Judge Embeddings: Toward Vector Representations of Legal Belief

Elliott Ash and Daniel L. Chen

December 2017

Vector Representations for Language Analysis

- Recent advances in natural language processing have stemmed from using dense vectors to represent language relations:
  - Topic models for encoding relations between documents (e.g. LDA, Blei 2003)
  - Word embeddings for encoding relations between words and phrases (e.g. word2vec and glove, Mikolov et al 2013).
- This is an active research area with a cascade of extensions and variations:
  - Today I'll discuss applications of these techniques to legal language.

Vector Representations for Language Analysis

- Recent advances in natural language processing have stemmed from using dense vectors to represent language relations:
  - Topic models for encoding relations between documents (e.g. LDA, Blei 2003)
  - Word embeddings for encoding relations between words and phrases (e.g. word2vec and glove, Mikolov et al 2013).
- This is an active research area with a cascade of extensions and variations:
  - Today I'll discuss applications of these techniques to legal language.

Outline

Why word vectors?

- A basic function of word vectors is as an efficient dimension reduction method, where a wide sparse matrix is reduced to a thin dense matrix, and then used in down-stream prediction tasks.
- In addition, once words are represented as vectors, we can use linear algebra to understand the relationships between words:
  - Words that are geometrically close to each other are similar: e.g. "student" and "pupil."
- More intriguingly, embeddings algebra can depict conceptual, analogical relationships between words.
  - Consider the analogy: **man is to king as woman is to** ____
  - With embeddings, we have

$$vec(king) - vec(man) + vec(woman) \approx vec(queen)$$

- Trained on a corpus of statutes (Ash 2016), we have

$$vec[\text{"corporate income tax"}] - vec[\text{"corporation"}] + vec[\text{"person"}]$$

$$\approx vec[\text{"personal income tax"}].$$

Why word vectors?

- A basic function of word vectors is as an efficient dimension reduction method, where a wide sparse matrix is reduced to a thin dense matrix, and then used in down-stream prediction tasks.
- In addition, once words are represented as vectors, we can use linear algebra to understand the relationships between words:
  - Words that are geometrically close to each other are similar: e.g. "student" and "pupil."
- More intriguingly, embeddings algebra can depict conceptual, analogical relationships between words.
  - Consider the analogy: **man is to king as woman is to** ____
  - With embeddings, we have

$$vec(king) - vec(man) + vec(woman) \approx vec(queen)$$

- Trained on a corpus of statutes (Ash 2016), we have

$$vec[\text{"corporate income tax"}] - vec[\text{"corporation"}] + vec[\text{"person"}]$$

$$\approx vec[\text{"personal income tax"}].$$

Why word vectors?

- A basic function of word vectors is as an efficient dimension reduction method, where a wide sparse matrix is reduced to a thin dense matrix, and then used in down-stream prediction tasks.
- In addition, once words are represented as vectors, we can use linear algebra to understand the relationships between words:
  - Words that are geometrically close to each other are similar: e.g. "student" and "pupil."
- More intriguingly, embeddings algebra can depict conceptual, analogical relationships between words.
  - Consider the analogy: **man is to king as woman is to ____**
  - With embeddings, we have

$$vec(king) - vec(man) + vec(woman) \approx vec(queen)$$

- Trained on a corpus of statutes (Ash 2016), we have

$$vec["\text{corporate income tax}"] - vec["\text{corporation}"] + vec["\text{person}"]$$

$$\approx vec["\text{personal income tax}"].$$

Why word vectors?

- A basic function of word vectors is as an efficient dimension reduction method, where a wide sparse matrix is reduced to a thin dense matrix, and then used in down-stream prediction tasks.
- In addition, once words are represented as vectors, we can use linear algebra to understand the relationships between words:
  - Words that are geometrically close to each other are similar: e.g. "student" and "pupil."
- More intriguingly, embeddings algebra can depict conceptual, analogical relationships between words.
  - Consider the analogy: **man is to king as woman is to ____**
  - With embeddings, we have

$$vec(king) - vec(man) + vec(woman) \approx vec(queen)$$

- Trained on a corpus of statutes (Ash 2016), we have

$$vec[\text{"corporate income tax"}] - vec[\text{"corporation"}] + vec[\text{"person"}]$$

$$\approx vec[\text{"personal income tax"}].$$

Why word vectors?

- A basic function of word vectors is as an efficient dimension reduction method, where a wide sparse matrix is reduced to a thin dense matrix, and then used in down-stream prediction tasks.
- In addition, once words are represented as vectors, we can use linear algebra to understand the relationships between words:
  - Words that are geometrically close to each other are similar: e.g. "student" and "pupil."
- More intriguingly, embeddings algebra can depict conceptual, analogical relationships between words.
  - Consider the analogy: **man is to king as woman is to** _ _ _ _
  - With embeddings, we have

$$vec(king) - vec(man) + vec(woman) \approx vec(queen)$$

- Trained on a corpus of statutes (Ash 2016), we have

$$vec[\text{"corporate income tax"}] - vec[\text{"corporation"}] + vec[\text{"person"}]$$

$$\approx vec[\text{"personal income tax"}].$$

Word Function $\longleftrightarrow$ Word Neighbors

- "You shall know a word by the company it keeps"

  - J.R. Firth, Papers in Linguistics,1957

- “He filled the **wampimuk**, passed it around and we all drunk some.”
- “The defendant was convicted of **wampimuk** and sentenced to life in prison.”

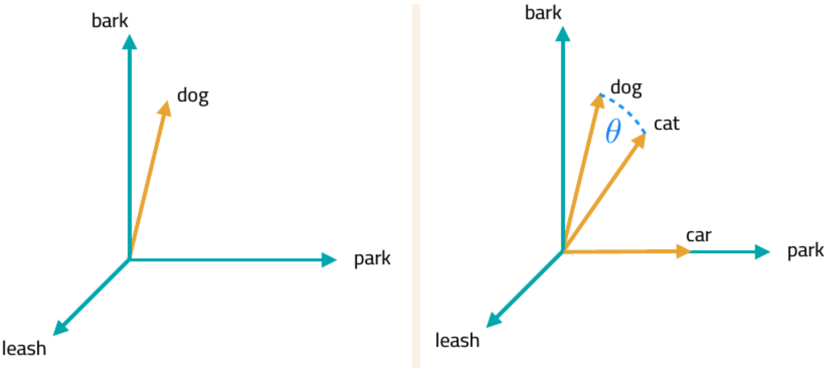
Word Function $\longleftrightarrow$ Word Neighbors

- "You shall know a word by the company it keeps"

  - J.R. Firth, Papers in Linguistics,1957

- "He filled the **wampimuk**, passed it around and we all drunk some."
- "The defendant was convicted of **wampimuk** and sentenced to life in prison."

Words as Vectors



- Use cosine similarity as a measure of relatedness:

$$\cos\theta = \frac{v_1 \cdot v_2}{||v_1||||v_2||}$$

Most similar words to dog, depending on window size

More paradigmatic ← → More syntagmatic

2-word window	30-word window
cat	kennel
horse	puppy
fox	pet
pet	bitch
rabbit	terrier
pig	rottweiler
animal	canine
mongrel	cat
sheep	bark
pigeon	alsatian

- Small windows pick up substitutable words; large windows pick up topics.

Generalized Embeddings

- Embeddings models have been extended from words to phrases, sentences, and documents (e.g. Le and Mikolov 2014).
  - Document embeddings are different from topic models because the vector dimensions have a geometric (rather than topic-share) interpretation
- More generalized uses of embeddings include shopping cart embeddings, which can identify complements and substitutes (Blei 2016).
- We want to treat a judicial opinion, or a judge, as a rich object with language and metadata features – embeddings can acommodate this.

Generalized Embeddings

- Embeddings models have been extended from words to phrases, sentences, and documents (e.g. Le and Mikolov 2014).
  - Document embeddings are different from topic models because the vector dimensions have a geometric (rather than topic-share) interpretation
- More generalized uses of embeddings include shopping cart embeddings, which can identify complements and substitutes (Blei 2016).
- We want to treat a judicial opinion, or a judge, as a rich object with language and metadata features – embeddings can acommodate this.

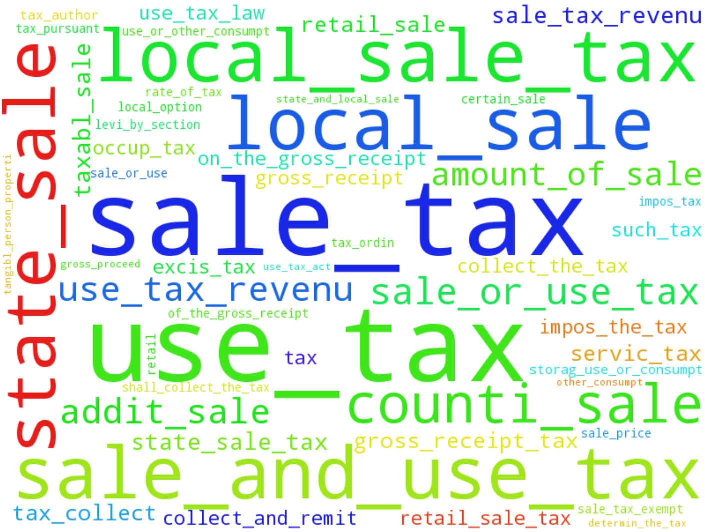
Generalized Embeddings

- Embeddings models have been extended from words to phrases, sentences, and documents (e.g. Le and Mikolov 2014).
  - Document embeddings are different from topic models because the vector dimensions have a geometric (rather than topic-share) interpretation
- More generalized uses of embeddings include shopping cart embeddings, which can identify complements and substitutes (Blei 2016).
- We want to treat a judicial opinion, or a judge, as a rich object with language and metadata features – embeddings can acommodate this.

Outline

Outline

Which laws are close to "sales tax"?



Outline

1. Introduction to Word Embeddings
2. Related Work
   - Ash (2016): Classifying Tax Statutes
   - Rudolph and Blei (2017): Dynamic Word Embeddings
   - Implicit Bias in Language
3. Legal Applications
   - Judge Embeddings Demonstration
   - Law & Economics (Ash-Chen-Naidu 2017)
   - What about originalism?
   - Implicit Bias in Judiciary
4. Concluding Thoughts
5. Appendix Slides

Rudolph and Blei (2017)

- Train word embeddings on the U.S. Congressional Record, 1858-2009.
- Dynamic word embeddings model:
  - Captures how the meaning of words evolves over time.
  - The innovation is to include "year" in the embedding model, and allow word vectors to drift over time.

Rudolph and Blei (2017)

- Train word embeddings on the U.S. Congressional Record, 1858-2009.
- Dynamic word embeddings model:
  - Captures how the meaning of words evolves over time.
  - The innovation is to include "year" in the embedding model, and allow word vectors to drift over time.

Word Meaning Changes

computer		bush	
1858	**1986**	**1858**	**1990**
computer	computer	bush	bush
draftsman	software	barberry	cheney
draftsmen	computers	rust	nonsense
copyist	copyright	bushes	nixon
photographer	technological	borer	reagan
computers	innovation	eradication	george
copyists	mechanical	grasshoppers	headed
janitor	hardware	cancer	criticized
accountant	technologies	tick	clinton
bookkeeper	vehicles	eradicate	blindness

Drift in word "prostitution"

prostitution				
1930	**1945**	**1962**	**1988**	**1990**
prostitution	prostitution	prostitution	harassment	prostitution
punishing	indecent	indecent	intimidation	servitude
immoral	vile	harassment	prostitution	harassment
bootlegging	immoral	intimidation	counterfeit	intimidation
riotous	induces	sexual	illegal	trafficking
forbidden	incite	vile	trafficking	harassing
anarchists	abortion	counterfeit	indecent	apprehended
assemblage	forbid	anarchists	disregard	killings
forbid	harboring	mobs	anarchists	labeled
abet	assemblage	lawbreakers	punishing	naked

Outline

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

- "We replicated a spectrum of known biases, as measured by the Implicit Association Test, using a widely used, purely statistical machine-learning model trained on a standard corpus of text from the World Wide Web. . . "

Word Embedding Association Test

- Target words:
  - programmer, engineer, scientist, ...
  - nurse, teacher, librarian, ...

  - caress, freedom, health, love, peace, cheer, friend, . . .
  - abuse, crash, filth, murder, sickness, accident, . . .
- Attribute words:
  - man, male, ...
  - woman, female, ...

  - white, caucasian, european, . . .
  - black, african, negro, . . .
- WEAT Test:
  - Compute similarities between all target words and all attribute words
  - Compute mean target-attribute clustering

Word Embedding Association Test

- Target words:
  - programmer, engineer, scientist, ...
  - nurse, teacher, librarian, ...

  - caress, freedom, health, love, peace, cheer, friend, . . .
  - abuse, crash, filth, murder, sickness, accident, . . .
- Attribute words:
  - man, male, ...
  - woman, female, ...

  - white, caucasian, european, . . .
  - black, african, negro, . . .
- WEAT Test:
  - Compute similarities between all target words and all attribute words
  - Compute mean target-attribute clustering

Word Embedding Association Test

- Target words:
  - programmer, engineer, scientist, ...
  - nurse, teacher, librarian, ...

  - caress, freedom, health, love, peace, cheer, friend, . . .
  - abuse, crash, filth, murder, sickness, accident, . . .
- Attribute words:
  - man, male, ...
  - woman, female, ...

  - white, caucasian, european, . . .
  - black, african, negro, . . .
- WEAT Test:
  - Compute similarities between all target words and all attribute words
  - Compute mean target-attribute clustering

Word Embedding Association Test

- Target words:
  - programmer, engineer, scientist, ...
  - nurse, teacher, librarian, ...

  - caress, freedom, health, love, peace, cheer, friend, . . .
  - abuse, crash, filth, murder, sickness, accident, . . .
- Attribute words:
  - man, male, ...
  - woman, female, ...

  - white, caucasian, european, . . .
  - black, african, negro, . . .
- WEAT Test:
  - Compute similarities between all target words and all attribute words
  - Compute mean target-attribute clustering

Word Embedding Association Test

- Target words:
  - programmer, engineer, scientist, ...
  - nurse, teacher, librarian, ...

  - caress, freedom, health, love, peace, cheer, friend, . . .
  - abuse, crash, filth, murder, sickness, accident, . . .
- Attribute words:
  - man, male, ...
  - woman, female, ...

  - white, caucasian, european, . . .
  - black, african, negro, . . .
- WEAT Test:
  - Compute similarities between all target words and all attribute words
  - Compute mean target-attribute clustering

Bolukbasi et al (NIPS 2016)

- Geometrically, **gender association is shown to be captured by a direction in the word embedding**.
- Gender-neutral words are linearly separable from gender-definition words in the word embedding space.
- "Using these properties, we provide a methodology for modifying an embedding to remove gender stereotypes, such as the association between the words receptionist and female, while maintaining desired associations such as between the words queen and female."

Bolukbasi et al (NIPS 2016)

- Geometrically, **gender association is shown to be captured by a direction in the word embedding**.
- Gender-neutral words are linearly separable from gender-definition words in the word embedding space.
- "Using these properties, we provide a methodology for modifying an embedding to remove gender stereotypes, such as the association between the words receptionist and female, while maintaining desired associations such as between the words queen and female."

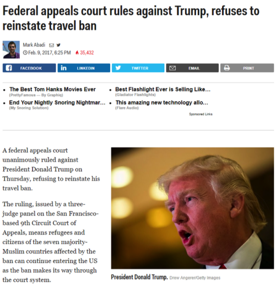
Bolukbasi et al (NIPS 2016)

- Geometrically, **gender association is shown to be captured by a direction in the word embedding**.
- Gender-neutral words are linearly separable from gender-definition words in the word embedding space.
- "Using these properties, we provide a methodology for modifying an embedding to remove gender stereotypes, such as the association between the words receptionist and female, while maintaining desired associations such as between the words queen and female."

Outline

U.S. Courts as "Natural Laboratory"

- Do schools of thought matter for policymaking?
  - We have recently seen the importance of US federal courts ruling against Trump.
  - These courts involve expert decision-making with far-reaching implications.



Federal appeals court rules against Trump, refuses to reinstate travel ban

Mark Abadi
Feb. 9, 2017, 6:25 PM 35,432

FACEBOOK LINKEDIN TWITTER EMAIL PRINT

- The Best Tom Hanks Movies Ever (PrettyFamous — By Graphiq)
- Best Flashlight Ever is Selling Like... (Gladiator Flashlights)
- End Your Nightly Snoring Nightmar... (My Snoring Solution)
- This amazing new technology allo... (Flare Audio)

Sponsored Links

A federal appeals court unanimously ruled against President Donald Trump on Thursday, refusing to reinstate his travel ban.

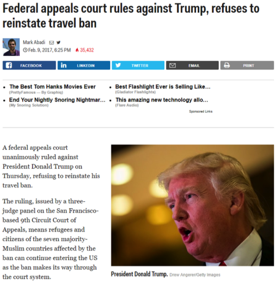
The ruling, issued by a three-judge panel on the San Francisco-based 9th Circuit Court of Appeals, means refugees and citizens of the seven majority-Muslim countries affected by the ban can continue entering the US as the ban makes its way through the court system.

President Donald Trump. Drew Angerer/Getty Images

- Judges exercise power and discretion in policymaking. (e.g. Epstein et al. 2013)
  - **Interpret**, apply, create law and legal precedent under uncertainty.
    - **Subjective** decision-making creates a role for schools of thinking.
  - e.g. Originalism, Critical Legal Theory, **Law and Economics**.
- Can embeddings models help us measure schools of thought?

U.S. Courts as "Natural Laboratory"

- Do schools of thought matter for policymaking?
  - We have recently seen the importance of US federal courts ruling against Trump.
  - These courts involve expert decision-making with far-reaching implications.



Federal appeals court rules against Trump, refuses to reinstate travel ban

Mark Abadi

Feb. 9, 2017, 6:25 PM 35,432

FACEBOOK LINKEDIN TWITTER EMAIL PRINT

- The Best Tom Hanks Movies Ever (PrettyFamous — By Graphiq)
- Best Flashlight Ever is Selling Like... (Gladiator Flashlights)
- End Your Nightly Snoring Nightmar... (My Snoring Solution)
- This amazing new technology allo... (Flare Audio)

Sponsored Links

A federal appeals court unanimously ruled against President Donald Trump on Thursday, refusing to reinstate his travel ban.

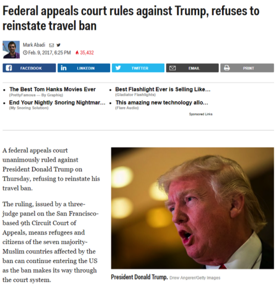
The ruling, issued by a three-judge panel on the San Francisco-based 9th Circuit Court of Appeals, means refugees and citizens of the seven majority-Muslim countries affected by the ban can continue entering the US as the ban makes its way through the court system.

President Donald Trump. Drew Angerer/Getty Images

- Judges exercise power and discretion in policymaking. (e.g. Epstein et al. 2013)
  - **Interpret**, apply, create law and legal precedent under uncertainty.
    - **Subjective** decision-making creates a role for schools of thinking.
  - e.g. Originalism, Critical Legal Theory, **Law and Economics**.
- Can embeddings models help us measure schools of thought?

U.S. Courts as "Natural Laboratory"

- Do schools of thought matter for policymaking?
    - We have recently seen the importance of US federal courts ruling against Trump.
    - These courts involve expert decision-making with far-reaching implications.



Federal appeals court rules against Trump, refuses to reinstate travel ban

Mark Abadi
Feb. 9, 2017, 6:25 PM 35,432

FACEBOOK LINKEDIN TWITTER EMAIL PRINT

- The Best Tom Hanks Movies Ever (PrettyFamous — By Graphiq)
- Best Flashlight Ever is Selling Like... (Gladiator Flashlights)
- End Your Nightly Snoring Nightmar... (My Snoring Solution)
- This amazing new technology allo... (Flare Audio)

Sponsored Links

A federal appeals court unanimously ruled against President Donald Trump on Thursday, refusing to reinstate his travel ban.

The ruling, issued by a three-judge panel on the San Francisco-based 9th Circuit Court of Appeals, means refugees and citizens of the seven majority-Muslim countries affected by the ban can continue entering the US as the ban makes its way through the court system.

President Donald Trump. Drew Angerer/Getty Images

- Judges exercise power and discretion in policymaking. (e.g. Epstein et al. 2013)
    - **Interpret**, apply, create law and legal precedent under uncertainty.
        - **Subjective** decision-making creates a role for schools of thinking.
    - e.g. Originalism, Critical Legal Theory, **Law and Economics**.
- Can embeddings models help us measure schools of thought?

Legal Embeddings

- How can embeddings models be used to understand the law – legal language and legal reasoning?
  - Are legal ideologies also encoded in the vector space?
  - Is there a vector direction for "law and economics"? For originalism?
  - Once we know this vector direction, can we say

$$\text{``}\mathit{Ginsburg}\text{''} + \text{``}\mathit{Economics}\text{''} = \text{``}\mathit{Gorsuch}\text{''}$$

Legal Embeddings

- How can embeddings models be used to understand the law – legal language and legal reasoning?
  - Are legal ideologies also encoded in the vector space?
  - Is there a vector direction for "law and economics"? For originalism?
  - Once we know this vector direction, can we say

$$\text{``}\textit{Ginsburg}\text{''} + \text{``}\textit{Economics}\text{''} = \text{``}\textit{Gorsuch}\text{''}$$

Bird's Eye View

- To fix ideas:
  - opinion i,
  - written by judge j, with characteristics X_j
  - at time t
  - in court/jurisdiction c.
- An opinion is a vector of features Y_i:
  - ruling (affirm/reverse)
  - text features of the opinion
  - set of citations to previous opinions.
- We also have D_i, a vector of (text and metadata) features describing the trial-court opinion
- We want to model

$$Y_i \sim F(D_i, X_j, c, t)$$

where $F(\cdot)$ is some distribution over opinion features we can approximate using deep neural nets or some other machine learning technique.

Bird's Eye View

- To fix ideas:
  - opinion i,
  - written by judge j, with characteristics X_j
  - at time t
  - in court/jurisdiction c.
- An opinion is a vector of features Y_i:
  - ruling (affirm/reverse)
  - text features of the opinion
  - set of citations to previous opinions.
- We also have D_i, a vector of (text and metadata) features describing the trial-court opinion
- We want to model

$$Y_i \sim F(D_i, X_j, c, t)$$

where $F(\cdot)$ is some distribution over opinion features we can approximate using deep neural nets or some other machine learning technique.

Bird's Eye View

- To fix ideas:
  - opinion i,
  - written by judge j, with characteristics X_j
  - at time t
  - in court/jurisdiction c.
- An opinion is a vector of features Y_i:
  - ruling (affirm/reverse)
  - text features of the opinion
  - set of citations to previous opinions.
- We also have D_i, a vector of (text and metadata) features describing the trial-court opinion
- We want to model

$$Y_i \sim F(D_i, X_j, c, t)$$

where $F(\cdot)$ is some distribution over opinion features we can approximate using deep neural nets or some other machine learning technique.

Bird's Eye View

- To fix ideas:
  - opinion i,
  - written by judge j, with characteristics X_j
  - at time t
  - in court/jurisdiction c.
- An opinion is a vector of features Y_i:
  - ruling (affirm/reverse)
  - text features of the opinion
  - set of citations to previous opinions.
- We also have D_i, a vector of (text and metadata) features describing the trial-court opinion
- We want to model

$$Y_i \sim F(D_i, X_j, c, t)$$

where $F(\cdot)$ is some distribution over opinion features we can approximate using deep neural nets or some other machine learning technique.

What would this model do?

$$Y_i \sim F(D_i, X_j, c, t)$$

- This model could be used to simulate counterfactuals:
  - How would the decision in a case change by switching out the authoring judge j?
  - How would the style of language change for a different circuit c?
- We want to represent the output, Y_i, and the inputs, D_i, and X_j, as embeddings:
  - for prediction accuracy
  - to exploit geometric relations between data points

What would this model do?

$$Y_i \sim F(D_i, X_j, c, t)$$

- This model could be used to simulate counterfactuals:
  - How would the decision in a case change by switching out the authoring judge j?
  - How would the style of language change for a different circuit c?
- We want to represent the output, Y_i, and the inputs, D_i, and X_j, as embeddings:
  - for prediction accuracy
  - to exploit geometric relations between data points

Outline

Our Data

- 380,000 cases from Federal Circuit Courts.
- Biographical features of the 268 judges in our sample
- For the demonstrations, we took 212,101 opinions for 1970-2013.
  - We added 3,647 Supreme Court case opinions from 1970-2013.

Approach: doc2vec

- We trained doc2vec on the corpus of opinions, treating a paragraph as a document.
- Case level data:
  - Take the average of the vectors of the paragraph of the opinion
  - These vectors can predict the court decision (for or against a government agency) with 70% accuracy.
- Judge-time data:
  - We de-meaned case level vectors by topic
  - We constructed judge-level vectors for five-year time windows

Approach: doc2vec

- We trained doc2vec on the corpus of opinions, treating a paragraph as a document.
- Case level data:
  - Take the average of the vectors of the paragraph of the opinion
  - These vectors can predict the court decision (for or against a government agency) with 70% accuracy.
- Judge-time data:
  - We de-meaned case level vectors by topic
  - We constructed judge-level vectors for five-year time windows

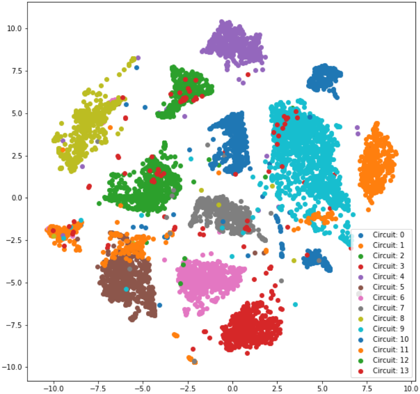
Approach: doc2vec

- We trained doc2vec on the corpus of opinions, treating a paragraph as a document.
- Case level data:
  - Take the average of the vectors of the paragraph of the opinion
  - These vectors can predict the court decision (for or against a government agency) with 70% accuracy.
- Judge-time data:
  - We de-meaned case level vectors by topic
  - We constructed judge-level vectors for five-year time windows

Approach: doc2vec

- We trained doc2vec on the corpus of opinions, treating a paragraph as a document.
- Case level data:
    - Take the average of the vectors of the paragraph of the opinion
    - These vectors can predict the court decision (for or against a government agency) with 70% accuracy.
- Judge-time data:
    - We de-meaned case level vectors by topic
    - We constructed judge-level vectors for five-year time windows

Visual Structure of Judge Embeddings



- Circuits cluster together.
- There is spread of Supreme Court (red dots) across the clusters.
- 11th Circuit (orange dots) is split into multiple clusters; these judges overlap with the 5th Circuit (brown dots); the 11th Circuit split off from the 5th Circuit in 1982 and uses pre-1982 5th Circuit cases as precedent.

Potential Refinements

- Down-weighting or exclusion of identifying or personal language (e.g. "Ginsburg", "Scalia")
- Up-weighting of ideological language (e.g. "First Amendment", "optimal deterrence")
- Integration of citation network information (citing Ginsburg vs Scalia)

Outline

Manne Program: Economics Institute for Judges

"FROM THE BEGINNING, THE JUDGES DEFERRED TO THEIR TEACHERS," wrote a *New York Times* reporter. Below, Nobel Laureate Milton Friedman elaborates a point at an LEC

Economics Institute for Federal Judges. Thirty-nine judges were graduated from this intensive two-week

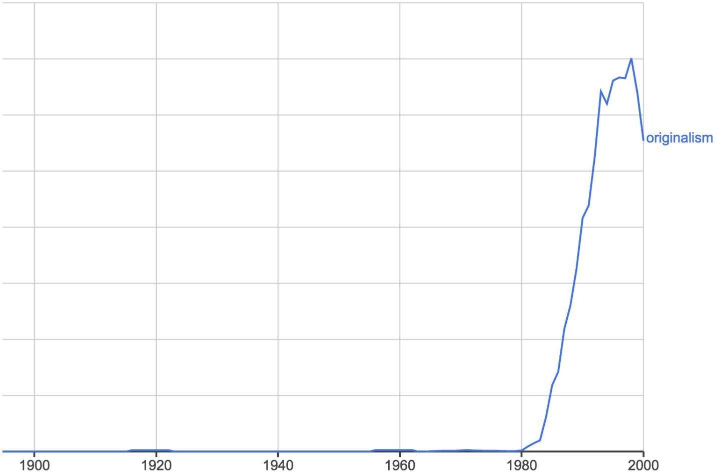
program of study of market economics in the Center's fourth year, bringing to 58 the total number of

Impact of Economics Judges – Highlights

- Summary Correlations
  - Economics Training correlated with Economics Style
    - ◂ Both are independently correlated (but not synonymous) with Republican Party
  - Economics Trained Judges vote against regulation and reject criminal appeals
    - ◂ Economics is more predictive than Republican Party
- Economics Judges' Impact on Economics Cases
  - ◂ Event study
- Economics Judges Impact on Criminal Cases
  - ◂ Training immediately increases sentence lengths in event study

Outline

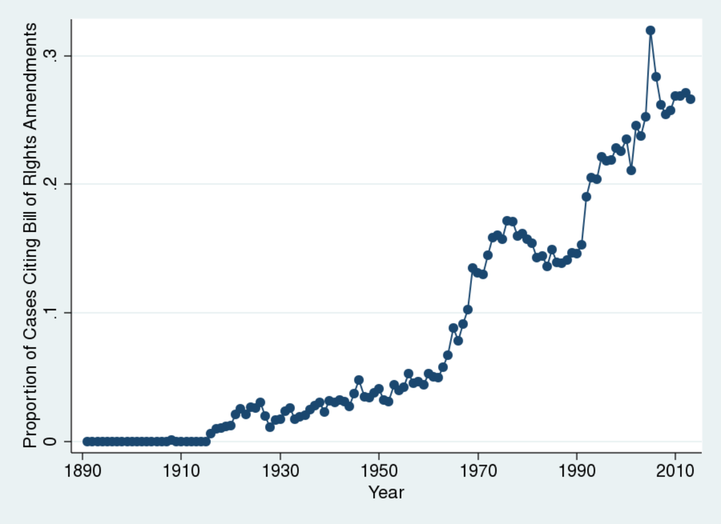
Originalism



- The word "originalism" was coined by Paul Brest in 1980: *"By “originalism” I mean the familiar approach to constitutional adjudication that accords binding authority to the text of the Constitution or the intentions of its adopters."*
- Is there a vector direction for originalist principles?

Measuring Originalism

Figure: Trend in Citing Bill of Rights Amendments



Most Originalist Circuit Court Judges

Rank	Judge	Originalism Score
1	DUNCAN, ALLYSON	6.76
2	RAWLINSON, JOHN	6.08
3	SYKES, DIANE S.	5.29
4	**SCALIA, ANTONIN**	**5.13**
5	PARKER, BARRINGTON	4.76
6	MARCUS, STANLEY	4.33
7	LINN, RICHARD	3.88
8	LEMMON, DAL	3.78
9	GRABER, SUSAN	3.43
10	HARDIMAN, THOMAS	3.36
11	WESLEY, RICHARD	3.19
12	SACK, ROBERT DAVID	3.17
13	CLEVENGER, RAYMOND	3.13
14	MCKEAGUE, DAVID	2.77
15	**GARLAND, MERRICK**	**2.67**
16	KETHLEDGE, RAYMOND	2.30
17	**GORSUCH, NEIL M.**	**2.28**
18	CLAY, ERIC L.	2.24
	...	
	SOTOMAYOR, SONIA	**0.26**
	POSNER, RICHARD A.	**-0.4**

Outline

Word Embedding Association Test

Sentiment Attribute Words	
joy, love, peace, wonderful,	agony, terrible, horrible, nasty,
pleasure, friend, laughter, happy	evil, war, awful, failure

Implicit Sexism Target Words	
male, man, boy, brother,	female, woman, girl, sister,
he, him, his, son	she, her, hers, daughter

Implicit Racism Target Words	
european, white, caucasian	black, african, negro

- Compute "Assocation" as the average word-vector similarities between a group of target words and a group of attribute words.

$$\text{Implicit Sexism} = \frac{\text{Male-Pleasant Association}}{\text{Male-Unpleasant Association}} \Big/ \frac{\text{Female-Pleasant Association}}{\text{Female-Unleasant Association}}$$

$$\text{Implicit Racism} = \frac{\text{White-Pleasant Association}}{\text{White-Unpleasant Association}} \Big/ \frac{\text{Black-Pleasant Association}}{\text{Black-Unleasant Association}}$$

- We compute judge WEAT scotes by training a Word2Vec model separately by judge

Word Embedding Association Test

Sentiment Attribute Words	
joy, love, peace, wonderful,	agony, terrible, horrible, nasty,
pleasure, friend, laughter, happy	evil, war, awful, failure

Implicit Sexism Target Words	
male, man, boy, brother,	female, woman, girl, sister,
he, him, his, son	she, her, hers, daughter

Implicit Racism Target Words	
european, white, caucasian	black, african, negro

- Compute "Assocation" as the average word-vector similarities between a group of target words and a group of attribute words.

$$\text{Implicit Sexism} = \frac{\text{Male-Pleasant Association}}{\text{Male-Unpleasant Association}} \Big/ \frac{\text{Female-Pleasant Association}}{\text{Female-Unleasant Association}}$$

$$\text{Implicit Racism} = \frac{\text{White-Pleasant Association}}{\text{White-Unpleasant Association}} \Big/ \frac{\text{Black-Pleasant Association}}{\text{Black-Unleasant Association}}$$

- We compute judge WEAT scotes by training a Word2Vec model separately by judge

Word Embedding Association Test

Sentiment Attribute Words	
joy, love, peace, wonderful,	agony, terrible, horrible, nasty,
pleasure, friend, laughter, happy	evil, war, awful, failure

Implicit Sexism Target Words	
male, man, boy, brother,	female, woman, girl, sister,
he, him, his, son	she, her, hers, daughter

Implicit Racism Target Words	
european, white, caucasian	black, african, negro

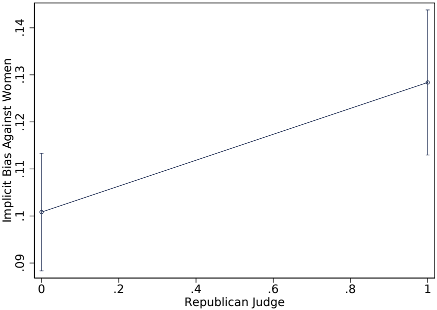
- Compute "Assocation" as the average word-vector similarities between a group of target words and a group of attribute words.

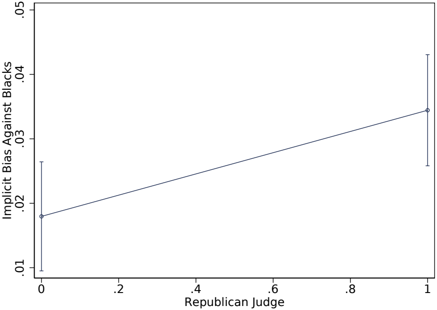
$$\text{Implicit Sexism} = \frac{\text{Male-Pleasant Association}}{\text{Male-Unpleasant Association}} \Big/ \frac{\text{Female-Pleasant Association}}{\text{Female-Unleasant Association}}$$

$$\text{Implicit Racism} = \frac{\text{White-Pleasant Association}}{\text{White-Unpleasant Association}} \Big/ \frac{\text{Black-Pleasant Association}}{\text{Black-Unleasant Association}}$$

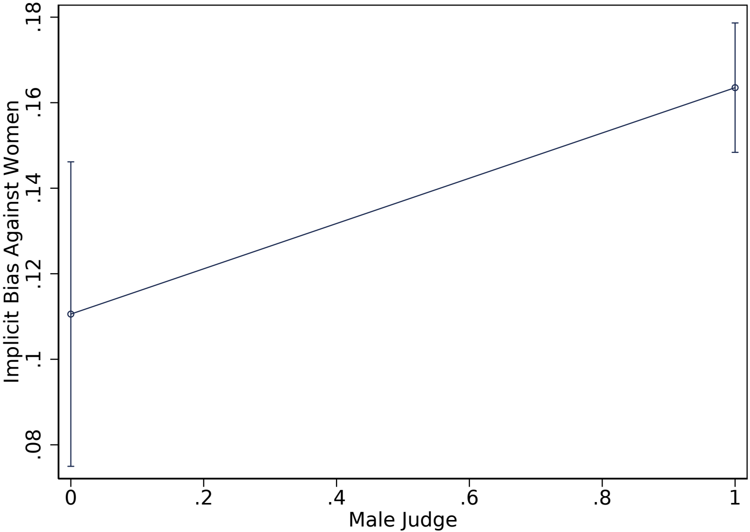
- We compute judge WEAT scotes by training a Word2Vec model separately by judge

Republican judges have higher gender bias and race bias

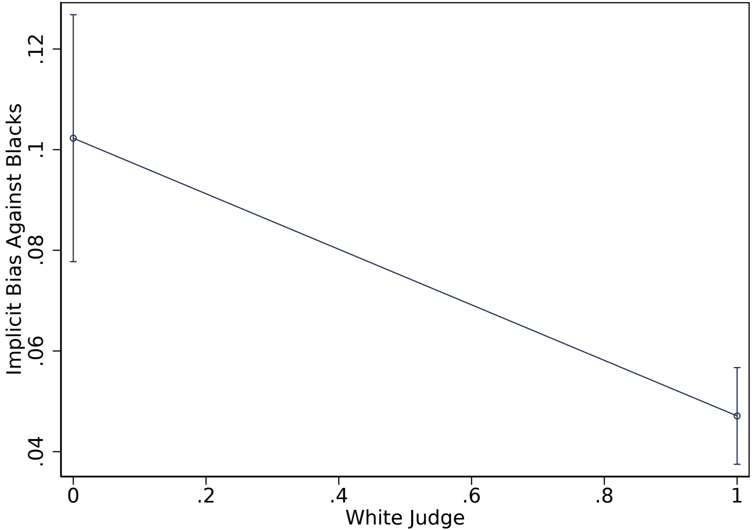




Male judges have higher gender bias than female judges



White judges have *lower* race bias than black judges



Trump nominees have high race and gender, but not government, bias

President Donald J. Trump's Supreme Court List

Amy Coney Barrett of Indiana, U.S. Court of Appeals for the Seventh Circuit

Keith Blackwell of Georgia, Supreme Court of Georgia

Charles Canady of Florida, Supreme Court of Florida

Steven Colloton of Iowa, U.S. Court of Appeals for the Eighth Circuit

Allison Eid of Colorado, U.S. Court of Appeals for the Tenth Circuit

Britt Grant of Georgia, Supreme Court of Georgia

Raymond Gruender of Missouri, U.S. Court of Appeals for the Eighth Circuit

Thomas Hardiman of Pennsylvania, U.S. Court of Appeals for the Third Circuit

Brett Kavanaugh of Maryland, U.S. Court of Appeals for the District of Columbia Circuit

Raymond Kethledge of Michigan, U.S. Court of Appeals for the Sixth Circuit

Joan Larsen of Michigan, U.S. Court of Appeals for the Sixth Circuit

Mike Lee of Utah, United States Senator

Thomas Lee of Utah, Supreme Court of Utah

Edward Mansfield of Iowa, Supreme Court of Iowa

Federico Moreno of Florida, U.S. District Court for the Southern District of Florida

Kevin Newsom of Alabama, U.S. Court of Appeals for the Eleventh Circuit

William Pryor of Alabama, U.S. Court of Appeals for the Eleventh Circuit

Margaret Ryan of Virginia, U.S. Court of Appeals for the Armed Forces

David Stras of Minnesota, U.S. Court of Appeals for the Eighth Circuit

Diane Sykes of Wisconsin, U.S. Court of Appeals for the Seventh Circuit

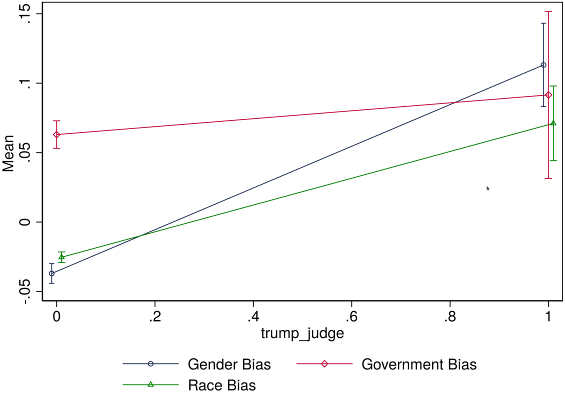
Amul Thapar of Kentucky, U.S. Court of Appeals for the Sixth Circuit

Timothy Tymkovich of Colorado, U.S. Court of Appeals for the Tenth Circuit

Robert Young of Michigan, Supreme Court of Michigan (Ret.)

Don Willett of Texas, Supreme Court of Texas

Patrick Wyrick of Oklahoma, Supreme Court of Oklahoma



Outline

Outlook

- This paper has explored recent advances in embeddings models and discussed their potential for legal scholarship.
    - There is clear potential for using these methods to understand better the relations between judges and to predict their decision-making.
- New corpora and new computational models will lead to a richer understanding of law and the legal system.

Outlook

- This paper has explored recent advances in embeddings models and discussed their potential for legal scholarship.
  - There is clear potential for using these methods to understand better the relations between judges and to predict their decision-making.
- New corpora and new computational models will lead to a richer understanding of law and the legal system.

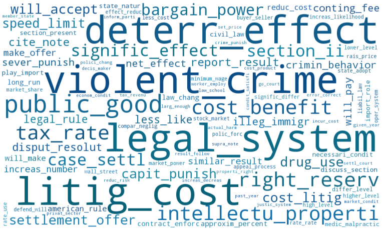
Outlook

- This paper has explored recent advances in embeddings models and discussed their potential for legal scholarship.
  - There is clear potential for using these methods to understand better the relations between judges and to predict their decision-making.
- New corpora and new computational models will lead to a richer understanding of law and the legal system.

Outline

Law-and-Economics Language

- All available JSTOR articles with JEL K (Law and Economics) (1991-2008)
  - Highest and lowest frequencies for two-grams in $\geq$ 1000 cases:



Most similar to Law-Econ Corpus

Least similar to Law-Econ Corpus

- Law-Econ: deterrent effect, cost-benefit, public goods, bargaining power, litigation costs
  - violent crime, criminal behavior, capital punishment, illegal immigration
- Non-LE: find reason, find fact, fail establish, substantive / sufficient / argue evidence
  - evidence and other constitutional theories of interpretation seem less salient

Scoring Judges By Economics Style

- E_g: relative frequencies for phrase g in JEL K
- $F_i = \{F_{i1}, F_{i2}, ..., F_{iP}\}$: relative frequencies for phrase g in case i
  - Economics Style of case i is cosine similarity to economics corpus (average econ score of its phrases):

$$z_i = \frac{F_i \cdot E}{\|F_i\| \, \|E\|}$$

- Score judges by their use of economics language: ◂ Methodology
  - Residualize z_i on circuit-year fixed effects to control for case portfolio
  - J_j: set of n_j cases authored by judge j. Economics Style of judge j is:

$$Z_{jt} = \frac{1}{n_j} \sum_{i \in J_j} z_i$$

Scoring Judges By Economics Style

- E_g: relative frequencies for phrase g in JEL K
- $F_i = \{F_{i1}, F_{i2}, ..., F_{iP}\}$: relative frequencies for phrase g in case i
  - Economics Style of case i is cosine similarity to economics corpus (average econ score of its phrases):

$$z_i = \frac{F_i \cdot E}{\|F_i\| \, \|E\|}$$

- Score judges by their use of economics language: ◂ Methodology
  - Residualize z_i on circuit-year fixed effects to control for case portfolio
  - J_j: set of n_j cases authored by judge j. Economics Style of judge j is:

$$Z_{jt} = \frac{1}{n_j} \sum_{i \in J_j} z_i$$

Scoring Judges By Economics Style

- E_g: relative frequencies for phrase g in JEL K
- $F_i = \{F_{i1}, F_{i2}, ..., F_{iP}\}$: relative frequencies for phrase g in case i
  - Economics Style of case i is cosine similarity to economics corpus (average econ score of its phrases):

$$z_i = \frac{F_i \cdot E}{\|F_i\| \, \|E\|}$$

- Score judges by their use of economics language: Methodology
  - Residualize z_i on circuit-year fixed effects to control for case portfolio
  - J_j: set of n_j cases authored by judge j. Economics Style of judge j is:

$$Z_{jt} = \frac{1}{n_j} \sum_{i \in J_j} z_i$$

Empirical Approach

- The coefficient γ gives the causal effect of judge-assignment
  - case i, judge j, court c, year t randomization check

$$Y_{ijct} = \alpha_{ct} + \gamma Z_{ijt} + X_j' \beta + \varepsilon_{ijct}$$

- Outcome Y_{ijct} measured four ways:
  - (1) $1 =$ conservative vote, $-1 =$ liberal vote (Songer-Auburn 5%, hand-labeled)
  - (2) Voting against government regulatory agencies (100%, machine-coded)
  - (3) Rejecting criminal appeals (100%, machine-coded)
    - from gov't in title of case, Π vs. △, for (2) Economics, Labor, and (3) Criminal Appeals cases
  - (4) Length of criminal sentence (100%, FOIA requested to include judge identity)
- Z_{ijt}, law-and-economics thinking of judge j:
  - Economics Style (leave-one-out mean $Z_{ijt} = \sum_{k \in J_i^j}^{J} \frac{z_k}{|J_i^j|}$)
  - Economics Training (1976-1999; 2000-2009; yr-by-yr to present)
  - Treatment is judge; so cluster by judge; weight to treat judge-years equally
- Controls
  - α_{ct}: court-year fixed effects Methodology
  - X_j: judge covariates, e.g. Republican (benchmark for Economics Training)

Empirical Approach

- The coefficient γ gives the causal effect of judge-assignment
  - case i, judge j, court c, year t randomization check

$$Y_{ijct} = \alpha_{ct} + \gamma Z_{ijt} + X_j'\beta + \varepsilon_{ijct}$$

- Outcome Y_{ijct} measured four ways:
  - (1) $1 =$ conservative vote, $-1 =$ liberal vote (Songer-Auburn 5%, hand-labeled)
  - (2) Voting against government regulatory agencies (100%, machine-coded)
  - (3) Rejecting criminal appeals (100%, machine-coded)
    - from gov't in title of case, Π vs. $\triangle$, for (2) Economics, Labor, and (3) Criminal Appeals cases
  - (4) Length of criminal sentence (100%, FOIA requested to include judge identity)
- Z_{ijt}, law-and-economics thinking of judge j:
  - Economics Style (leave-one-out mean $Z_{ijt} = \sum_{k \in J_i^j}^{J} \frac{z_k}{|J_i^j|}$)
  - Economics Training (1976-1999; 2000-2009; yr-by-yr to present)
  - Treatment is judge; so cluster by judge; weight to treat judge-years equally
- Controls
  - α_{ct}: court-year fixed effects Methodology
  - X_j: judge covariates, e.g. Republican (benchmark for Economics Training)

Empirical Approach

- The coefficient γ gives the causal effect of judge-assignment
  - case i, judge j, court c, year t randomization check

$$Y_{ijct} = \alpha_{ct} + \gamma Z_{ijt} + X_j' \beta + \varepsilon_{ijct}$$

- Outcome Y_{ijct} measured four ways:
  - (1) $1 =$ conservative vote, $-1 =$ liberal vote (Songer-Auburn 5%, hand-labeled)
  - (2) Voting against government regulatory agencies (100%, machine-coded)
  - (3) Rejecting criminal appeals (100%, machine-coded)
    - from gov't in title of case, Π vs. $\triangle$, for (2) Economics, Labor, and (3) Criminal Appeals cases
  - (4) Length of criminal sentence (100%, FOIA requested to include judge identity)
- Z_{ijt}, law-and-economics thinking of judge j:
  - Economics Style (leave-one-out mean $Z_{ijt} = \sum_{k \in J_i^j}^{J} \frac{z_k}{|J_i^j|}$)
  - Economics Training (1976-1999; 2000-2009; yr-by-yr to present)
  - Treatment is judge; so cluster by judge; weight to treat judge-years equally
- Controls
  - α_{ct}: court-year fixed effects Methodology
  - X_j: judge covariates, e.g. Republican (benchmark for Economics Training)

Empirical Approach

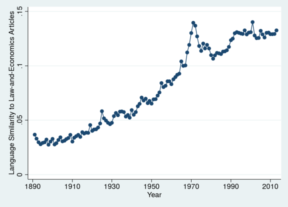
- The coefficient γ gives the causal effect of judge-assignment
  - case i, judge j, court c, year t ◂ randomization check

$$Y_{ijct} = \alpha_{ct} + \gamma Z_{ijt} + X_j'\beta + \varepsilon_{ijct}$$

- Outcome Y_{ijct} measured four ways:
  - (1) $1 =$ conservative vote, $-1 =$ liberal vote (Songer-Auburn 5%, hand-labeled)
  - (2) Voting against government regulatory agencies (100%, machine-coded)
  - (3) Rejecting criminal appeals (100%, machine-coded)
    - from gov't in title of case, Π vs. $\triangle$, for (2) Economics, Labor, and (3) Criminal Appeals cases
  - (4) Length of criminal sentence (100%, FOIA requested to include judge identity)
- Z_{ijt}, law-and-economics thinking of judge j:
  - Economics Style (leave-one-out mean $Z_{ijt} = \sum_{k \in J_i^j}^{J} \frac{z_k}{|J_i^j|}$)
  - Economics Training (1976-1999; 2000-2009; yr-by-yr to present)
  - Treatment is judge; so cluster by judge; weight to treat judge-years equally
- Controls
  - α_{ct}: court-year fixed effects ◂ Methodology
  - X_j: judge covariates, e.g. Republican (benchmark for Economics Training)

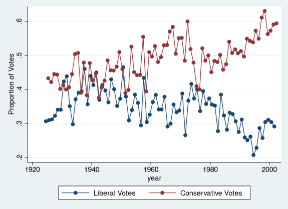
Increasing Conservativism in Federal Judiciary

Use randomly assigned judges to isolate causal effect of panel 1 on panels 2 and 3



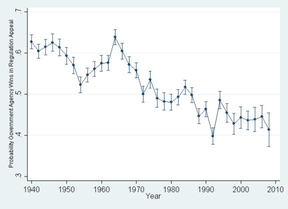
Language similarity to

law-and-economics articles



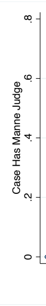
Conservative Votes

conservatism definition



Voting for government regulation

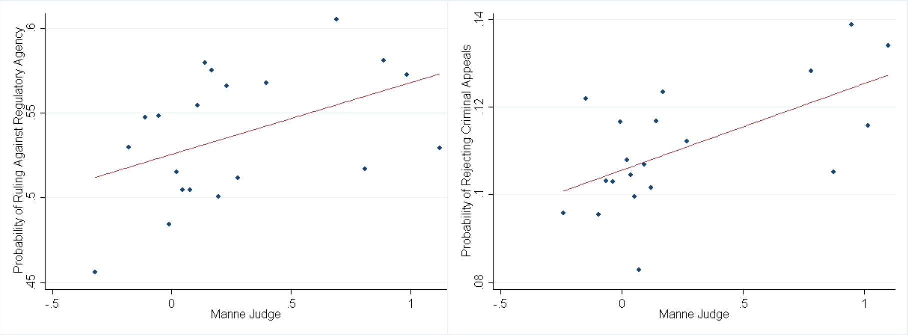
Lor



Benchmark Effect of Economics (vs. Republican)

	Ruling Against Regulatory Agency				Rejecting Criminal Appeal			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Econ Style	**0.00554****	**0.00533****			**0.00250***	**0.00222***		
	(0.00245)	**(0.00243)**			**(0.00132)**	**(0.00132)**		
Econ Training			**0.0364***	**0.0425****			**0.0199****	**0.0220*****
			(0.0208)	**(0.0212)**			**(0.00774)**	**(0.00781)**
Republican		-0.00752		-0.0333		-0.00963***		-0.0164***
		(0.00750)		(0.0208)		(0.00333)		(0.00630)
N	53977	53977	12320	12320	194070	194070	97824	97824
adj. R-sq	0.100	0.100	0.173	0.173	0.239	0.239	0.043	0.043
Circuit-Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Sample	All	All	Post 1991		All	All	Post 1991	

Benchmark Effect of Economics (vs. Republican)

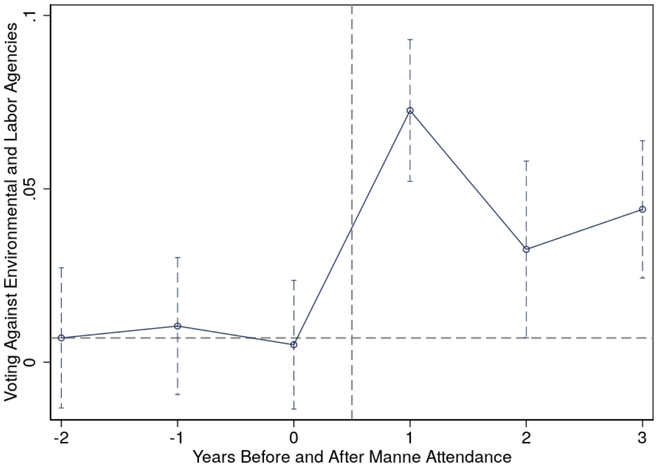


Economics Trained Judges vote against regulation and reject criminal appeals.

Binscatter: Probability vs. economics training, residualized on circuit-year fixed effects and Republican indicator

Highlights

Impact of Economics Judges on Environment/Labor



Residuals from regression of vote-against-government on circuit-year FEs, judge FEs, and party-year FEs, plotted by years before and after Manne attendance. Spikes give 90% confidence intervals. Alleviates selection concern.

Impact of Economics Judges on Regulation Cases

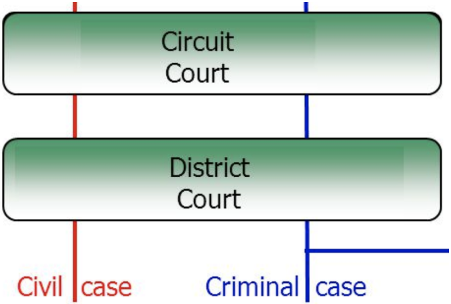
	# Uses of "Efficient"		
	(1)	(2)	(3)
Econ Training	-0.00407	**0.0494*****	
	(0.00455)	**(0.0188)**	
Econ Training *			**0.0495***
Post 1991			**(0.0272)**
N	45752	11372	72005
adj. R-sq	0.125	0.148	0.261
Circuit-Year FE	Y	Y	Y
Control	N	N	N
Judge FE	N	N	Y
Sample	Year < 1976	**Year > 1991**	All

Similar with Republican control. ◂ Highlights

Identifying Memetic Economic Phrases, All Cases

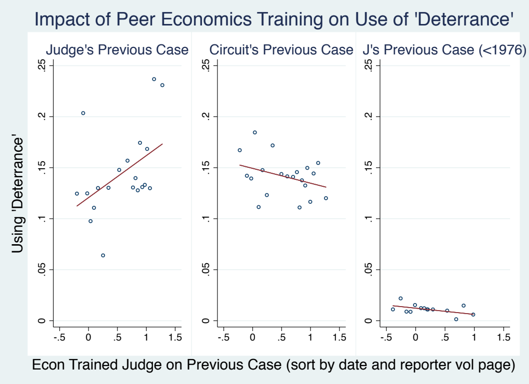
	# Uses of "Deterrence"			
Econ Training on	(1)	(2)	(3)	(4)
Next Case	-0.00412			
	(0.00730)			
This Case		**0.0161****		
		(0.00683)		
Previous Case			**0.0127***	
			(0.00692)	
Two Cases Ago				**0.0120***
				(0.00678)
N	353981	355504	354695	353928
adj. R-sq	0.009	0.010	0.010	0.010
Circuit-Year FE	Y	Y	Y	Y
Circuit Order	Y	Y	Y	Y
Sample	**Year > 1991**	**Year > 1991**	**Year > 1991**	**Year > 1991**
Order within	Judge	Judge	Judge	Judge
Cluster	Judge	Judge	Judge	Judge

Impact of Economics Judges, Criminal Cases



Federal courts handle the most serious criminal cases (8% of US prison population).

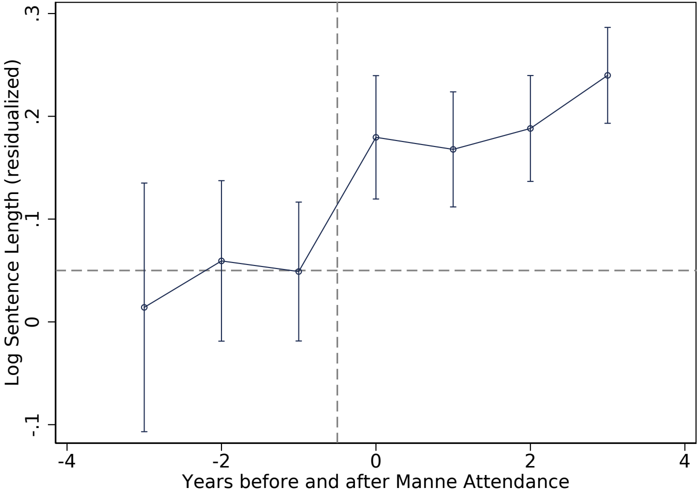
Impact of Peer Econ Judges on Criminal Case Reasoning



Previous judge case (median) 9 days ago; previous circuit case (median) 2 days ago. Exclude same day cases.

◂ Impacts on Criminal Cases

Manne Attendance on Criminal Sentencing (Event Study)



Sentence length (residualized on year fixed effects), plotted by years before and after Manne attendance, for judges who attended, 1992-2003.

Impacts on Criminal Cases

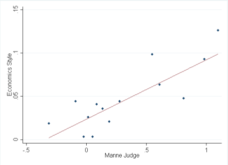
Implicit Attitudes

- The text of the opinions provide a window into rich representations of legal/political institutions, as we well as **human social psychology**.
- Caliskan, Bryson, and Narayanan (Science 2017) show that implicit gender and racial biases are embedded in human language.
  - We ask whether this implicit language bias varies across judges.

Implicit Attitudes

- The text of the opinions provide a window into rich representations of legal/political institutions, as we well as **human social psychology**.
- Caliskan, Bryson, and Narayanan (Science 2017) show that implicit gender and racial biases are embedded in human language.
  - We ask whether this implicit language bias varies across judges.

Summary Correlations



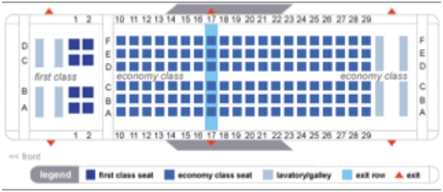
Economics Training correlated with Economics Style

	Republican		
	(1)	(2)	(3)
Economics Style	0.0367*		**0.0563****
	(0.0146)		**(0.0191)**
Economics Training		0.140**	**0.191****
		(0.0382)	**(0.0602)**
N	923866	410309	380085
adj. R-sq	0.137	0.082	0.099

0.2 Correlation between Economics Training and Republican Party

Highlights

Identification of Learning & Memetic Effects



The coefficient γ gives the causal effect of judge-assignment

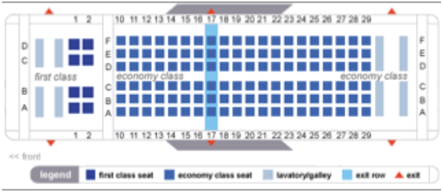
- case i, judge j, court c, year t

$$F_{ijct} = \alpha_{ct} + \gamma Z_{ijct} + X_j' \beta + \varepsilon_{ijct}$$

- Z_{ijct}, law-and-economics exposure:
  - γ_1: Presence of Economics Training on the **Previous Case of this Judge**
  - Presence of Economics Training on the **Previous Case in this Circuit**
  - γ_2: Presence of Economics Training on the **Previous Case of Judge on Topic**
  - Presence of Economics Training on the **Previous Case of Circuit on Topic**
    - Separately identify the impact within topic (γ_2) vs. across topic (γ_1)
- Active v. Passive Persuasion (Was previous case divided? $\hat{p}$(citation, reversal, dissent))

Impacts of Peer Training

Identification of Learning & Memetic Effects



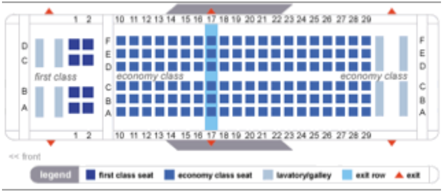
The coefficient γ gives the causal effect of judge-assignment

- case i, judge j, court c, year t

$$F_{ijct} = \alpha_{ct} + \gamma Z_{ijct} + X_j'\beta + \varepsilon_{ijct}$$

- Z_{ijct}, law-and-economics exposure:
  - γ_1· Presence of Economics Training on the **Previous Case of this Judge**
  - Presence of Economics Training on the **Previous Case in this Circuit**
  - γ_2· Presence of Economics Training on the **Previous Case of Judge on Topic**
  - Presence of Economics Training on the **Previous Case of Circuit on Topic**
    - Separately identify the impact within topic (γ_2) vs. across topic (γ_1)
- Active v. Passive Persuasion (Was previous case divided? $\hat{p}$(citation, reversal, dissent))

Impacts of Peer Training



The coefficient γ gives the causal effect of judge-assignment

- case i, judge j, court c, year t

$$F_{ijct} = \alpha_{ct} + \gamma Z_{ijct} + X_j'\beta + \varepsilon_{ijct}$$

- Z_{ijct}, law-and-economics exposure:
  - γ_1· Presence of Economics Training on the **Previous Case of this Judge**
  - Presence of Economics Training on the **Previous Case in this Circuit**
  - γ_2· Presence of Economics Training on the **Previous Case of Judge on Topic**
  - Presence of Economics Training on the **Previous Case of Circuit on Topic**
    - Separately identify the impact within topic (γ_2) vs. across topic (γ_1)
- Active v. Passive Persuasion (Was previous case divided? $\hat{p}$(citation, reversal, dissent))

◂ Impacts of Peer Training