediction of Appeal Success



Feature Importance	
Time Horizon Features	0.377804
Judge Features	0.277066
Respondent	0.177945
Trend Features	0.074494
Proceeding Features	0.060490
Location Features	0.042636

Another way to see importance of judge identity..

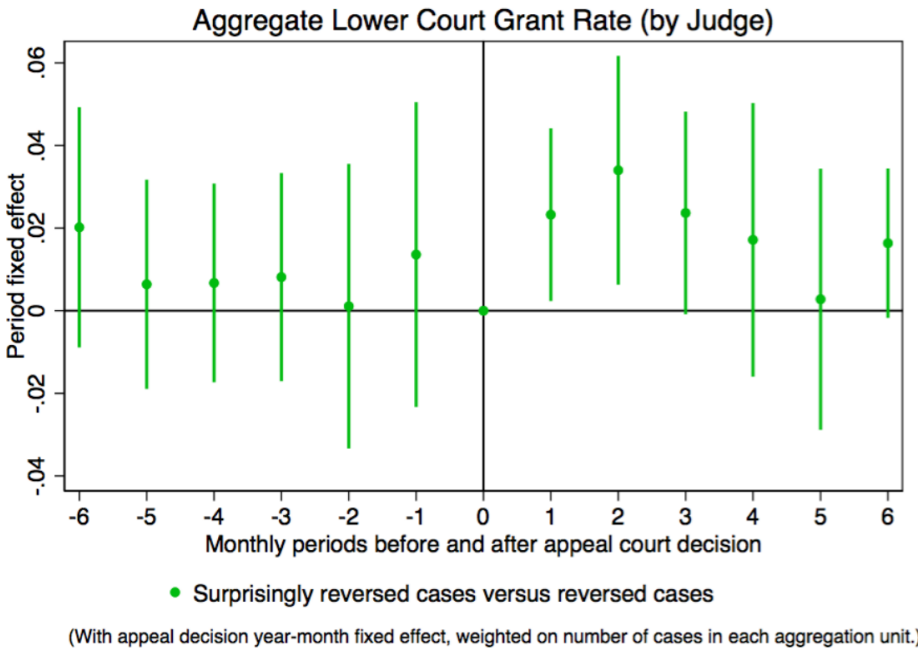
Inattentiveness to Appellate Reversal

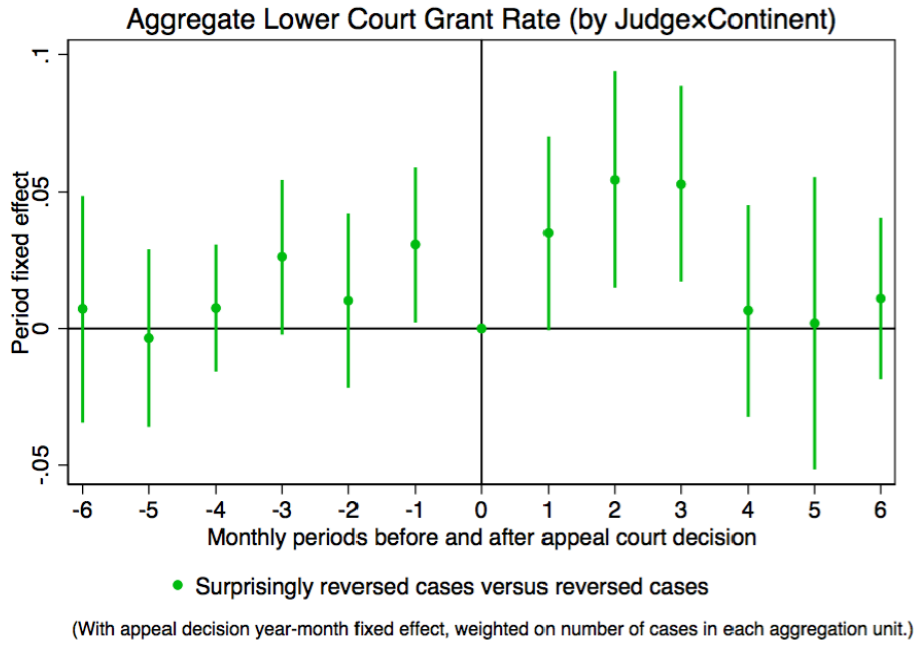
Measuring Inattention

1. Do we see judicial variation in responsiveness to reversal?

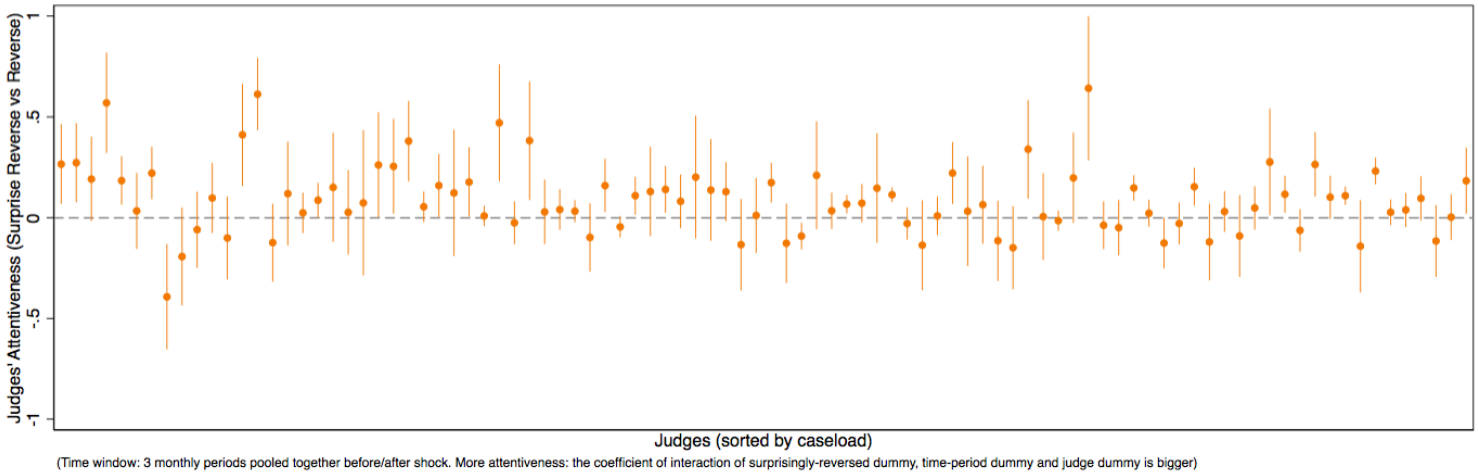
Within-judge $\triangle$grant rates after "surprising" reversals (model predicts affirm)

Effect of "Surprise" Appeal Rulings





Judges Vary in Responsiveness to Reversal



(Time window: 3 monthly periods pooled together before/after shock. More attentiveness: the coefficient of interaction of surprisingly-reversed dummy, time-period dummy and judge dummy is bigger)

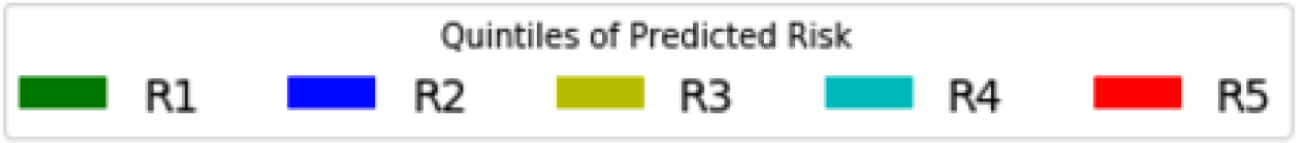
Do implicit rankings by judges differ by attentiveness?

Flipping a coin or.. Having a threshold

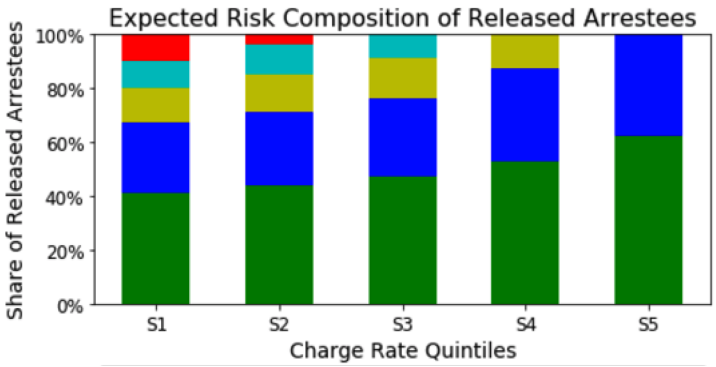
Implicit risk rankings of asylees closer to random?

How the judges rank the risk of asylees is unobserved. But, we can assess their implicit risk ranking by **comparing the distribution of outcomes of the asylees denied by the (randomly assigned) "strict" and the "lenient" judges.**

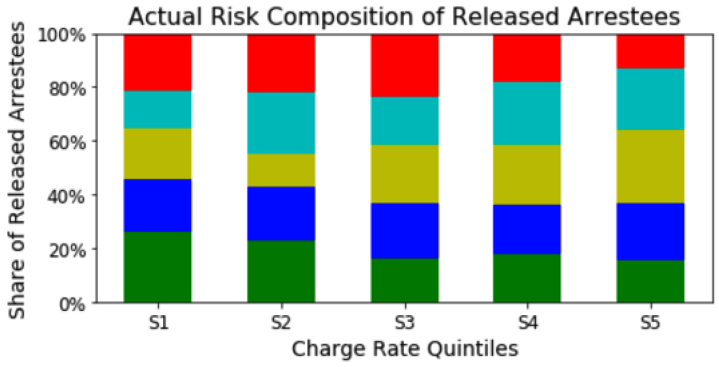
A conceptual example..



Robot Prosecutors



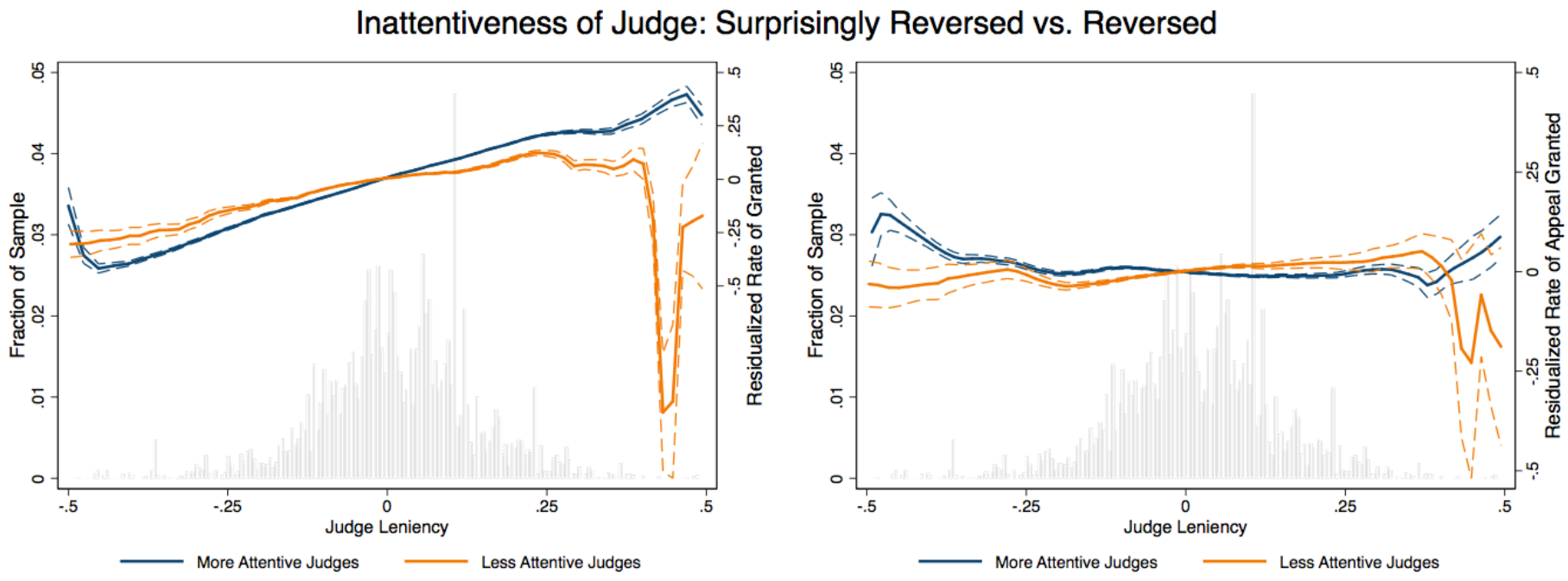
Human Prosecutors



- If defendants released based only on risk score, the harshest prosecutors would only be releasing low-risk defendants.
- Distribution of risk scores for released defendants is similar for most lenient and least lenient prosecutors.

- Are the lenient asylum judges, only denying the 'riskiest' applicants
  - i.e., seeing the lowest reversal rates (of their asylum denials)?

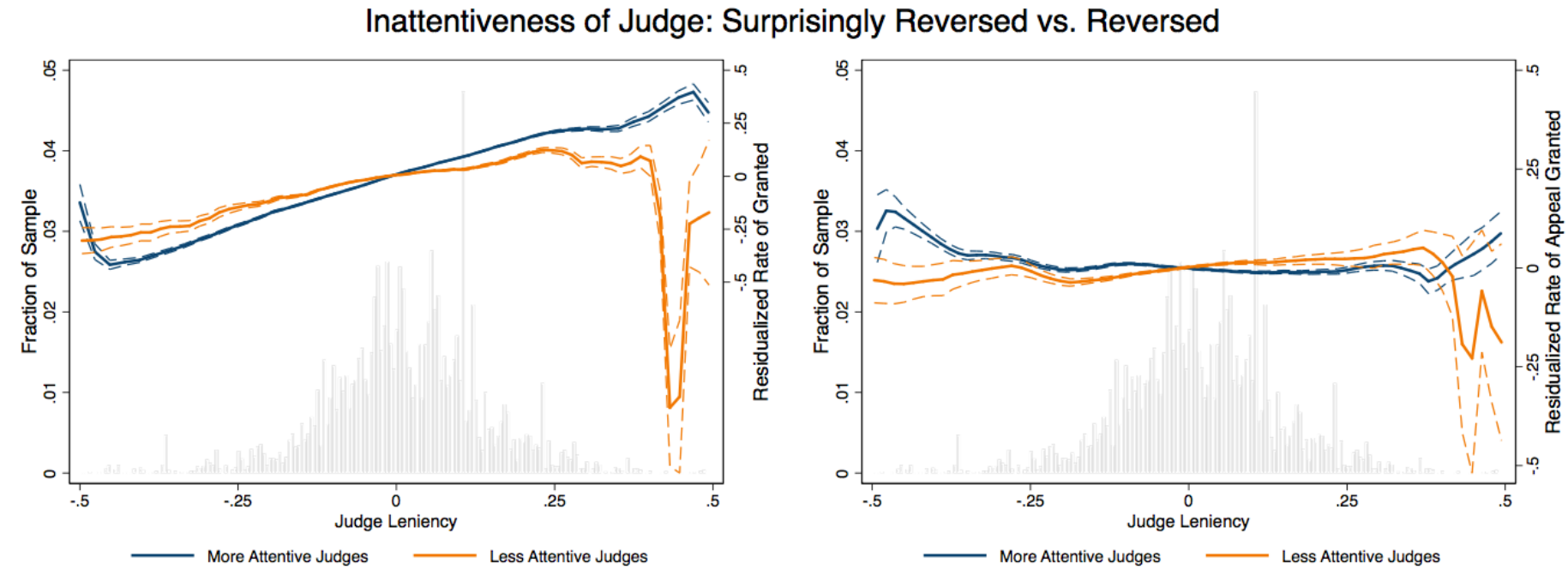
Left figure: Judges have strong habits



(Time window: 3 monthly periods pooled together before/after shock. More attentiveness: the coefficient of interaction of surprisingly reversed dummy and time-period dummy is bigger)

A judge who is generally lenient in other cases is likely to be lenient in a given case.

Right figure: Assess implicit risk ranking

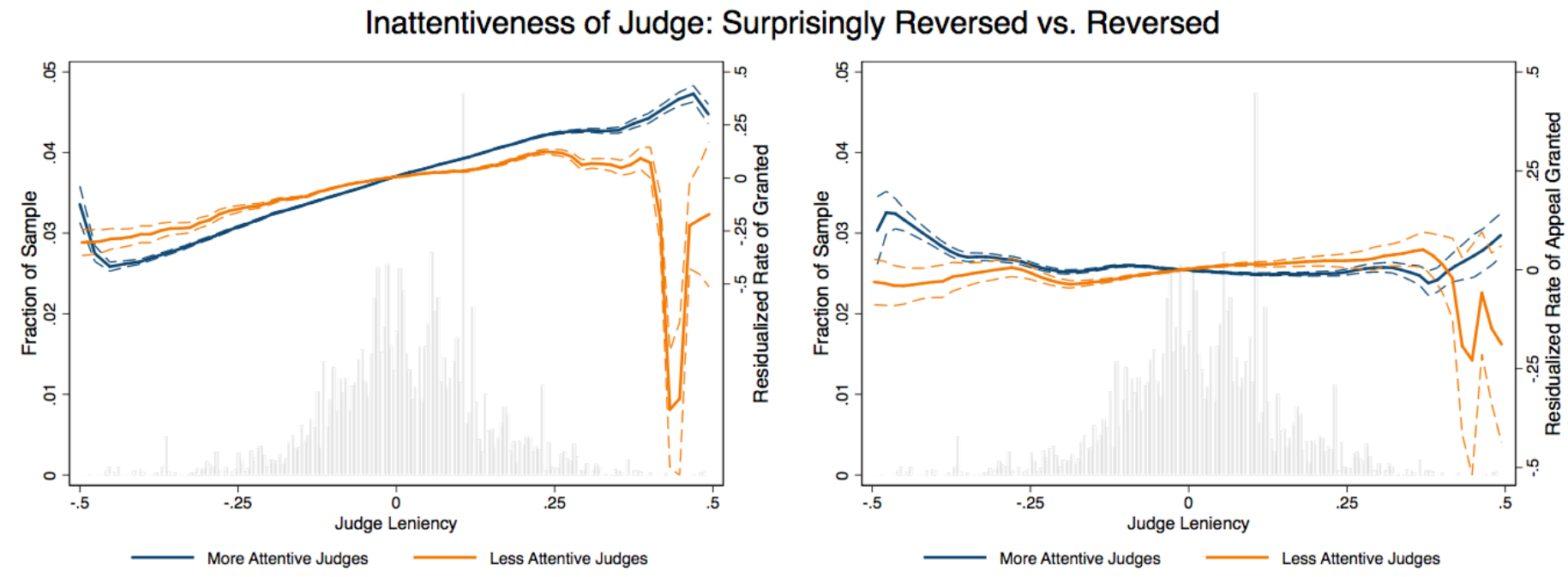


(Time window: 3 monthly periods pooled together before/after shock. More attentiveness: the coefficient of interaction of surprisingly reversed dummy and time-period dummy is bigger)

If judges are 'ordering' their asylees, the most lenient judge letting in the most applicants should be rejecting only the "least safe" applicants.

Their appeal rate should be lower.

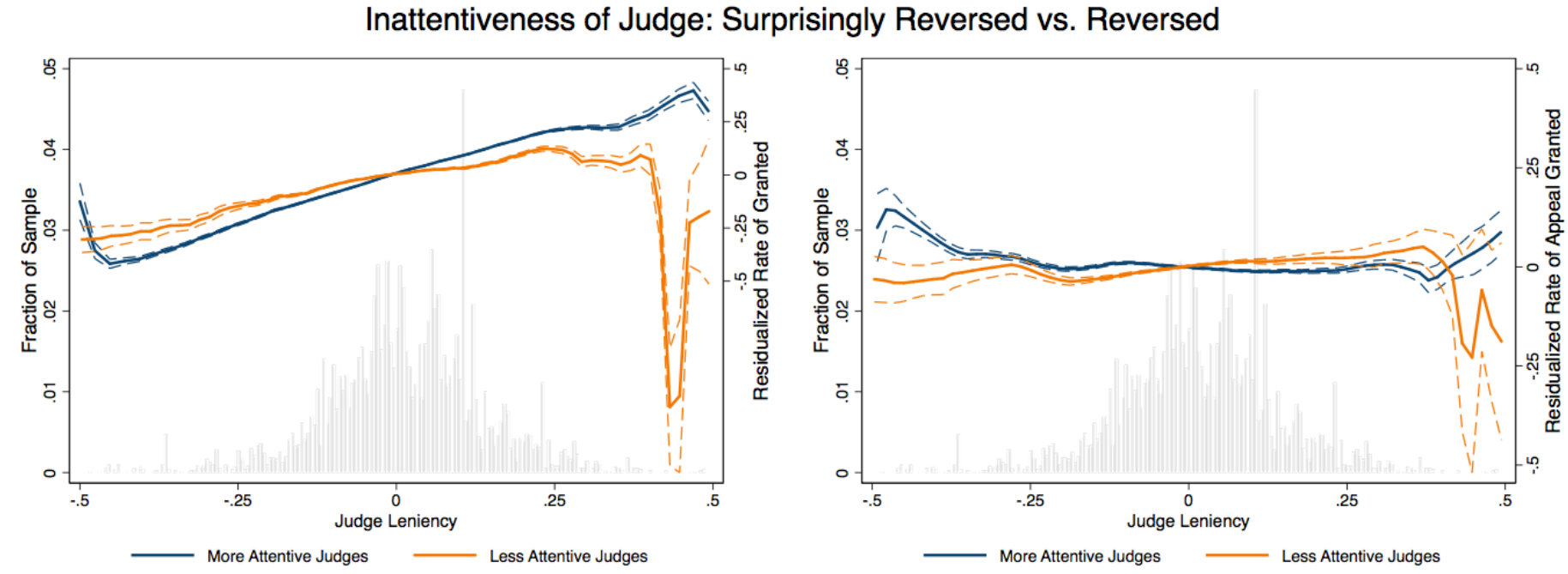
Right figure: Assess implicit risk ranking



(Time window: 3 monthly periods pooled together before/after shock. More attentiveness: the coefficient of interaction of surprisingly reversed dummy and time-period dummy is bigger)

We observe this "diagonal" for the more attentive judges.

Right figure: Assess implicit risk ranking



(Time window: 3 monthly periods pooled together before/after shock. More attentiveness: the coefficient of interaction of surprisingly reversed dummy and time-period dummy is bigger)

.. but not the less attentive judges, potentially more prone to other extraneous factors biasing their decisions

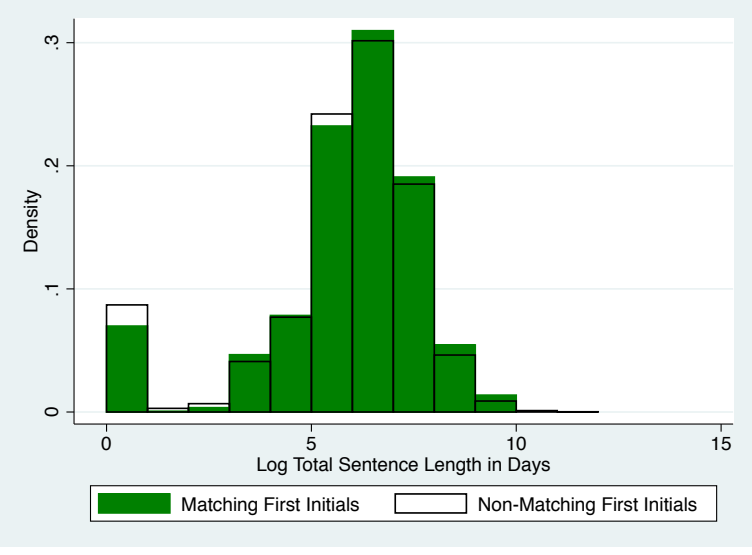
Is indifference greater for some individuals?

Difference in Indifference: Asylum

Dependent variable	Granted Asylum		
Sample	All	With Lawyer	Without Lawyer
	(1)	(2)	(3)
Upset Loss (Loss X Predicted Win)	-0.066***	-0.007	-0.067**
	(0.022)	(0.011)	(0.030)
Upset Loss (Loss X Predicted Win)	0.061**		
X Lawyer	(0.023)		
Close Loss (Loss X Predicted Close)	-0.046**	0.008	-0.045**
	(0.022)	(0.011)	(0.021)
Close Loss (Loss X Predicted Close)	0.054**		
X Lawyer	(0.024)		
Upset Win (Win X Predicted Loss)	-0.023	-0.001	-0.036
	(0.035)	(0.015)	(0.032)
Upset Win (Win X Predicted Loss)	0.020		
X Lawyer	(0.036)		

Unrepresented Parties in Asylum Bear Brunt of Mood Effects

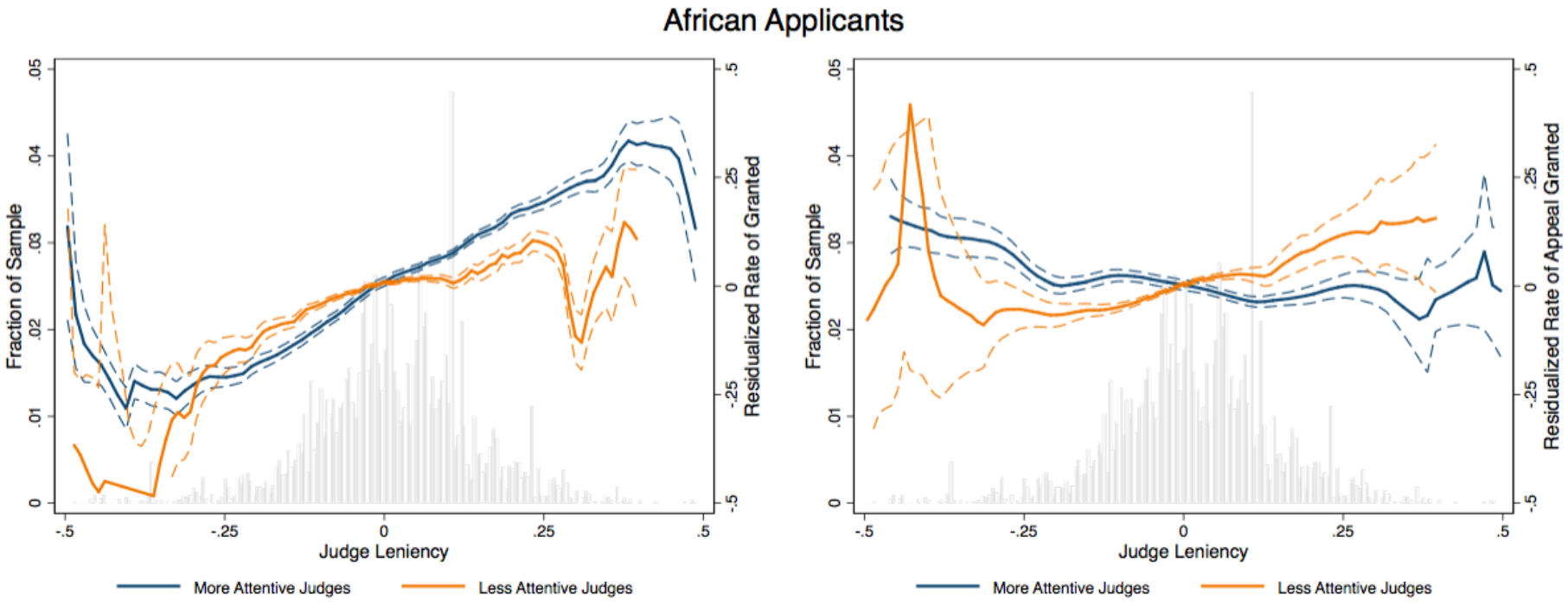
Difference in Indifference: Sentencing



	(1)	(2)	(3)	(4)	(5)
	Log of Total Sentence in Days				
Mean of dep. var.	5.75				
First Letter Match	0.102**	0.0480	0.0308	0.0864	0.0889*
	(0.0442)	(0.0537)	(0.0930)	(0.0477)	(0.0498)
Defendants Sample:	Negro	Not Negro	Black	All	All
Judges Sample:	All	All	All	Black	White
Judge Fixed Effects	Y	Y	Y	Y	Y
Month-Sentence FE	Y	Y	Y	Y	Y
Case Class FE	Y	Y	Y	Y	Y
Observations	33020	15840	10208	13441	35419
R-squared	0.443	0.492	0.539	0.457	0.462

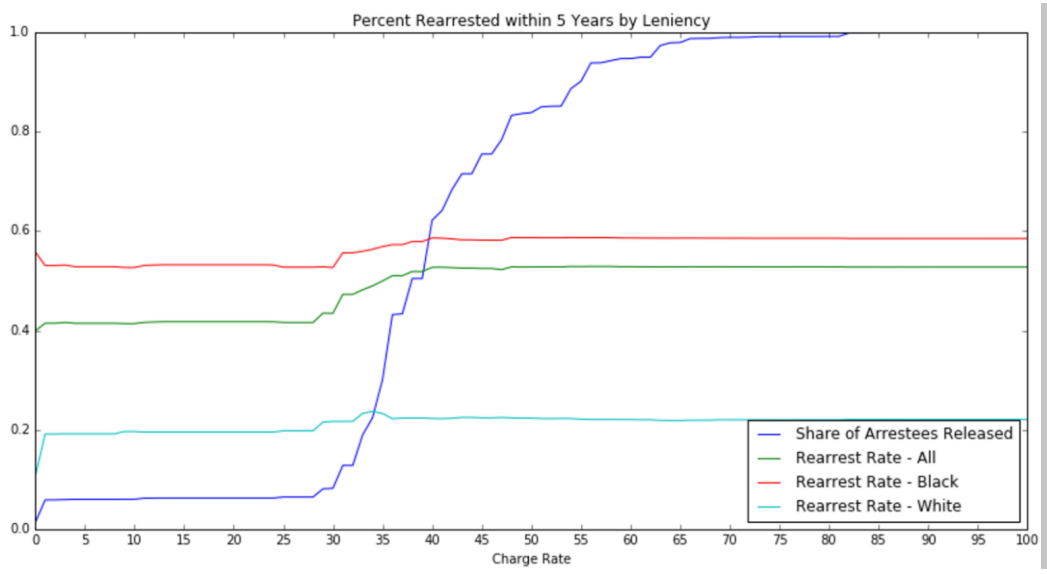
- First letter name effects: 8% longer sentence lengths
- C1: effects are more salient for african americans classified (by police) as "N"
- C3: effects are small and insignificant for those classified as "B"
  - Power of recognition, from new labels obtained through social movements

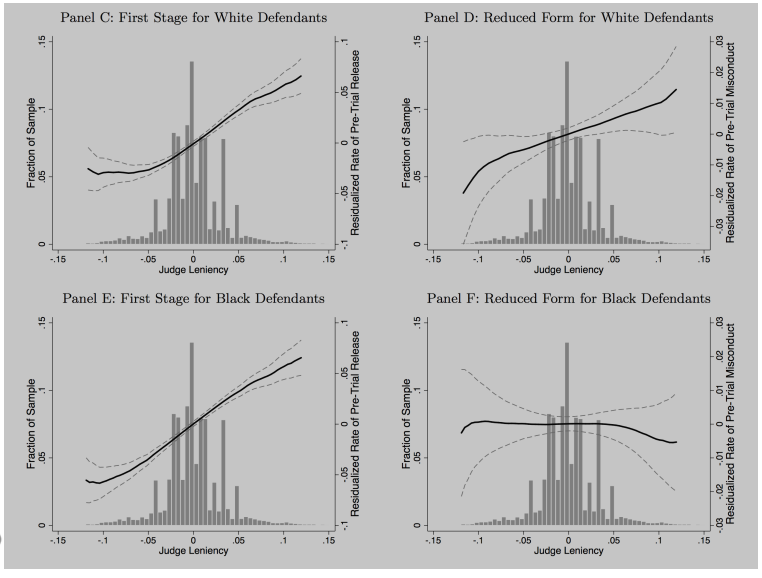
Difference in Indifference: Asylum Risk Rankings



The less attentive judges have a "wrong slope" that appears more inconsistent for applicants from the Global South.

Difference in Indifference: Sentencing Risk Rankings





Likewise, in sentencing, wrong slope for Blacks (left figure)

- Bail judges released along "right" diagonal for Whites but not Blacks (right figure)

Using ML to Diagnose Judicial Inattention

1. Early predictability
2. Behavioral anomalies
3. If systematic indifference, judge identity might predict appeal
4. Inattentiveness to appellate reversals
5. Implicit risk rankings of asylees closer to random
6. Is indifference greater for some refugees (e.g., from Global South)?
7. Can we use judicial analytics to increase recognition & dignity?

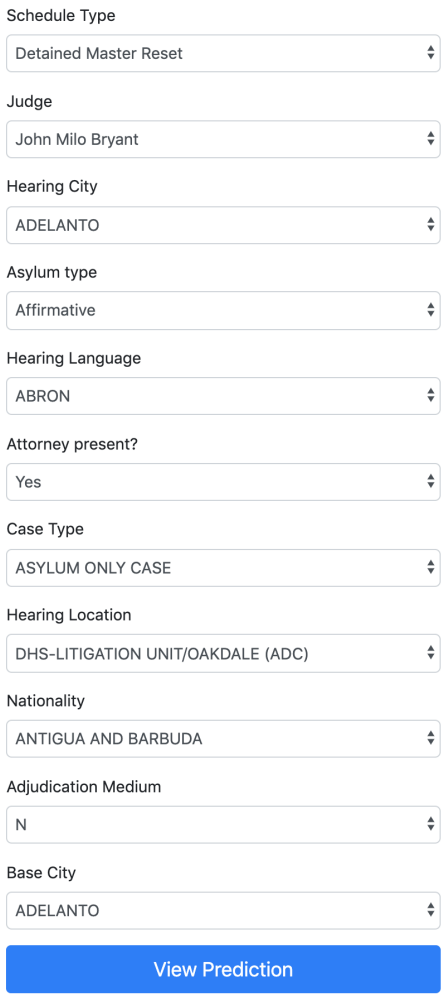
Judicial Analytics for Recognition and Dignity

US Circuit	District	SCOTUS	Asylum	New Orleans DA

India	Kenya	Philippines	Croatia	Czech	Chile / Peru
Implicit Bias	Do behavioral biases replicate?				
In-group Bias	In-group Bias			Interpellation	Impligit Egoism

- Personalized nudges for judges (instead of checklists) to increase justice?
  - Based off recent decisions and environment: "be less indifferent"

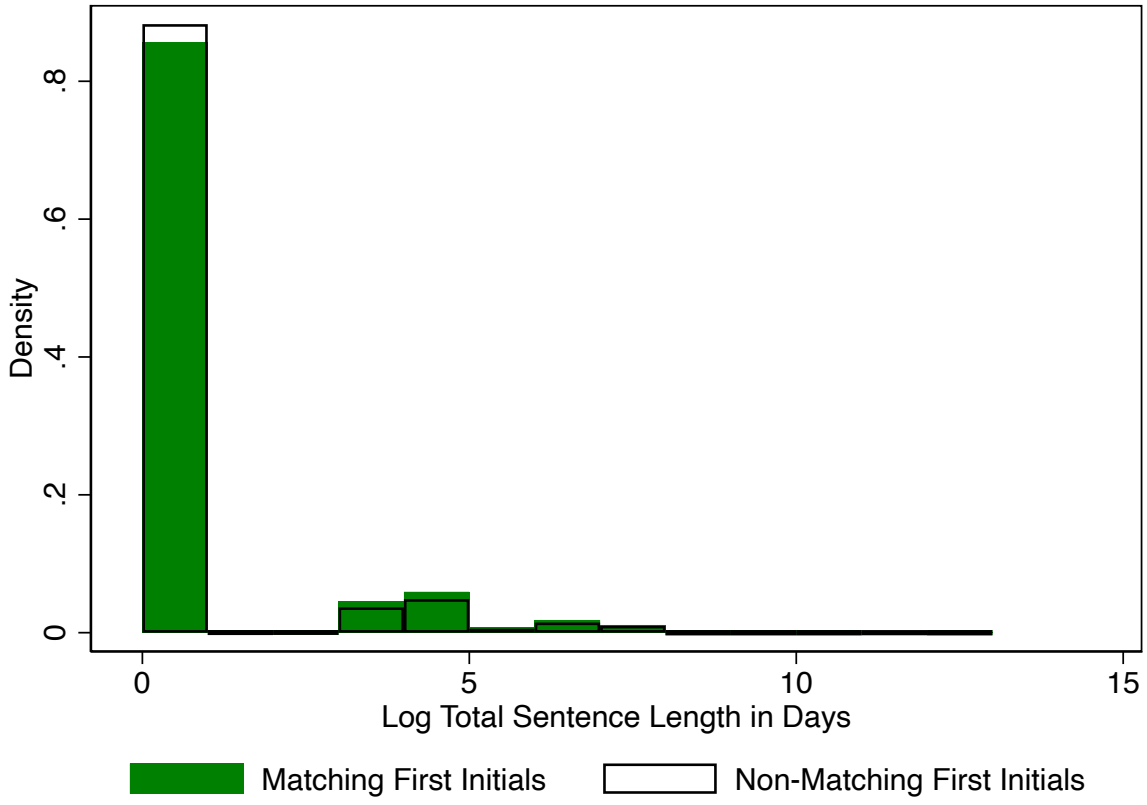
Prediction App (Beta): https://floating-lake-11821.herokuapp.com/



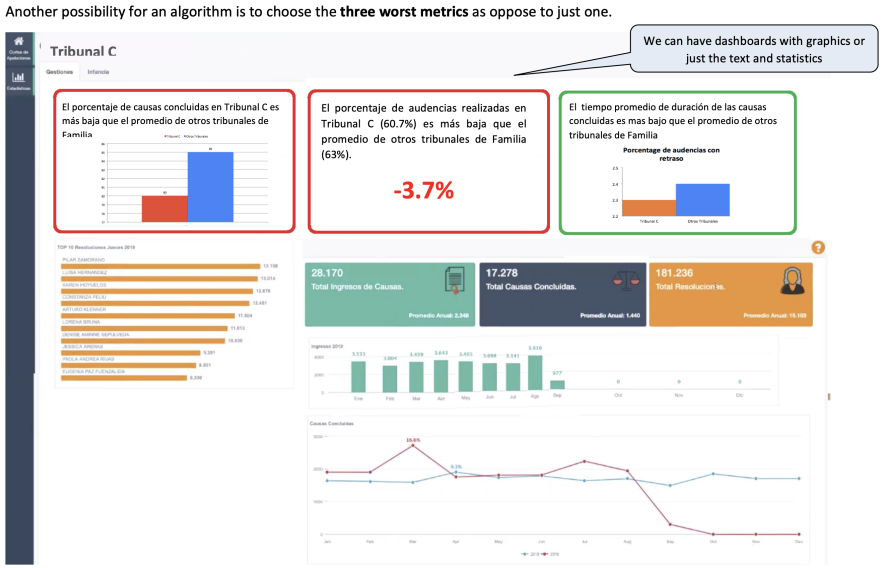
- Assess effects on trust and perceived indifference of lawmakers
- and applications, decisions, reversals, speed, disparities, etc.

Name Effects Replicate in Chile

Judges assign longer sentence lengths to defendants whose first initial matches their own (3.3M sentences)



Chilean App (pilot) to leverage self-image motives



- Rich data ecosystem
  - case-level data for civil cases, criminal cases, and appeals from 2015 on.
  - characteristics of case and parties, outcome, appeal, and reversal if any.
  - Logins to a dashboard used by judges and other court staff to check statistics on their own and peers' performance.
  - Human Resources Data on all 13,368 judiciary employees.
  - Firms' data linkable by tax identifier to cases to assess economic impacts.
  - Court user and staff surveys to assess trust and perceived indifference.

Incremental AI Addresses Common Criticisms of AI in Law

- Potential Bias
  - Stage 0: assess judges vs. a bootstrapped judge (predicted decision-maker)
- Reduced Autonomy
  - Stage 1: use AI as support tool, setting default
- Erosion of Learning
  - Stage 2/4: pointing out when predicted to error + platform of experts
- Transparency
  - Stage 3: interpretable ML for explanability
- Status Quo Bias
  - Stage 3b: AI can ask why user deviates
- Adversarial Attack
  - only shown to judges, not to litigants

- (1) self-image (predicted self), (2) self-improvement (nudges),
- (3) self-understanding (why), (3b) self-expression (explaining), (4) ego (self vs. others)

Data Explorer

- https://explore-ecourts.herokuapp.com/
- App for accountability, level playing field, access to justice