

Building a Robot Judge: Data Science for the Law

1. Introduction

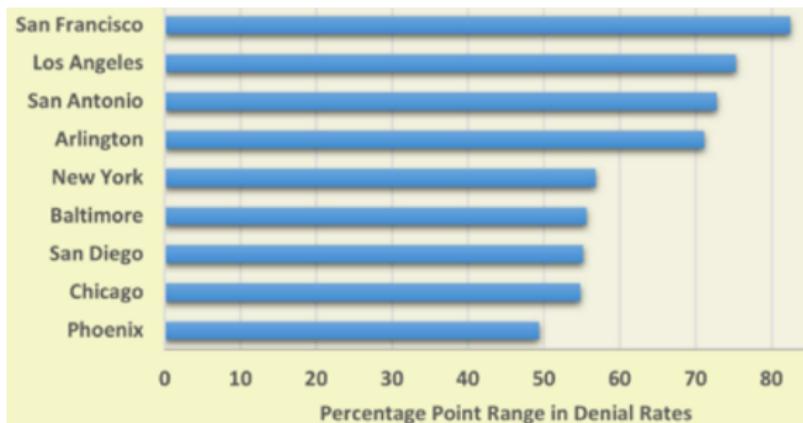
Elliott Ash





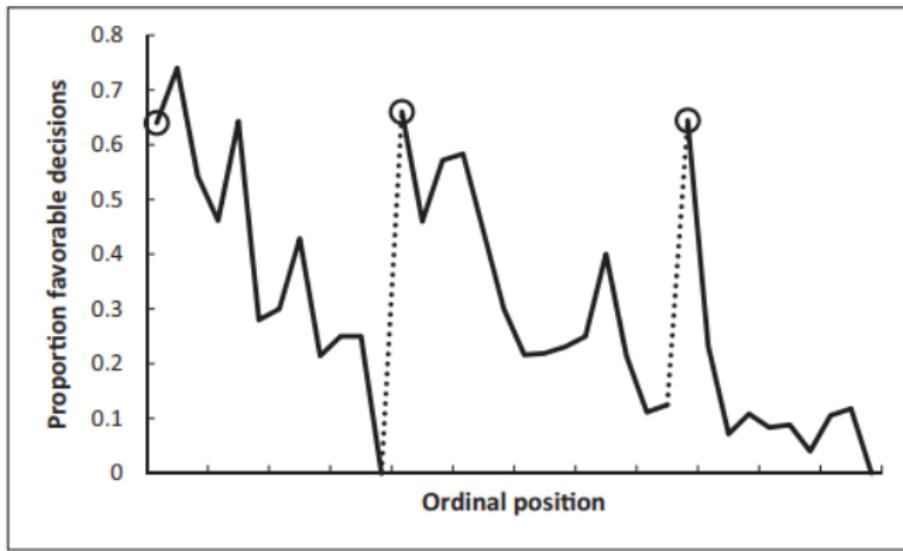
Judge, Jury, or EXEcute File?

U.S. Asylum Courts: Disparities in Grant Rates



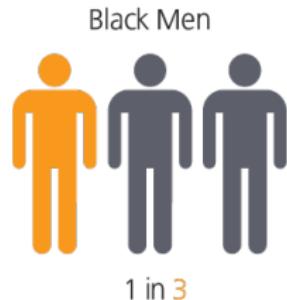
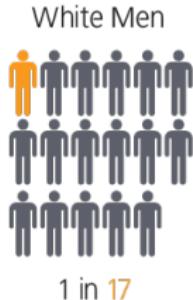
- ▶ In San Francisco, one judge grants 90.6% of asylum requests, while another judge grants just 2.9%!

Jailing Decisions Before/After Lunch Breaks



Source: Danziger et al, PNAS 2011, Israel judges deciding on parole.

Lifetime Likelihood of Imprisonment of U.S. Residents Born in 2001



Source: The Sentencing Project.

Judge, Jury, or EXEcute File?



Welcome

- ▶ This course focuses on applications of **machine learning** and **natural language processing** in law.
- ▶ Goals:
 - ▶ Predict and analyze legal decisions.
 - ▶ Analyze language in legal documents.
 - ▶ Applications of causal inference approaches to decisions and texts.

Lecture Times

- ▶ Mondays, 1:15pm-2:45pm
 - ▶ LFW C5
- ▶ Office hours:
 - ▶ After lecture, or by appointment.

Online Course Materials

- ▶ Course Syllabus:
 - ▶ https://docs.google.com/document/d/1d0H0gqYZHe0gL6Zu7fq-HqGyH5x8_f4k2zgGNPhg6jM/edit?usp=sharing
- ▶ Course Repo:
 - ▶ https://github.com/ellliottt/robot_judge_2019

Readings

- ▶ The material in the slides is based on the readings.
 - ▶ A lot is skipped, so it would be reasonable to focus on the slides for study and to refer to texts as needed.
- ▶ *Natural Language Processing in Python*
(<http://www.nltk.org/book/>)
 - ▶ Chapters 1, 2, 3, 5, 7, 8
- ▶ *Hands-on Machine Learning with Scikit-learn & TensorFlow*
(O'Reilly 2017)
 - ▶ Chapters 2, 3, 4, 7, 8 (code and text)
 - ▶ Chapters 10, 11, 13, 14, 15 (text, not code)
- ▶ See syllabus for other readings.

Python

- ▶ Python is ideal for text data and machine learning.
 - ▶ I recommend Anaconda 3.6: continuum.io/downloads
- ▶ See the course web site for download instructions by platform.
- ▶ I ask that the problem sets be submitted as Jupyter notebooks.
- ▶ If you prefer to use a different programming language, we can discuss.

Problem Sets

- ▶ four problem sets, 10 points each.
- ▶ see due dates on syllabus.
- ▶ implement major methods in prediction and text analysis from class.
- ▶ 40% of grade.

Course Project

- ▶ The main course product is a research paper.
 - ▶ Can be done individually or in groups of two
 - ▶ In consultation with project advisors, form a research design using methods learned in the course.
- ▶ Deliverables:
 - ▶ 1/2 page description of topic (April 8th)
 - ▶ 2+ page proposal (May 20th, 10% of grade)
 - ▶ 10+ page paper (August 19th, 50% of grade)
- ▶ 2 extra credits available for course project (5 credits total):
 - ▶ 20+ pages instead of 10+ pages.
 - ▶ higher expectations for completeness and innovativeness.
 - ▶ should be a significant research contribution and of publication quality.

Office Hours Etc

- ▶ I will be available to meet after lectures.
- ▶ Can also set up appointments by email: ashe@ethz.ch.
- ▶ We can talk about the course material, your research, anything you want.

Class Survey Results

Social Science meets Data Science

- ▶ We are seeing a **revolution in social science**:
 - ▶ **new datasets**: administrative microdata, digitization of text archives, social media
 - ▶ **new methods**: causal inference, natural language processing, machine learning
- ▶ These trends are bringing social science *closer to the natural sciences*.

Law, Institutions, and Big Data

- ▶ Where are these trends most salient?
 - ▶ **law and political economy**
- ▶ The social phenomena of interest – **legal and political institutions** – are composed of thousands, potentially millions, of lines of **unstructured text**.
 - ▶ We cannot read them – somehow we must teach the computers to read them for us.

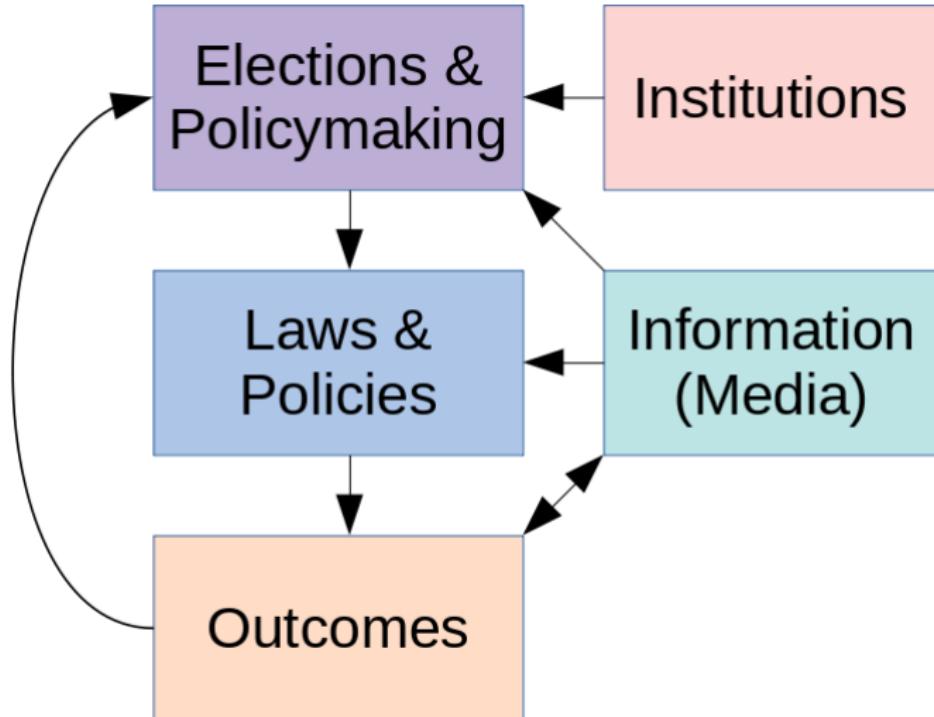
Elections &
Policymaking

Institutions

Laws &
Policies

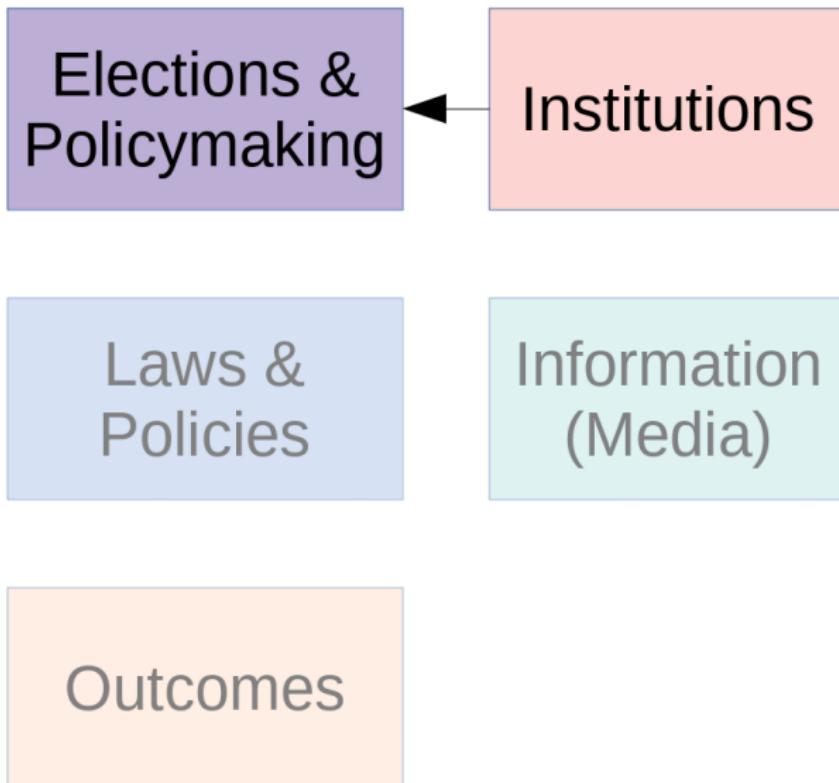
Information
(Media)

Outcomes



Do Electoral Institutions Affect Judges?

Ash and MacLeod (2019)



In some U.S. jurisdictions, judges are elected

Ash and MacLeod (2019)



- ▶ How do elections affect judge decision-making?
 - ▶ Do we choose **democracy** or **bureaucracy**?

Experiment: Elections vs. Merit System

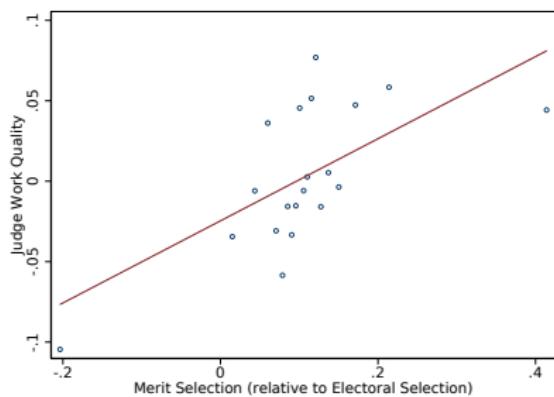
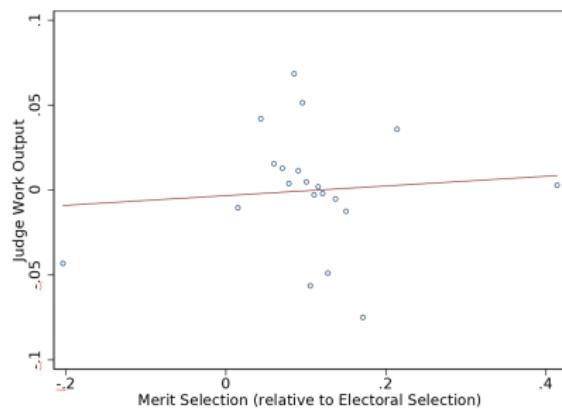
Ash and MacLeod (2019)

Elections	Judges are selected and retained through elections.
Merit commissions	Judges selected by expert panels, and then have tenure.

- ▶ 12 states moved from elections to the merit system between 1960 and 1990.
 - ▶ AZ (1973-), CO (1967-), FL (1977-), IA (1963-), IN (1971-), KS (1968-), MD (1977-), NE (1963-), OK (1968-), SD (1981-), UT (1985-), WY (1973-).
- ▶ How did these reforms affect...
 - ▶ ... work output (constructed from **text of opinions**)?
 - ▶ ... work quality (constructed from **citation network** between judges)?

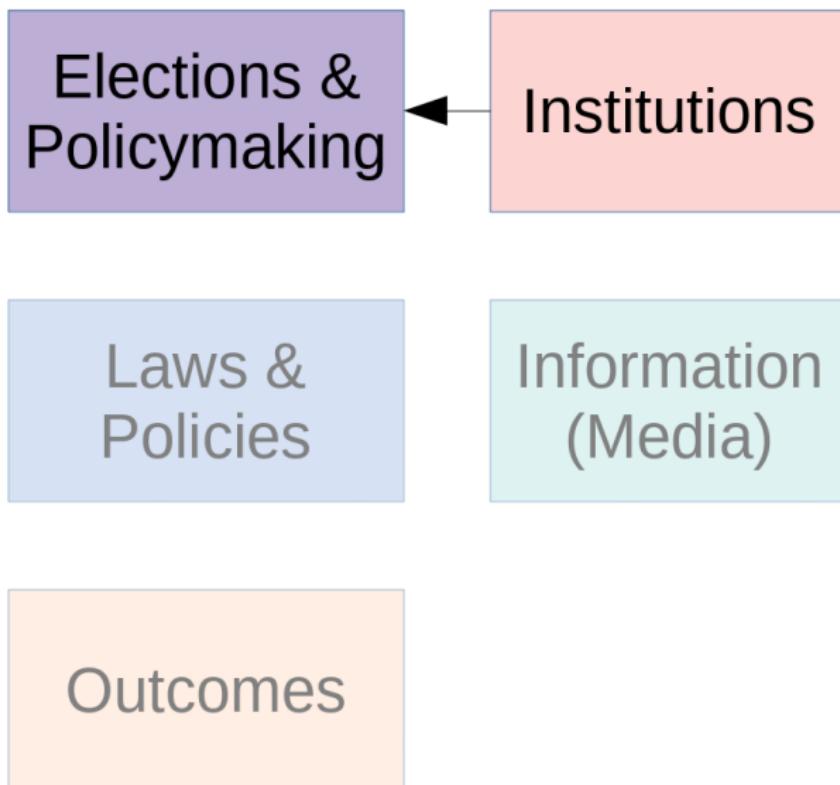
Merit judges do same amount of work, but work is higher-quality.

Ash and MacLeod (2019)



Do Electoral Institutions Affect Parliamentarians?

Ash, Morelli, and Osnabrugge (2018)



Majoritarian vs. Proportional Representation

Ash, Morelli, and Osnabrugge (2018)

- ▶ **Majoritarian (first past the post):** two parties, single party controls parliament.
 - ▶ United States, United Kingdom
- ▶ **Proportional representation:** many minority parties, coalition governments.
 - ▶ Germany, Switzerland

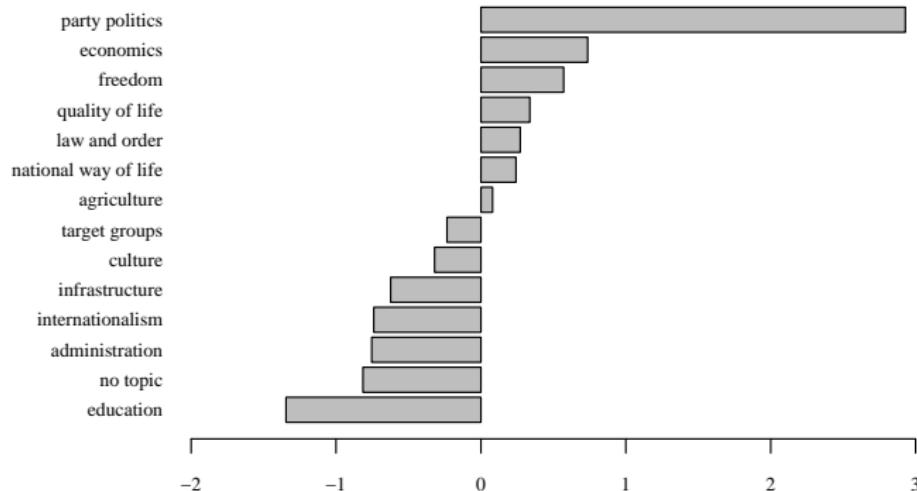
Experiment: Electoral Reform in New Zealand

Ash, Morelli, and Osnabrugge (2018)

- ▶ A 1993 reform in New Zealand moved from majoritarian to proportional representation.
- ▶ How did it affect speech topics in the New Zealand Parliament?
 - ▶ construct a measure of **attention to policies**, using **text of parliamentary speeches**.
 - ▶ economy, education, crime, foreign policy, etc.

Change in Parliament Attention due to Reform

Ash, Morelli, and Osnabrugge (2018)



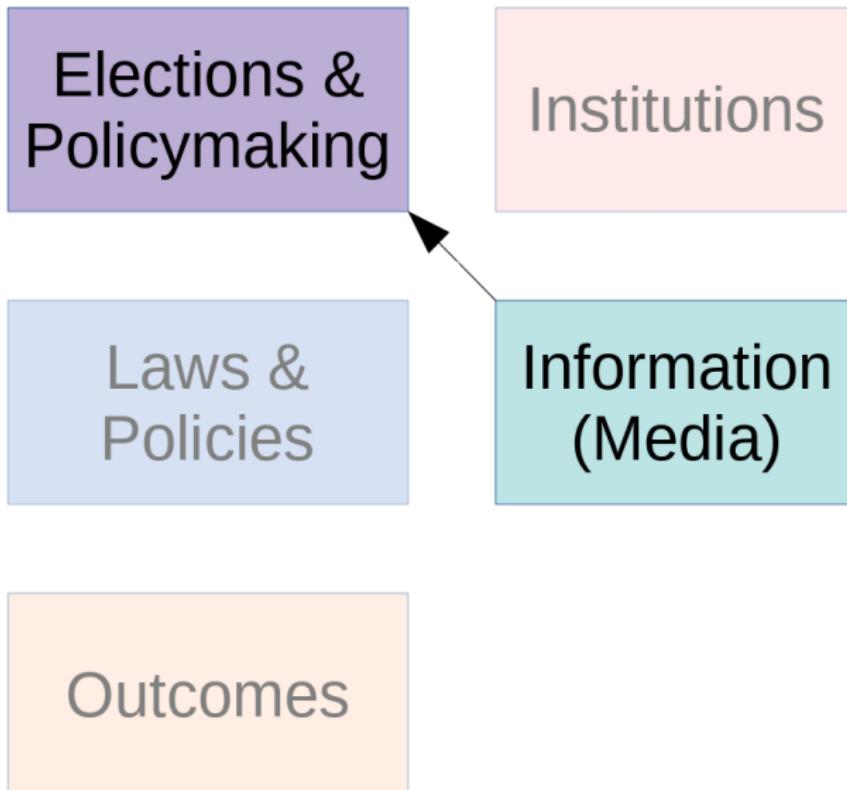
Example “Party Politics” Speech

Ash, Morelli, and Osnabrugge (2018)

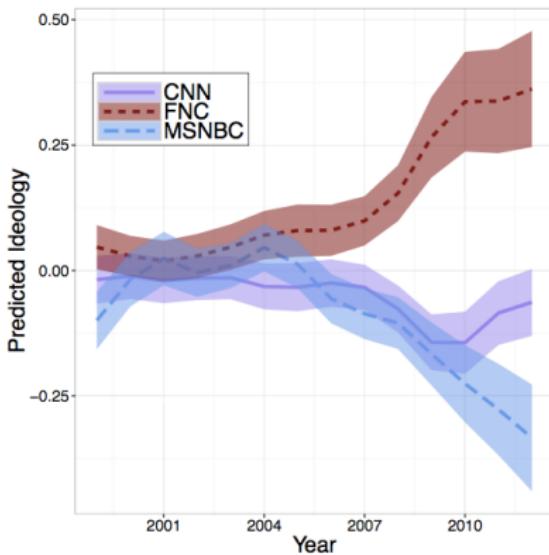
“I have seen seven Opposition leaders in my time, but I have never seen a leader as relentlessly negative as Helen Clark. . . . How could anybody be so negative, day in, day out? It could get into the Guinness Book of Records. She does not have a positive word to say about anything. It is all negative, negative, negative.”

- ▶ Parliamentarian Richard Prebble, 15 Feb 1999

Does News Media Affect Voters?



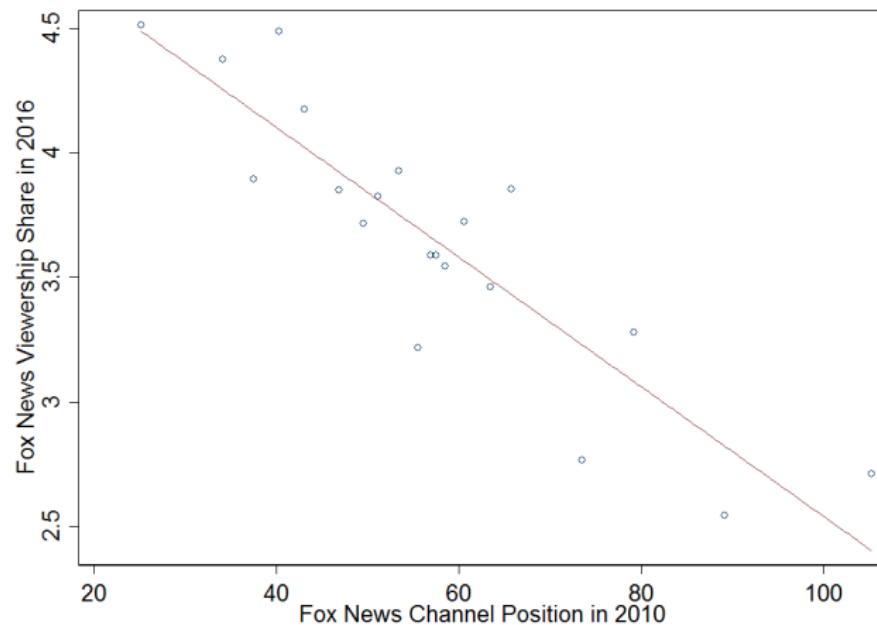
Fox News Channel is Politically Conservative



Martin and Yurukoglu (2017): Estimated ideology based on phrase usage for CNN, Fox News Channel (FNC), and MSNBC. Higher is more conservative.

Low Fox Channel Number → High Fox Viewership

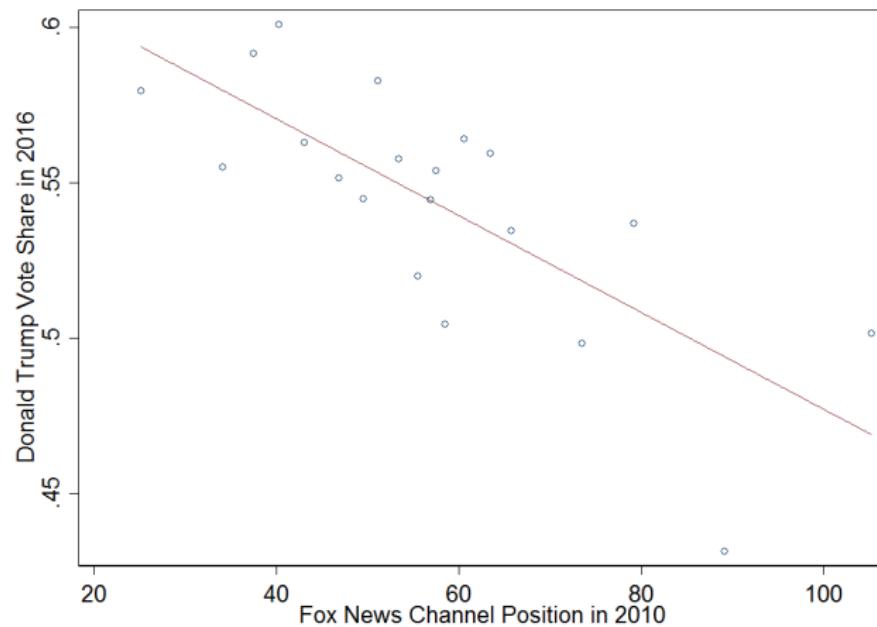
Ash and Galletta (2018)



Average Fox news viewership share plotted against Fox News channel position.

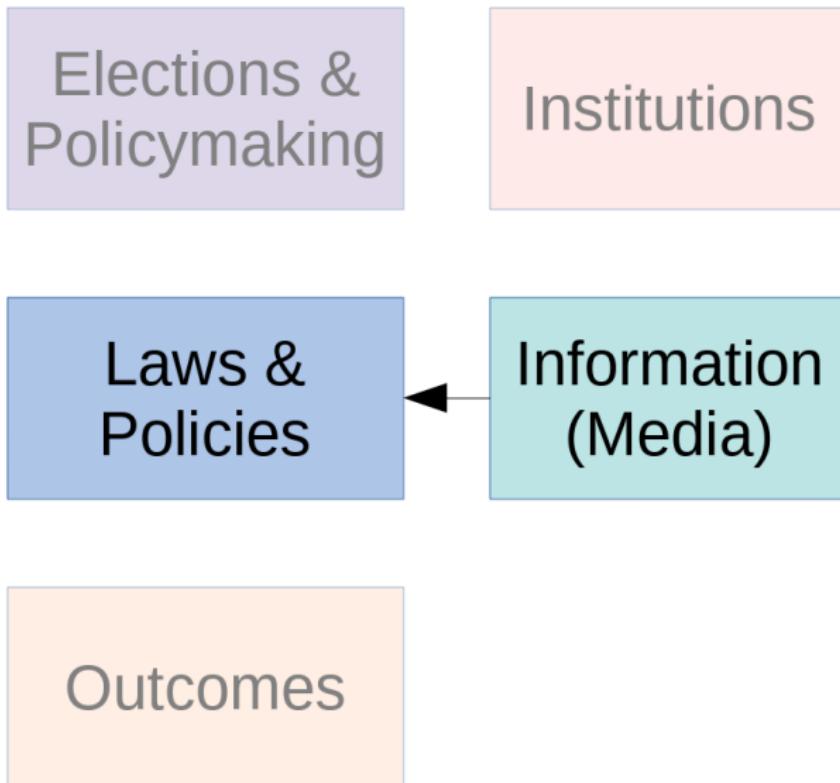
Low Fox Channel Number → More Trump Votes

Ash and Galletta (2018)



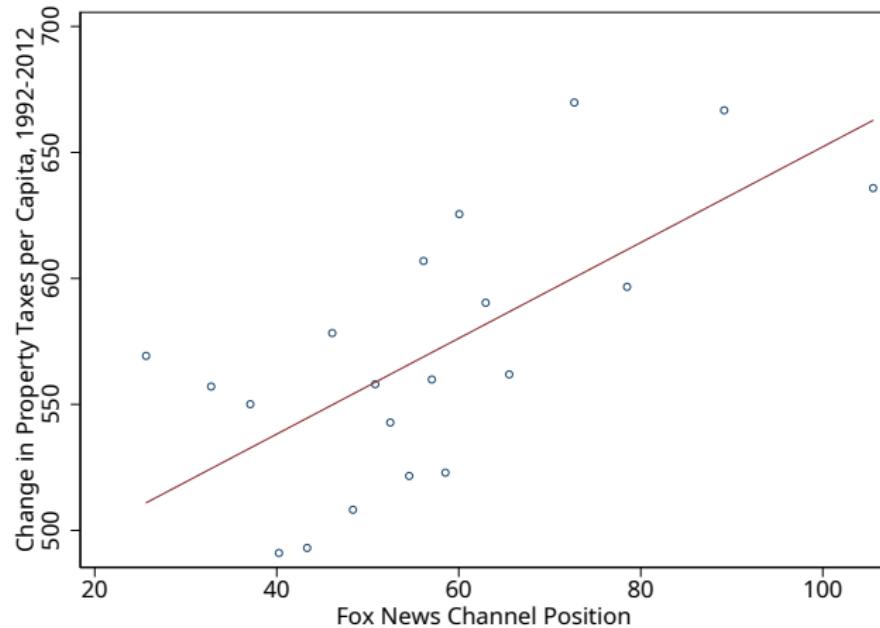
Ash and Galletta (2018). 2SLS estimates: 1.0% increase in Fox News viewer share increases Trump vote share by 6.0% (s.e.=1.1), $p < .01$, Cragg-Donald F-stat = 44.69.

Does News Media affect Policy?



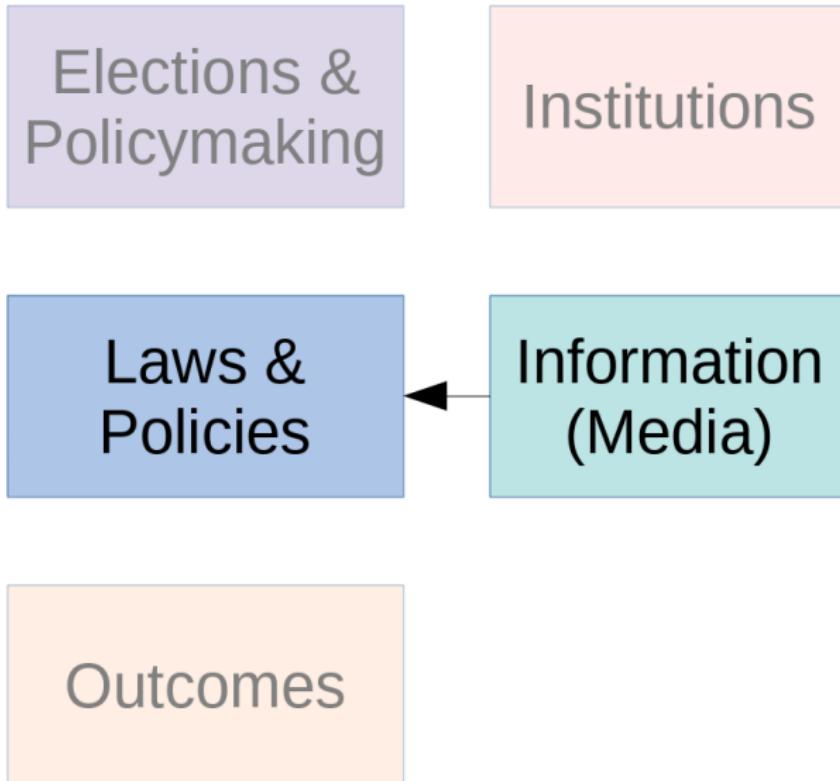
Low Fox Channel Number → Low Property Taxes

Ash and Galletta (2018)



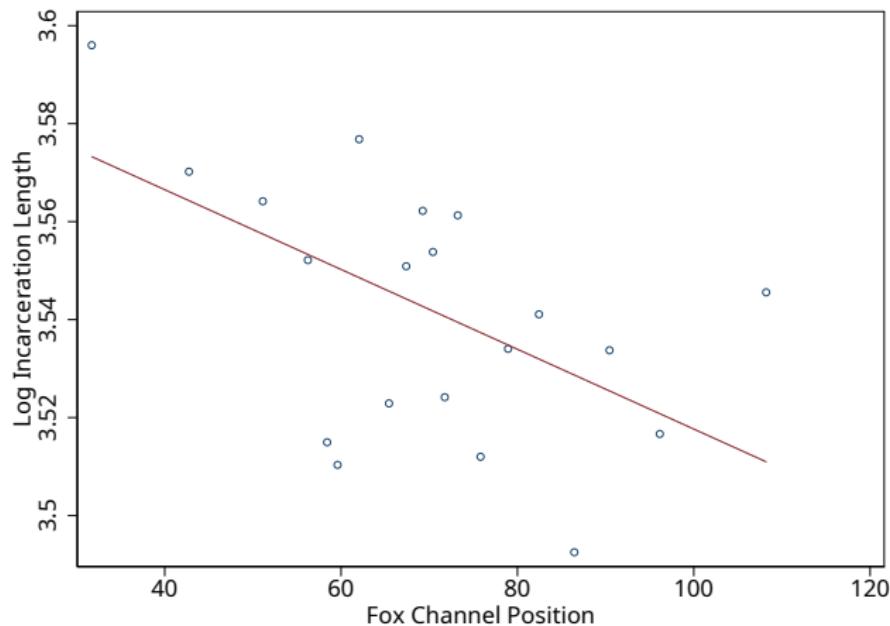
Binscatter diagram from Ash and Galletta (2018); 2SLS estimates: .01% increase in Fox News viewer share decreases Property Taxes per Capita, 1992-2012, by -\$144 (s.e.=\\$52), $p = .006$, Cragg-Donald F-stat = 26.7.

What about Judges?



Low Fox Channel Number → Harsh Sentences

Ash and Poyker (2018)



Binscatter diagram from Ash and Poyker (2018); 2SLS effect of Fox News Viewer Share on Log Sentence Length is 0.075 (0.048), $p=.048$, Cragg-Donald F-stat = 9.7.

...to
attent
amend
gree re

The New York Times

01
1 WO

...going to the bone.

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

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.

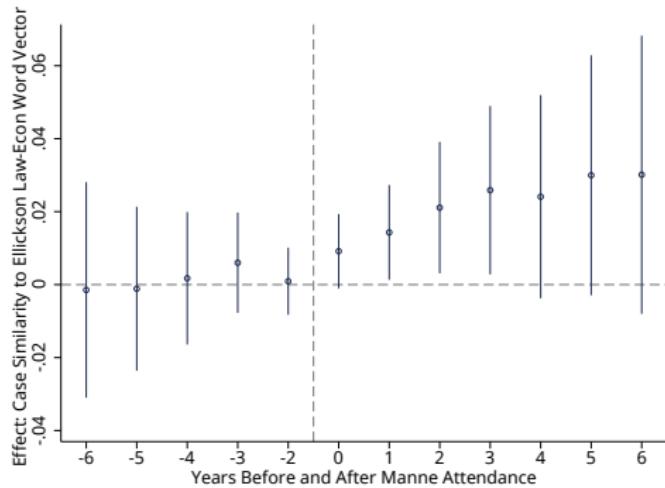
not to relate the theoretical studies cases now pending in Federal court. "One has to be very cautious in dealing with Federal judges," said Henry Mancini, director of the center. "Our goal has been to give them the most recent thinking in economic theory and enable them better understand the testimony of expert witnesses and lawyers."



COURTESY OF FLORIDA COUNTY BAR ASSOCIATION FOR ROBERT JELLINE. STUDENTS AT WORKSHOP ON ECONOMICS FOR THE COURTS FROM LSC. ADDRESS DR. HOWARD MORSE.

Economics Training → Econ Language in Cases

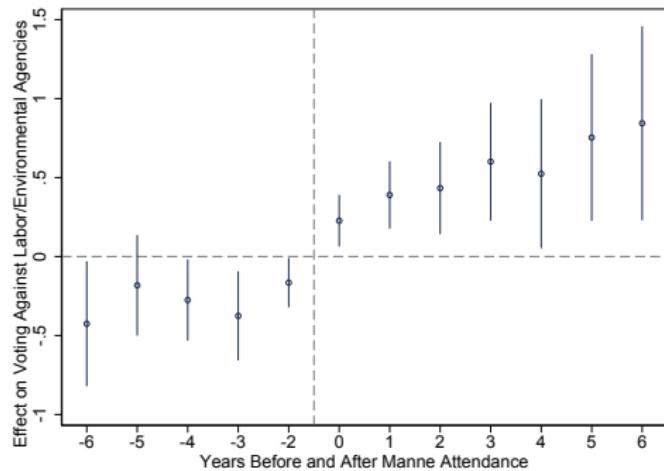
Ash, Chen, and Naidu (2018)



Left panel: List of words related to economics based on word embedding distance to corpus of expert-selected phrases. Right panel: Use of these words in written opinions before and after economics training attendance.

Economics Training → Rulings Against Agencies

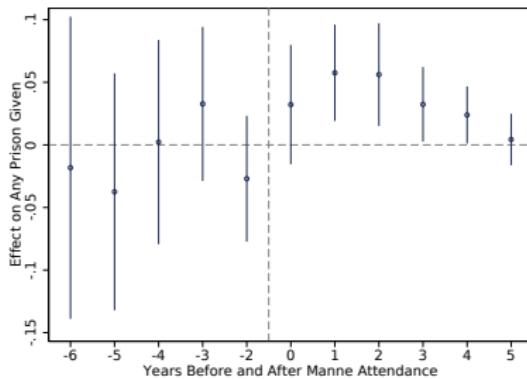
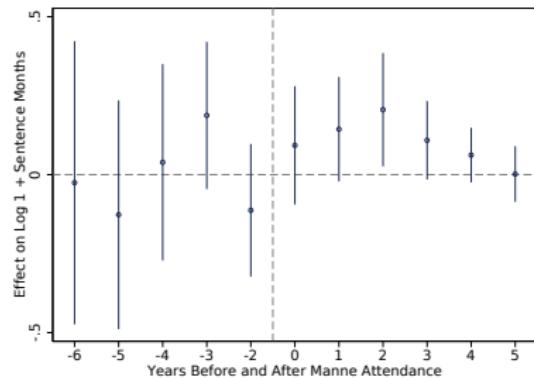
Ash, Chen, and Naidu (2018)



Probability of ruling against environmental regulation agencies (e.g. Environmental Protection Agency) and labor rights agencies (e.g. National Labor Relations Board), before and after economics training attendance.

Economics Training → Harsher Sentencing

Ash, Chen, and Naidu (2018)



Length of criminal sentences imposed (left panel), and probability of giving any prison sentence (right panel), before and after economics training attendance.



Dave Hoffman @HoffProf · Oct 18

If this paper is correct in magnitude or significance--which I think it is not--@ProfWrightGMU should demand a huge raise and the funders of @georgemasonlaw ought to triple down on their investment, because it would be the most successful education program ever.

Matt Grossmann ✅ @MattGrossmann

Judges attending economics training later use more economics language, render more conservative verdicts in economics cases, rule against regulatory agencies more often, & render longer criminal sentences. Effects diffuse across judiciary

Show this thread

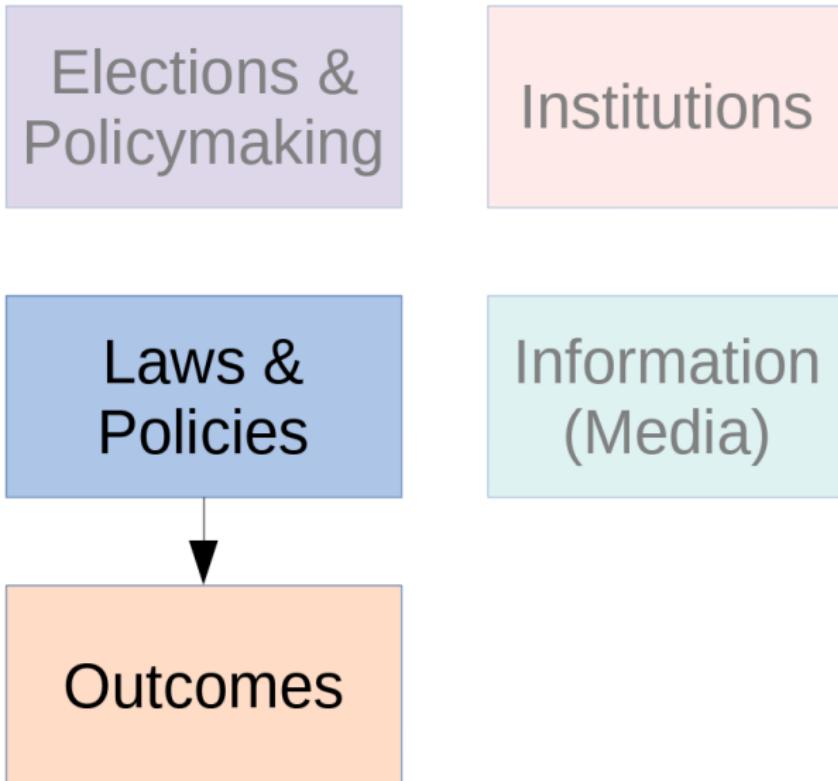
2

4

16

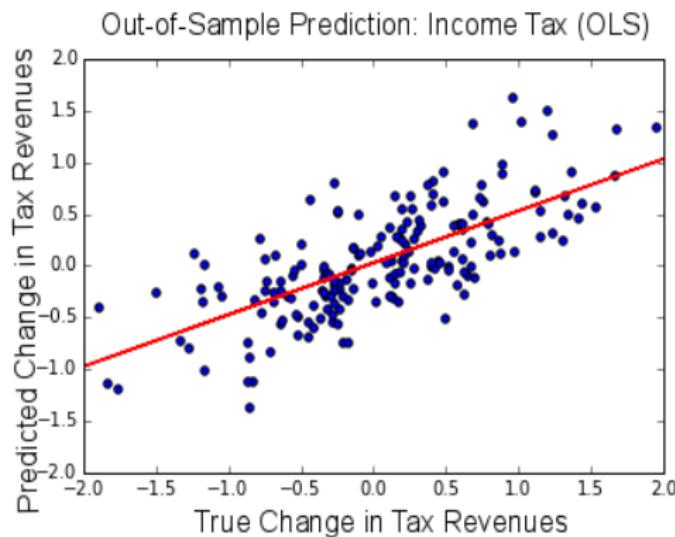


What about outcomes?



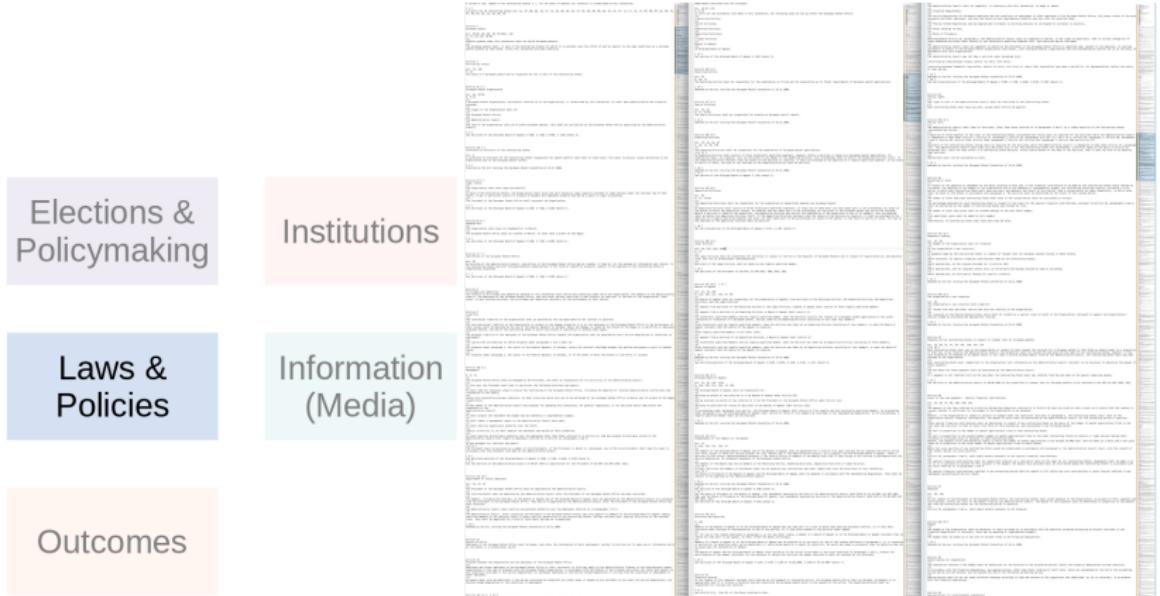
Predicting Tax Revenues Using Tax Code Text

Ash (2018)

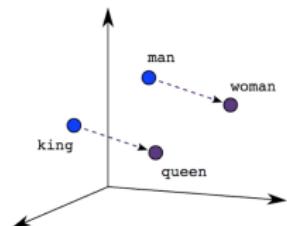


- ▶ Predicted change in state income tax revenue based on enacted statutes (vertical axis), plotted against true change in tax revenue (horizontal axis) (correlation between truth and prediction: 0.89).

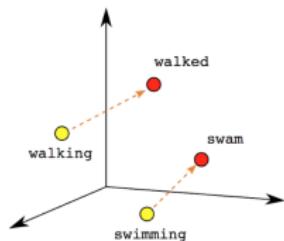
Laws are Embedded in Text



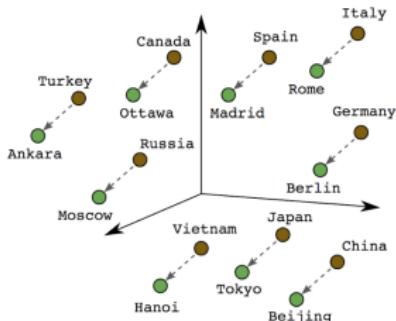
Word Embedding: Language as Data



Male-Female



Verb Tense



Country-Capital

Word Embedding is a technology from computational linguistics that represents words and phrases as vectors in a geometric space, where locations and directions encode meaning.

Analyzing Group Bias with Word Embeddings

Caliskan, Bryson, and Narayanan (*Science* 2017)

- ▶ Groups:
 - ▶ **Flowers:** aster, clover, hyacinth, marigold, poppy, azalea, crocus, iris, orchid, rose, bluebell, daffodil, lilac, pansy, tulip, buttercup, daisy, lily, peony, violet, carnation, gladiola, magnolia, petunia, zinnia.
 - ▶ **Insects:** ant, caterpillar, flea, locust, spider, bedbug, centipede, fly, maggot, tarantula, bee, cockroach, gnat, mosquito, termite, beetle, cricket, hornet, moth, wasp, blackfly, dragonfly, horsefly, roach, weevil.
- ▶ Attributes:
 - ▶ **Pleasant:** caress, freedom, health, love, peace, cheer, friend, heaven, loyal, pleasure, diamond, gentle, honest, lucky, rainbow, diploma, gift, honor, miracle, sunrise, family, happy, laughter, paradise, vacation.
 - ▶ **Unpleasant:** abuse, crash, filth, murder, sickness, accident, death, grief, poison, stink, assault, disaster, hatred, pollute, tragedy, divorce, jail, poverty, ugly, cancer, kill, rotten, vomit, agony, prison.
- ▶ **Flower** words are close to **pleasant** words; **insect** words are close to **unpleasant** words.

Word Associations in Judge Language

Ash, Chen, and Ornaghi (2018)

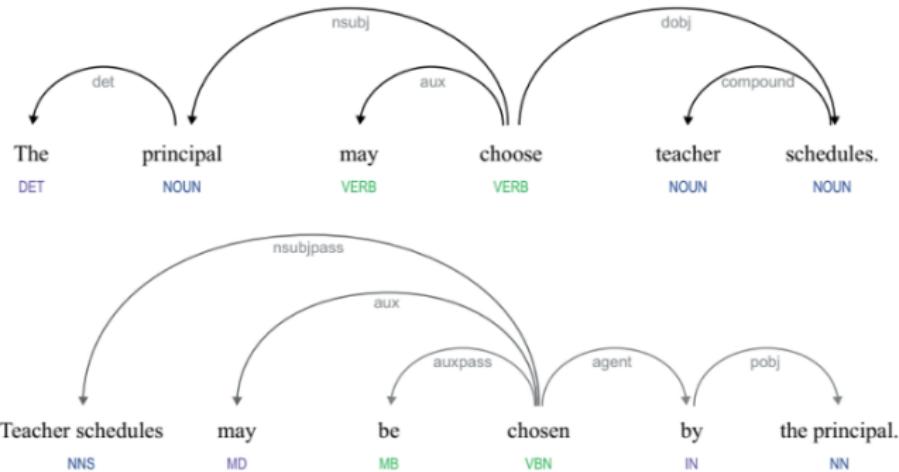
- ▶ Male (left) and female (right) adjectives:



- ▶ White-race (left) and black-race (right) adjectives:



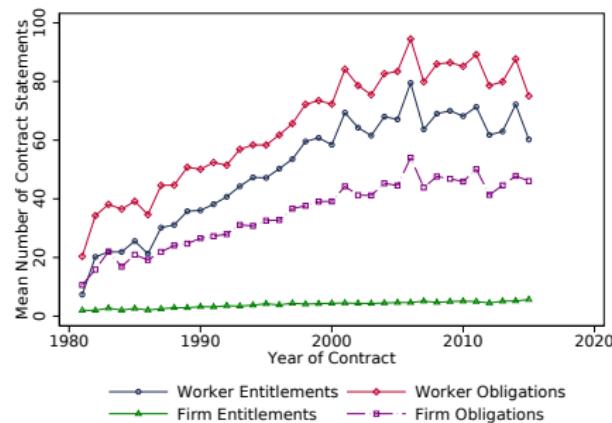
Extracting Information from Legal Texts



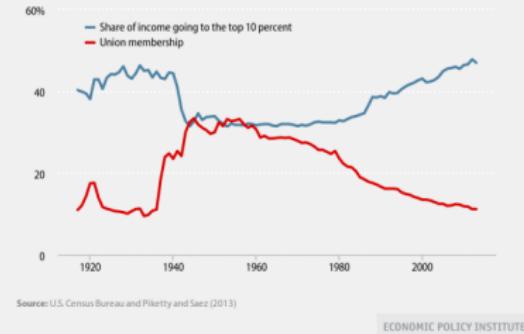
- ▶ **Syntactic dependency parsers** allow computers to read texts and parse subjects, actions, and other useful information.
 - ▶ In particular, modal verbs **shall**, **must**, **will**, **may**, and **can** encode **obligations** and **entitlements** in legal language.

Measuring Worker Authority in Labor Contracts

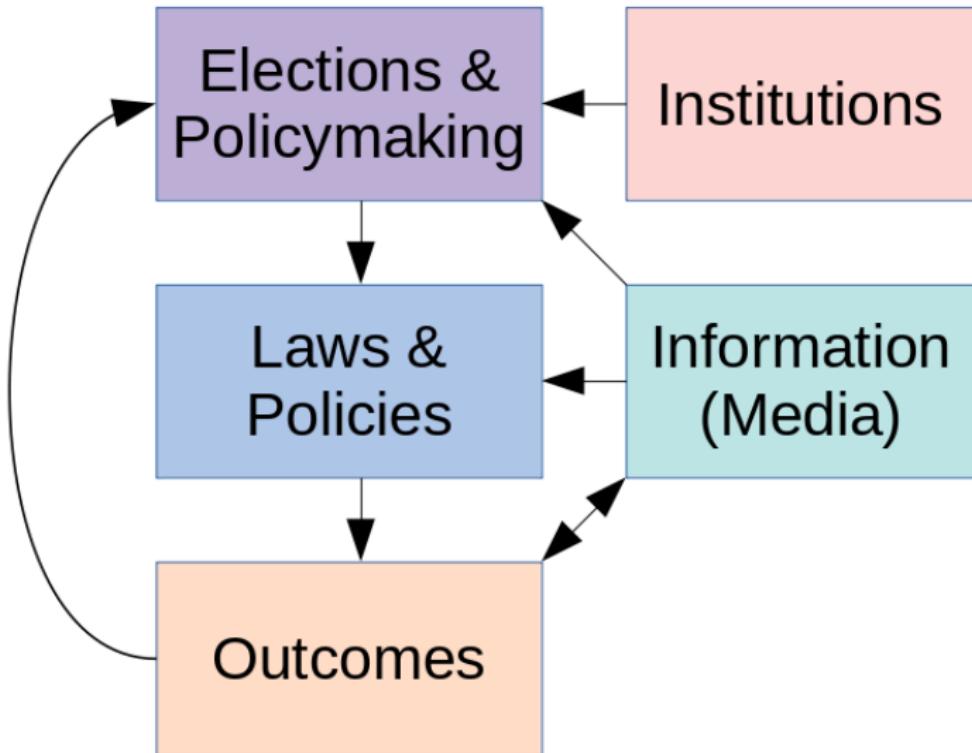
Ash, MacLeod, and Naidu (2018)



Union membership and share of income going to the top 10%



- ▶ Canadian Labor contracts, 1981-2015, consist of:
 - ▶ obligations for both firms and workers
 - ▶ entitlements for workers, but not firms
- ▶ Worker entitlements decrease strike responses.





Could your next lawyer be a **robot**? Tech firms making case for ...

Miami Herald - 4 hours ago

Litigators, **Judges**, and big firm partners aren't likely to feel heat from **robots** any time soon, but lower level staffers and recent law school graduates may find themselves competing with algorithms. A 2016 report from the consultant group McKinsey found that 22% of a lawyer's tasks and 35% of a law clerk's ...



How Artificial Intelligence Is Transforming the Legal Services ...

Analytics India Magazine - 4 Apr 2018

Wusong Technology is one of the leaders in **China**, working on digitizing the way Courts function using AI-enabled **robot**-chatbot called "Fa Xiaotao", which ... analytics to visualize case data in a networked map of which cases are cited in other cases and whether they lead to a positive or adverse **verdict**.



Robot **judges** could soon be helping with court cases

The Independent - 24 Oct 2016

An artificial intelligence (AI) **Judge** has accurately predicted most **verdicts** of the European Court of Human Rights, and might soon be making important decisions about cases. Scientists built an artificial intelligence computer that was able to look at legal evidence as well as considering ethical questions to ...



Conviction by computer is go, confirms **UK Ministry of Justice**

The Register - 8 Feb 2017

Petty criminals in **Britain** will soon be found guilty and sentenced by computers, under new government plans. Originally floated last year in a ...

What does a (robot) judge do?

- ▶ Takes **evidence**, and makes **decisions**.

Algorithms for legal decisions

1. Collect data on **all** cases (evidence and decision).
2. Algorithm learns to **replicate** what judge would decide.
 - ▶ Evaluate performance in held-out data.

- ▶ In study of prosecutor charge decisions, we obtained **88% accuracy**.

Source: Amaranto, Ash, Chen, Ren, and (2018). New Orleans District Attorney's Office.

- ▶ On appeal, **human judges affirmed about 85% of the time**.

Source: U.S. District Court civil cases, Administrative Office U.S. Courts, 2017.

- ▶ **Human doctors** are correct in diagnosis about **80% of the time**.

Source: BMJ Quality and Safety, 2015, based on autopsy reviews.

Predicting U.S. Asylum Court Decisions

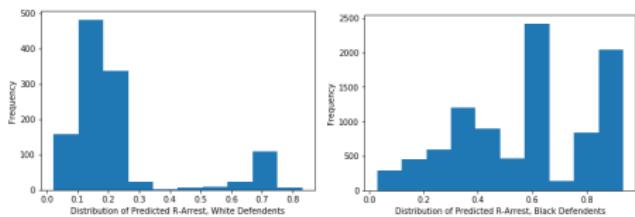
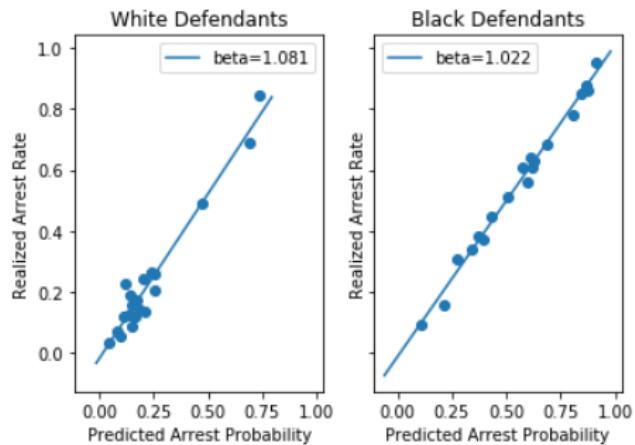
		Predicted	
		Denied	Granted
True	Denied	195,223	65,798
	Granted	73,269	104,406

Accuracy = 68.3%, F1 = 0.60

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

Predicting Re-Arrest

Amaranto, Ash, Chen, Ren, and Roper (2018)



Is 88% accurate enough?

- ▶ Maybe, maybe not...
 - ▶ But at some point (95%? 99%?), we might be happy to apply these predictions as law in future cases.
 - ▶ A 2014 audit found that Baltimore speeding cameras were only 90% accurate.



- ▶ Advantages of consistency:
 - ▶ Decisions will no longer depend on which judge you are assigned, whether he/she has had lunch, or whether it's hot out (Heyes and Saberian 2017).



“Siri, guilty or innocent?”

Source: skrubu.net.

- ▶ The **robot clerk** could provide predictions/advice based on what judges have done previously.

- ▶ **Algorithms can correct the biases of individual judges:**
 - ▶ all defendants get the same decision for the same evidence.
- ▶ **What about *systematic* biases?**
 - ▶ e.g., those leading to racial disparities.
 - ▶ There is a risk these could be reproduced or even amplified by the algorithm.
 - ▶ But algorithms can also be used to **detect** systematic bias, to **understand** it – and therefore to help **reduce** it.

Correlation vs. Causation

- ▶ A new literature at the intersection of machine learning and econometrics has begun to address the correlation-vs-causation problem:
 - ▶ Orthogonalized machine learning (Chernozhukov et al 2017):
 - ▶ “Orthogonalize” (residualize) treatment and outcome variables on set of instruments before prediction.
 - ▶ “Deep” instrumental variables (Hartford et al. 2017):
 - ▶ Use deep neural networks to estimate/predict first and second stage.
- ▶ In particular, can use **random assignment of judges** to cases to approximate **randomized experiments**.

Interpretable Machine Learning

- ▶ Key point:
 - ▶ Standard machine learning techniques cannot be interpreted easily.
- ▶ Clients and judges want to understand the model
 - ▶ Lawyers have told me that this is the major barrier to law firms automating more tasks – clients don't like a black box.
- ▶ Other models/approaches improve interpretability:
 - ▶ Random Forests provide feature importance ranking.
 - ▶ Structural Topic Model gives predictions and topics.
 - ▶ LIME and related tools can help interpret any model (Ribeiro et al 2016).

Robot Justice?

- ▶ **Human judges:**
 - ▶ have a concept of justice:
 - ▶ can rule on new types of cases, and in response to changing laws/norms.
- ▶ **Robot judges:**
 - ▶ help human judges reach that concept of justice.