"Judicial Decisions of the European Court of Human Rights: Looking into the Crystal Ball" by Masha Medvedeva, Michel Vols, and Martijn Wieling

Discussion prepared by Daniel L. Chen

May 2018

Research Question

N-grams to predict decisions

- Average accuracy of 75% in predicting violations
- Average accuracy of 70% in predicting future violations

1) But higher accuracy achieved by guessing "violation"

article	violation	non-violation
Article 2	559	161
Article 3	1446	595
Article 5	1511	393
Article 6	4828	736
Article 8	854	358
Article 10	394	142
Article 11	131	42
Article 13	1230	170
Article 14	195	239
	80%	**20%**

2) What is the social science or policy application?

Research Question

N-grams to predict decisions

- Average accuracy of 75% in predicting violations
- Average accuracy of 70% in predicting future violations

1) But higher accuracy achieved by guessing "violation"

article	violation	non-violation
Article 2	559	161
Article 3	1446	595
Article 5	1511	393
Article 6	4828	736
Article 8	854	358
Article 10	394	142
Article 11	131	42
Article 13	1230	170
Article 14	195	239
	80%	**20%**

2) What is the social science or policy application?

Research Question

N-grams to predict decisions

- Average accuracy of 75% in predicting violations
- Average accuracy of 70% in predicting future violations

1) But higher accuracy achieved by guessing "violation"

article	violation	non-violation
Article 2	559	161
Article 3	1446	595
Article 5	1511	393
Article 6	4828	736
Article 8	854	358
Article 10	394	142
Article 11	131	42
Article 13	1230	170
Article 14	195	239
	80%	**20%**

2) What is the social science or policy application?

Research Question

N-grams to predict decisions

- Average accuracy of 75% in predicting violations
- Average accuracy of 70% in predicting future violations

1) But higher accuracy achieved by guessing "violation"

article	violation	non-violation
Article 2	559	161
Article 3	1446	595
Article 5	1511	393
Article 6	4828	736
Article 8	854	358
Article 10	394	142
Article 11	131	42
Article 13	1230	170
Article 14	195	239
	80%	**20%**

2) What is the social science or policy application?

Research Question

N-grams to predict decisions

- Average accuracy of 75% in predicting violations
- Average accuracy of 70% in predicting future violations

1) But higher accuracy achieved by guessing "violation"

article	violation	non-violation
Article 2	559	161
Article 3	1446	595
Article 5	1511	393
Article 6	4828	736
Article 8	854	358
Article 10	394	142
Article 11	131	42
Article 13	1230	170
Article 14	195	239
	80%	**20%**

2) What is the social science or policy application?

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

1. Finding the degree of similarity between two words.
   `model.similarity('woman','man')`
   `0.73723527`
2. Finding odd one out.
   `model.doesnt_match('breakfast cereal dinner lunch';.split())`
   `'cereal'`
3. Amazing things like woman+king-man =queen
   `model.most_similar(positive=['woman','king'],negative=['man'],topn=1)`
   `queen: 0.508`
4. Probability of a text under the model
   `model.score(['The fox jumped over the lazy dog'.split()])`
   `0.21`

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

1. Finding the degree of similarity between two words.
   `model.similarity('woman','man')`
   `0.73723527`
2. Finding odd one out.
   `model.doesnt_match('breakfast cereal dinner lunch';.split())`
   `'cereal'`
3. Amazing things like woman+king-man =queen
   `model.most_similar(positive=['woman','king'],negative=['man'],topn=1)`
   `queen: 0.508`
4. Probability of a text under the model
   `model.score(['The fox jumped over the lazy dog'.split()])`
   `0.21`

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

1. Finding the degree of similarity between two words.
   `model.similarity('woman','man')`
   `0.73723527`
2. Finding odd one out.
   `model.doesnt_match('breakfast cereal dinner lunch';.split())`
   `'cereal'`
3. Amazing things like woman+king-man =queen
   `model.most_similar(positive=['woman','king'],negative=['man'],topn=1)`
   `queen: 0.508`
4. Probability of a text under the model
   `model.score(['The fox jumped over the lazy dog'.split()])`
   `0.21`

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

1. Finding the degree of similarity between two words.
   `model.similarity('woman','man')`
   `0.73723527`
2. Finding odd one out.
   `model.doesnt_match('breakfast cereal dinner lunch';.split())`
   `'cereal'`
3. Amazing things like woman+king-man =queen
   `model.most_similar(positive=['woman','king'],negative=['man'],topn=1)`
   `queen: 0.508`
4. Probability of a text under the model
   `model.score(['The fox jumped over the lazy dog'.split()])`
   `0.21`

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Alternative Approach: Word Embeddings

- (obama speaks media illinois) is orthogonal to (president greets press chicago) according to cosine similarity
- **word embeddings** capture contextual similarities between words

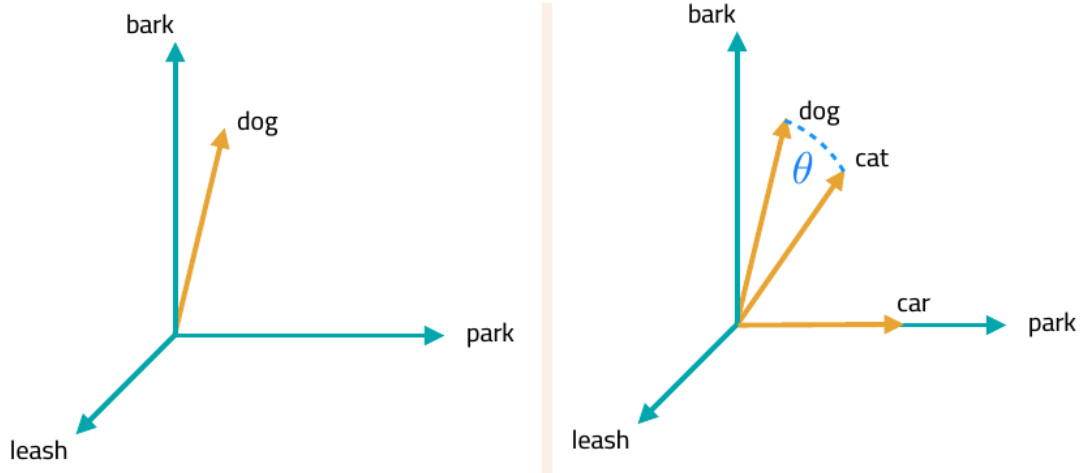
1. Finding the degree of similarity between two words.
```
model.similarity('woman','man')
0.73723527
```
2. Finding odd one out.
```
model.doesnt_match('breakfast cereal dinner
lunch';.split())
'cereal'
```
3. Amazing things like woman+king-man =queen
```
model.most_similar(positive=
['woman','king'],negative=['man'],topn=1)
queen: 0.508
```
4. Probability of a text under the model
```
model.score(['The fox jumped over the lazy
dog'.split()])
0.21
```

- Each word is mapped to one vector, often hundreds of dimensions
  - Contrast to 2B N-grams for sparse word representations
- If we know the words having similar meanings in different languages, word embeddings can be used to (Google) translate!

Relatedness between words

How does it work? Predict given a word using surrounding words

Words as Vectors



- Use cosine similarity as a measure of relatedness:

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

Relatedness between documents

Embeddings are a dimension-reduction approach in deep learning models for prediction (2B vocab v. 200 dimensions)

- identify closest documents
- allows vector math

("Judge Vectors: Spatial Representations of the Law using Document Embeddings"; Ash and Chen, 2018)

- *Everson vs. Board of Education* is to *Engel v. Vitale* as *Griswold v. Connecticut* is to *Roe v. Wade*.
    - application of the constitutional principle articulated in the former

Word embeddings isolate directions for gender, time, plural, etc.

- isolating directions for legal and political concepts
    - liberal vs. conservative, procedural vs. substantive, originalists vs. pragmatists, or economic analysis

Objective

- Predict a given word using its context *and* N-gram representation of whole document

Relatedness between documents

Embeddings are a dimension-reduction approach in deep learning models for prediction (2B vocab v. 200 dimensions)

- identify closest documents
- allows vector math

("Judge Vectors: Spatial Representations of the Law using Document Embeddings"; Ash and Chen, 2018)

- *Everson vs. Board of Education* is to *Engel v. Vitale* as *Griswold v. Connecticut* is to *Roe v. Wade*.
    - application of the constitutional principle articulated in the former

Word embeddings isolate directions for gender, time, plural, etc.

- isolating directions for legal and political concepts
    - liberal vs. conservative, procedural vs. substantive, originalists vs. pragmatists, or economic analysis

Objective

- Predict a given word using its context *and* N-gram representation of whole document

Relatedness between documents

Embeddings are a dimension-reduction approach in deep learning models for prediction (2B vocab v. 200 dimensions)

- identify closest documents
- allows vector math

("Judge Vectors: Spatial Representations of the Law using Document Embeddings"; Ash and Chen, 2018)

- *Everson vs. Board of Education* is to *Engel v. Vitale* as *Griswold v. Connecticut* is to *Roe v. Wade*.
    - application of the constitutional principle articulated in the former

Word embeddings isolate directions for gender, time, plural, etc.

- isolating directions for legal and political concepts
    - liberal vs. conservative, procedural vs. substantive, originalists vs. pragmatists, or economic analysis

Objective

- Predict a given word using its context *and* N-gram representation of whole document

Relatedness between documents

Embeddings are a dimension-reduction approach in deep learning models for prediction (2B vocab v. 200 dimensions)

- identify closest documents
- allows vector math

("Judge Vectors: Spatial Representations of the Law using Document Embeddings"; Ash and Chen, 2018)

- *Everson vs. Board of Education* is to *Engel v. Vitale* as *Griswold v. Connecticut* is to *Roe v. Wade*.
    - application of the constitutional principle articulated in the former

Word embeddings isolate directions for gender, time, plural, etc.

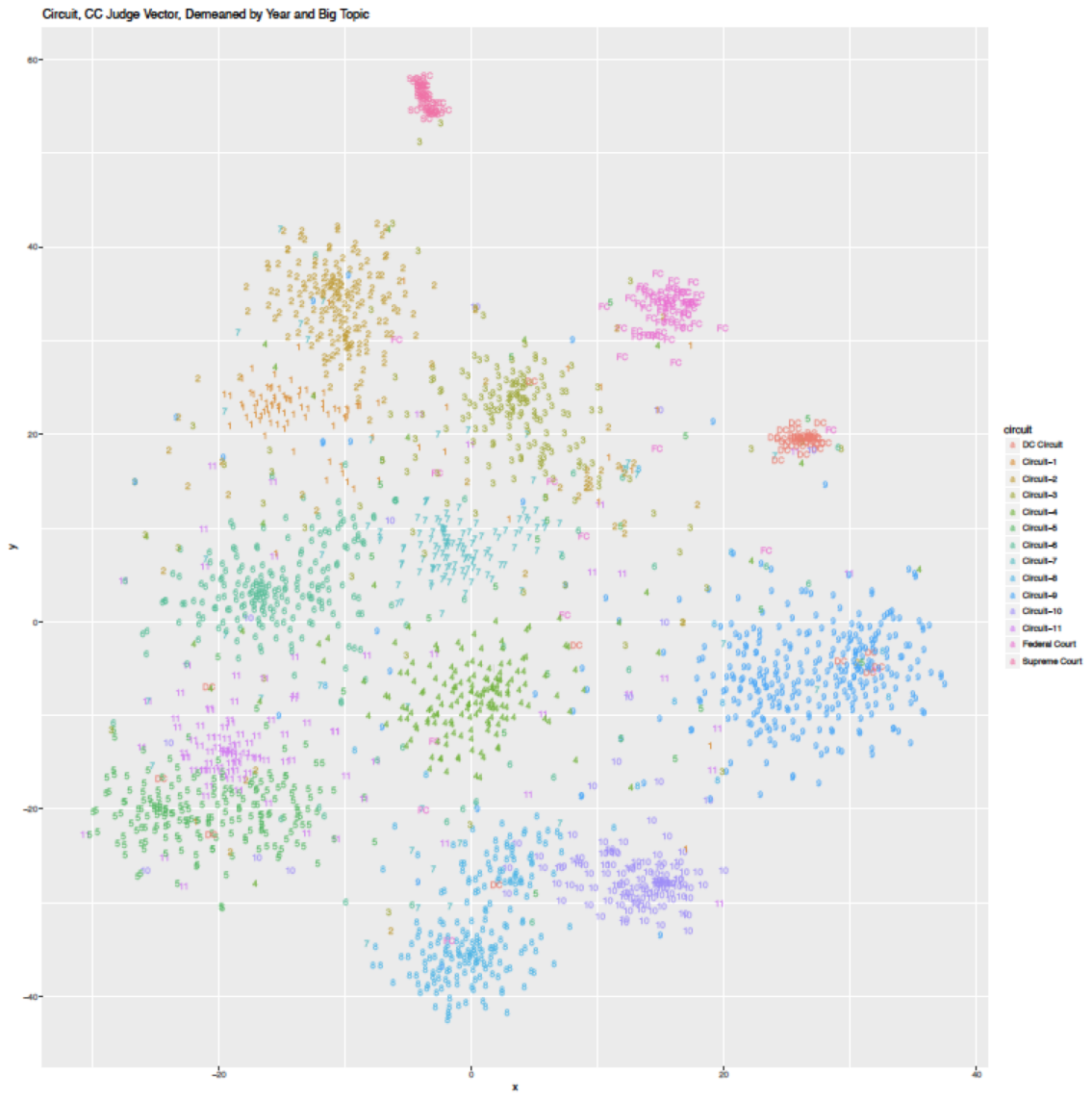
- isolating directions for legal and political concepts
    - liberal vs. conservative, procedural vs. substantive, originalists vs. pragmatists, or economic analysis

Objective

- Predict a given word using its context *and* N-gram representation of whole document

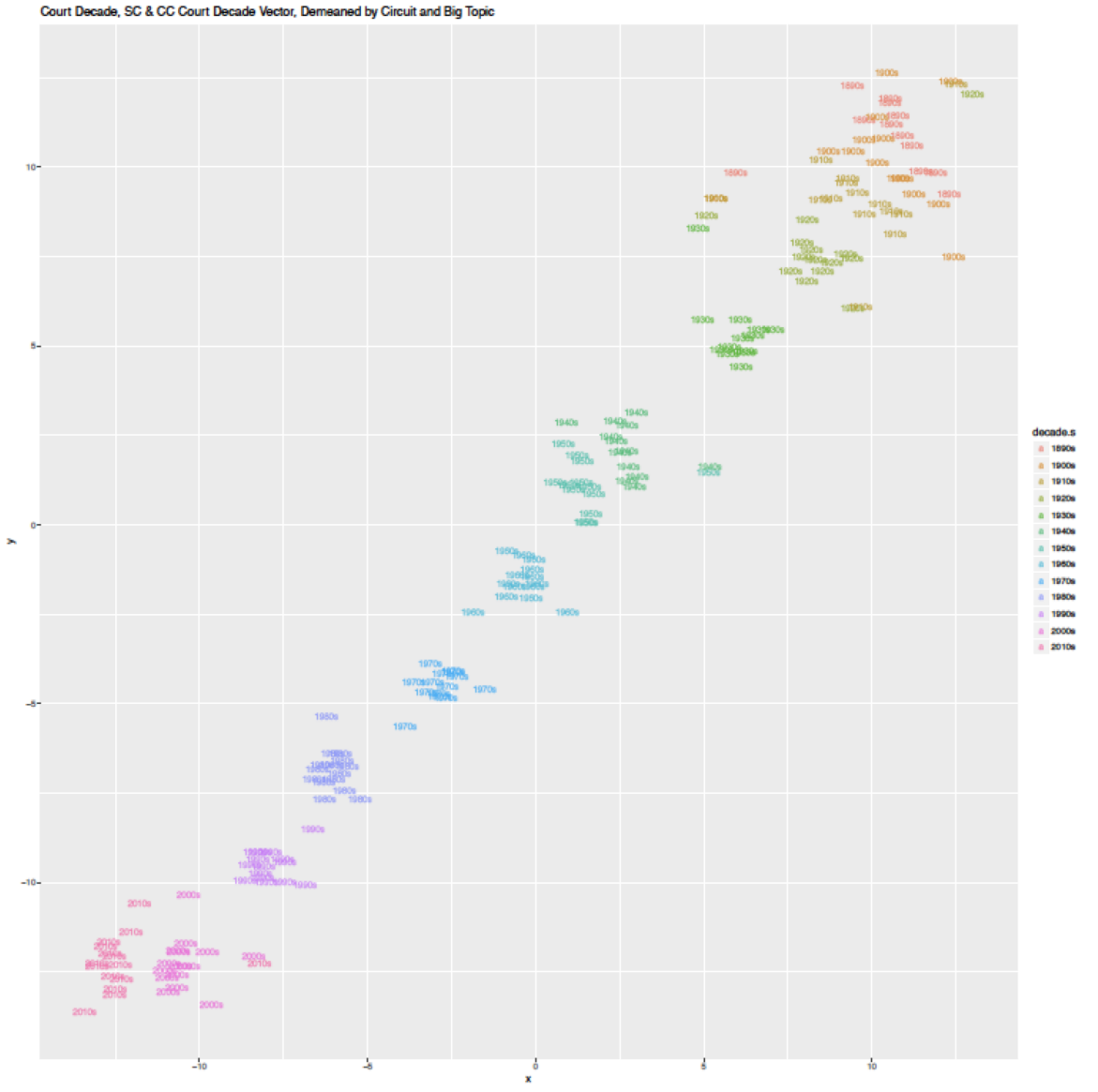
Visual Structure of Case Vectors by Circuit

Figure 1: Centered by Topic-Year, Averaged by Judge, Labeled by Court



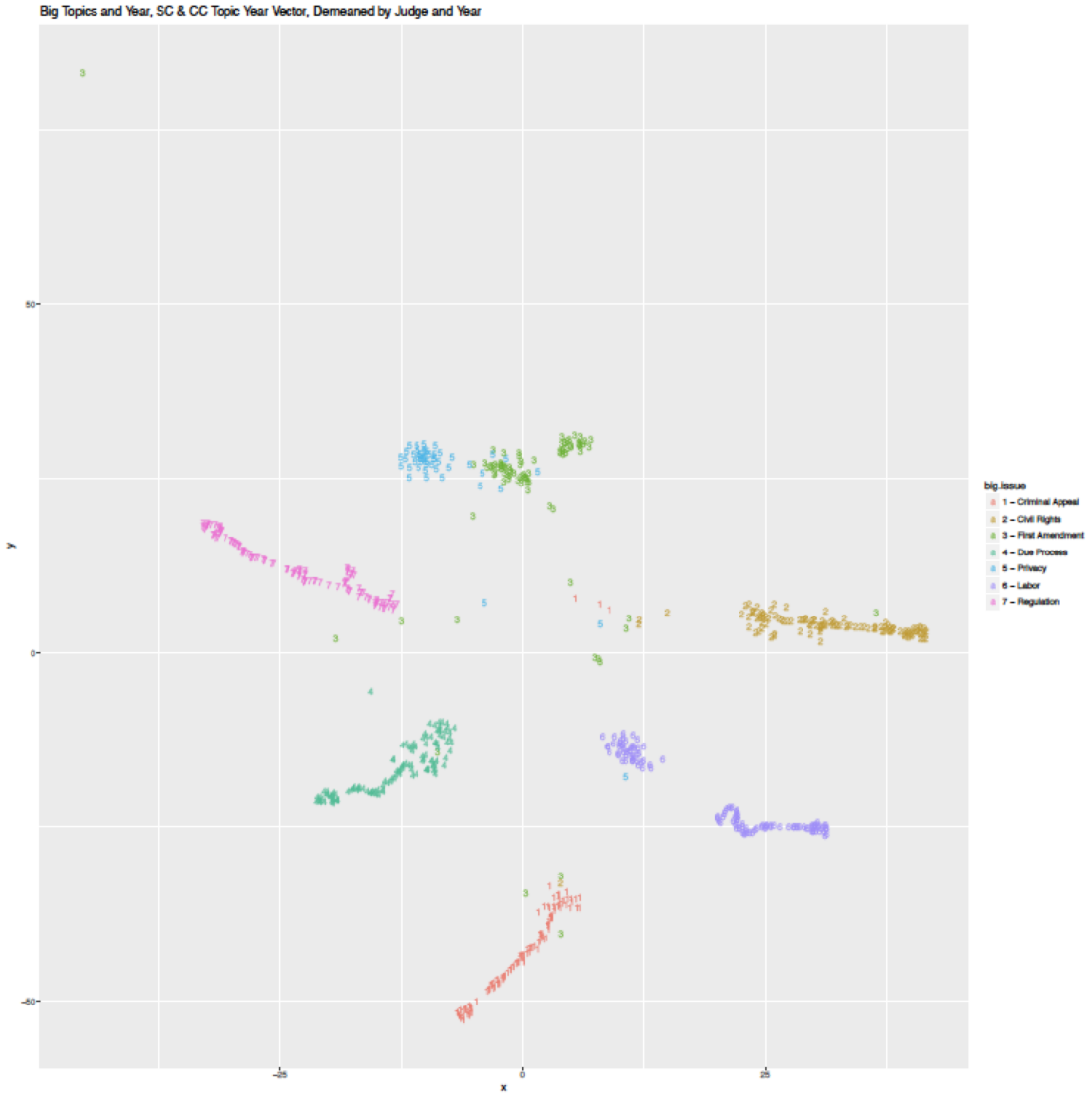
Case Vectors by Decade

Figure 2: Centered by Court-Topic, Averaged by Court-Year, Labeled by Decade



Case Vectors by Topic

Figure 3: Centered by Judge-Year, Averaged by Topic-Year, Labeled by Topic

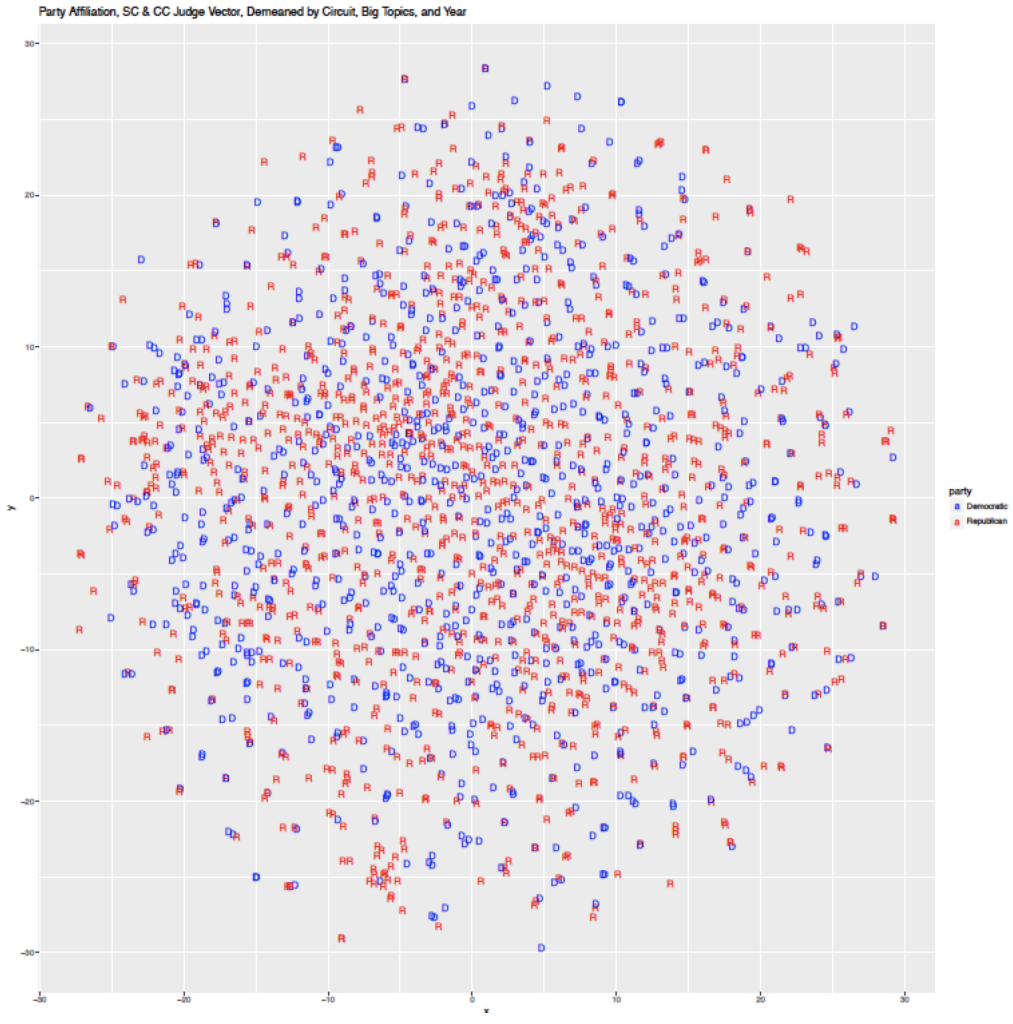


Case Vectors by Birth Cohort

Figure 5: Centered by Court-Topic-Year, Averaged by Judge, Labeled by Judge Birth Cohort

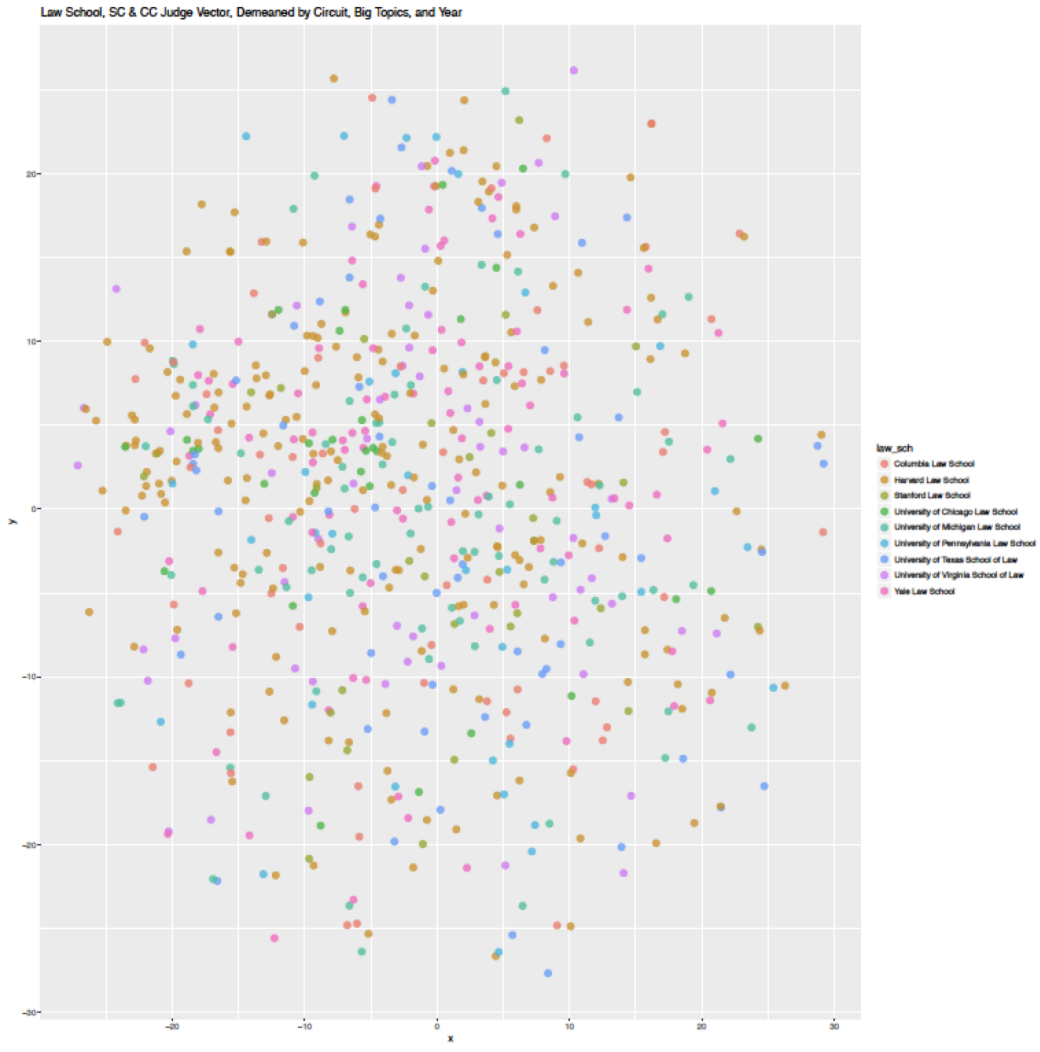
Case Vectors by Party

Figure 4: Centered by Court-Topic-Year, Averaged by Judge, Labeled by Political Party



Case Vectors by Law School

Figure 6: Centered by Court-Topic-Year, Averaged by Judge, Labeled by Law School Attended



Relatedness between judges

Circuit Judge Name	Similarity	Rank	Circuit Judge Name	Similarity	Rank
POSNER, RICHARD A.	1.000	1	TONE, PHILIP W.	0.459	16
EASTERBROOK, FRANK H.	0.663	2	SIBLEY, SAMUEL	0.459	17
SUTTON, JEFFREY S.	0.620	3	SCALIA, ANTONIN	0.456	18
NOONAN, JOHN T.	0.596	4	COLLOTON, STEVEN M.	0.445	19
NELSON, DAVID A.	0.592	5	DUNIWAY, BENJAMIN	0.438	20
CARNES, EDWARD E.	0.567	6	GIBBONS, JOHN J.	0.422	21
FRIENDLY, HENRY	0.566	7	BOGGS, DANNY J.	0.420	22
KOZINSKI, ALEX	0.563	8	*BREYER, STEPHEN G.*	0.414	23
GORSUCH, NEIL M.	0.559	9	GOODRICH, HERBERT	0.412	24
CHAMBERS, RICHARD H.	0.546	10	LOKEN, JAMES B.	0.410	25
FERNANDEZ, FERDINAND F.	0.503	11	WEIS, JOSEPH F.	0.408	26
EDMONDSON, JAMES L.	0.501	12	SCALIA, ANTONIN (SCOTUS)	0.406	27
KLEINFELD, ANDREW J.	0.491	13	BOUDIN, MICHAEL	0.403	28
WILLIAMS, STEPHEN F.	0.481	14	RANDOLPH, A. RAYMOND	0.397	29
KETHLEDGE, RAYMOND M.	0.459	15	MCCONNELL, MICHAEL W.	0.390	30

Document vectors demeaned by court, year, and topic, then aggregated by judge.

Contrast with N-Gram Approach

Law-Econ Style = Cosine distance to JSTOR JEL K

Rank	Judge	Law-Econ Style
1	CARDAMONE, RI.	2.85
2	DUNCAN, ALLYSON	2.69
3	MILLER, JUSTIN	2.57
4	SMITH, EDWARD	2.55
5	**GARLAND, MERRICK**	**2.33**
6	**WHITE, BYRON**	**2.25**
7	GARTH, LEONARD I	2.21
8	WOODROUGH, J.	2.13
9	O'SULLIVAN, CLIFF	2.00
10	ROBB, ROGER	1.78
11	PREGERSON, HARRY	1.77
12	STALEY, AUSTIN	1.64
13	HENDERSON, A.	1.50
14	MOTZ, DIANA	1.45
15	BIGGS, JOHN JR.	1.44
16	CLARK, CHARLES	1.44
17	REED, STANLEY	1.42
18	JACKSON, HOWELL	1.41
19	SIMONS, CHARLES	1.40
20	MILLER, SHACKEL.	1.38
21	WOODBURY, PETER	1.38
22	JONES, JOHN	1.27
23	HICKS, XENOPHON	1.25
24	SUHRHEINRICH, R.	1.24
25	**POSNER, RICHARD**	**1.23**
	GORSUCH, NEIL M.	**-0.84**
	SOTOMAYOR, SONIA	**-1.02**
	SCALIA, ANTONIN	**-1.28**

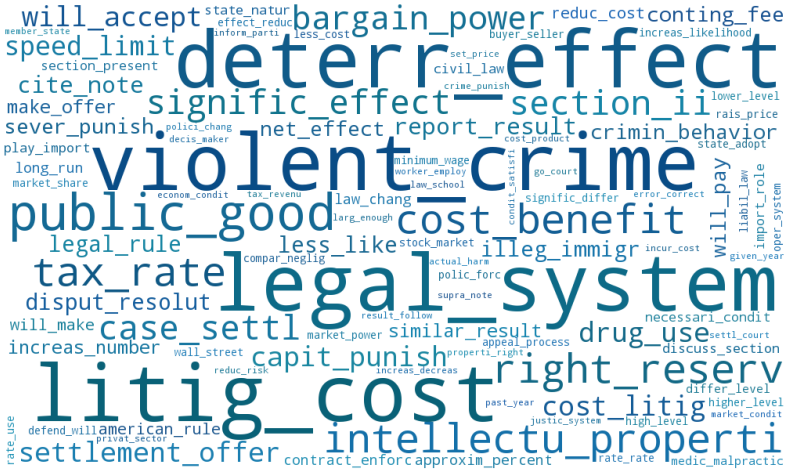
Contrast with N-Gram Approach

Law-Econ Style = Cosine distance to JSTOR JEL K

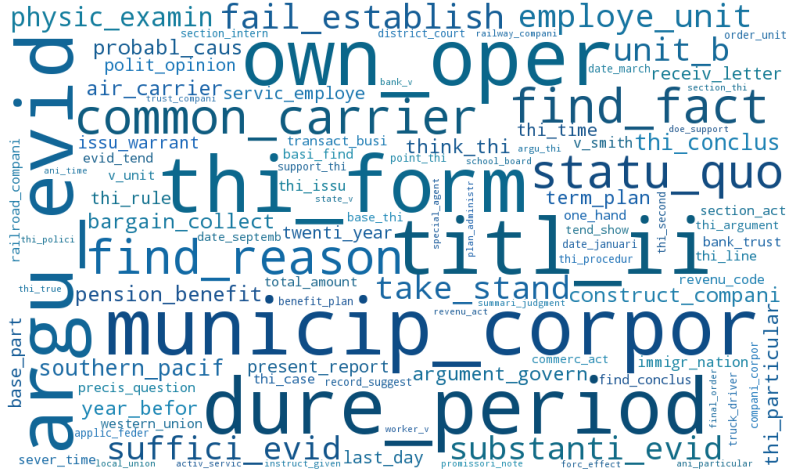
Rank	Judge	Law-Econ Style
1	CARDAMONE, RI.	2.85
2	DUNCAN, ALLYSON	2.69
3	MILLER, JUSTIN	2.57
4	SMITH, EDWARD	2.55
5	**GARLAND, MERRICK**	**2.33**
6	**WHITE, BYRON**	**2.25**
7	GARTH, LEONARD I	2.21
8	WOODROUGH, J.	2.13
9	O'SULLIVAN, CLIFF	2.00
10	ROBB, ROGER	1.78
11	PREGERSON, HARRY	1.77
12	STALEY, AUSTIN	1.64
13	HENDERSON, A.	1.50
14	MOTZ, DIANA	1.45
15	BIGGS, JOHN JR.	1.44
16	CLARK, CHARLES	1.44
17	REED, STANLEY	1.42
18	JACKSON, HOWELL	1.41
19	SIMONS, CHARLES	1.40
20	MILLER, SHACKEL.	1.38
21	WOODBURY, PETER	1.38
22	JONES, JOHN	1.27
23	HICKS, XENOPHON	1.25
24	SUHRHEINRICH, R.	1.24
25	**POSNER, RICHARD**	**1.23**
	GORSUCH, NEIL M.	**-0.84**
	SOTOMAYOR, SONIA	**-1.02**
	SCALIA, ANTONIN	**-1.28**

Law-and-Economics Language (N-gram)

- All JSTOR economics articles (1960-) JEL K (1990-) JLE (1960-)
  - 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

("Ideas Have Consequences: The Impact of Law and Economics on American Justice", Ash, Chen, Naidu)

Law-and-Economics Vectors

- externalit*, transaction_costs, efficien*, deterr*, cost_benefit, capital, game_theo, chicago_school, marketplace, law1economic, law2economic identified by Ellickson (2000)

- One of the sentences that is closest to “economics” in is: “The discussion then turned to economics.”

Law-and-Economics Vectors

- externalit*, transaction_costs, efficien*, deterr*, cost_benefit, capital, game_theo, chicago_school, marketplace, law1economic, law2economic identified by Ellickson (2000)

- One of the sentences that is closest to “economics” in is: “The discussion then turned to economics.”

Implicit (or Explicit) Attitudes

- Google translate
  - “he/she is a doctor”(turkish) -> “he is a doctor” (english)
  - “he/she is a nurse”(turkish) -> “she is a nurse” (english)
- The text of the opinions provide a window into rich representations of legal/political institutions, as we well as **human social psychology**.
- We ask whether gender and racial bias varies across judges.

Implicit (or Explicit) Attitudes

- Google translate
  - "he/she is a doctor"(turkish) -> "he is a doctor" (english)
  - "he/she is a nurse"(turkish) -> "she is a nurse" (english)
- The text of the opinions provide a window into rich representations of legal/political institutions, as we well as **human social psychology**.

- We ask whether gender and racial bias varies across judges.

Implicit (or Explicit) Attitudes

- Google translate
  - "he/she is a doctor"(turkish) -> "he is a doctor" (english)
  - "he/she is a nurse"(turkish) -> "she is a nurse" (english)
- The text of the opinions provide a window into rich representations of legal/political institutions, as we well as **human social psychology**.

- We ask whether gender and racial bias varies across judges.

Word Embedding Association Test (Science 2017)

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.

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

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

- Train Word2Vec separately by judge, following Caliskan et. al (2017).

("Measuring the Consequences of Implicit Bias Using Semantics Derived Automatically from Judicial Corpora", Ash, Chen, Ornaghi 2018)

Word Embedding Association Test (Science 2017)

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.

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

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

- Train Word2Vec separately by judge, following Caliskan et. al (2017).

("Measuring the Consequences of Implicit Bias Using Semantics Derived Automatically from Judicial Corpora", Ash, Chen, Ornaghi 2018)

Word Embedding Association Test (Science 2017)

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.

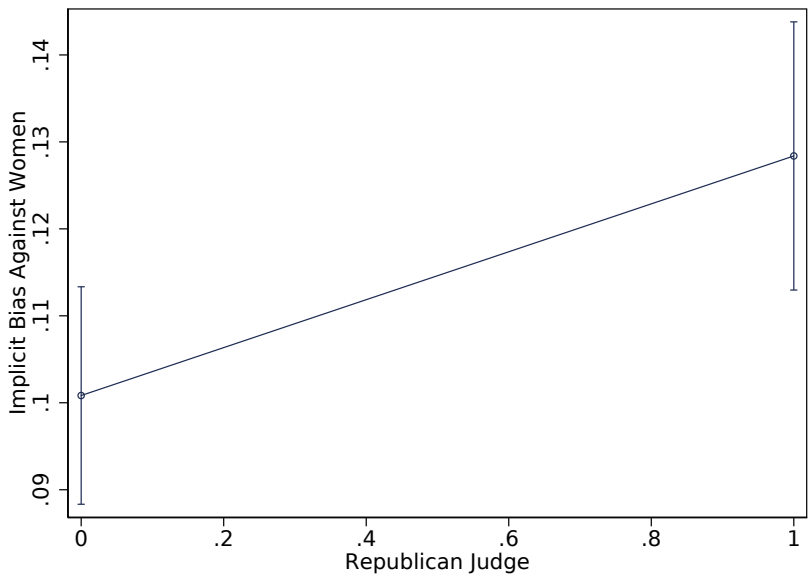
$$\text{Implicit Sexism} = \frac{\text{Male-Pleasant Association}}{\text{Male-Unpleasant Association}} - \frac{\text{Female-Pleasant Association}}{\text{Female-Unleasant Association}}$$

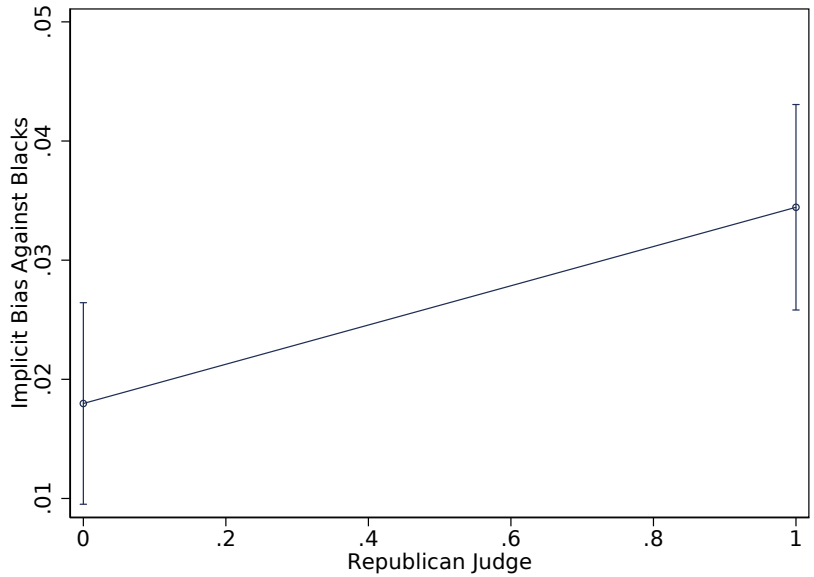
$$\text{Implicit Racism} = \frac{\text{White-Pleasant Association}}{\text{White-Unpleasant Association}} - \frac{\text{Black-Pleasant Association}}{\text{Black-Unleasant Association}}$$

- Train Word2Vec separately by judge, following Caliskan et. al (2017).

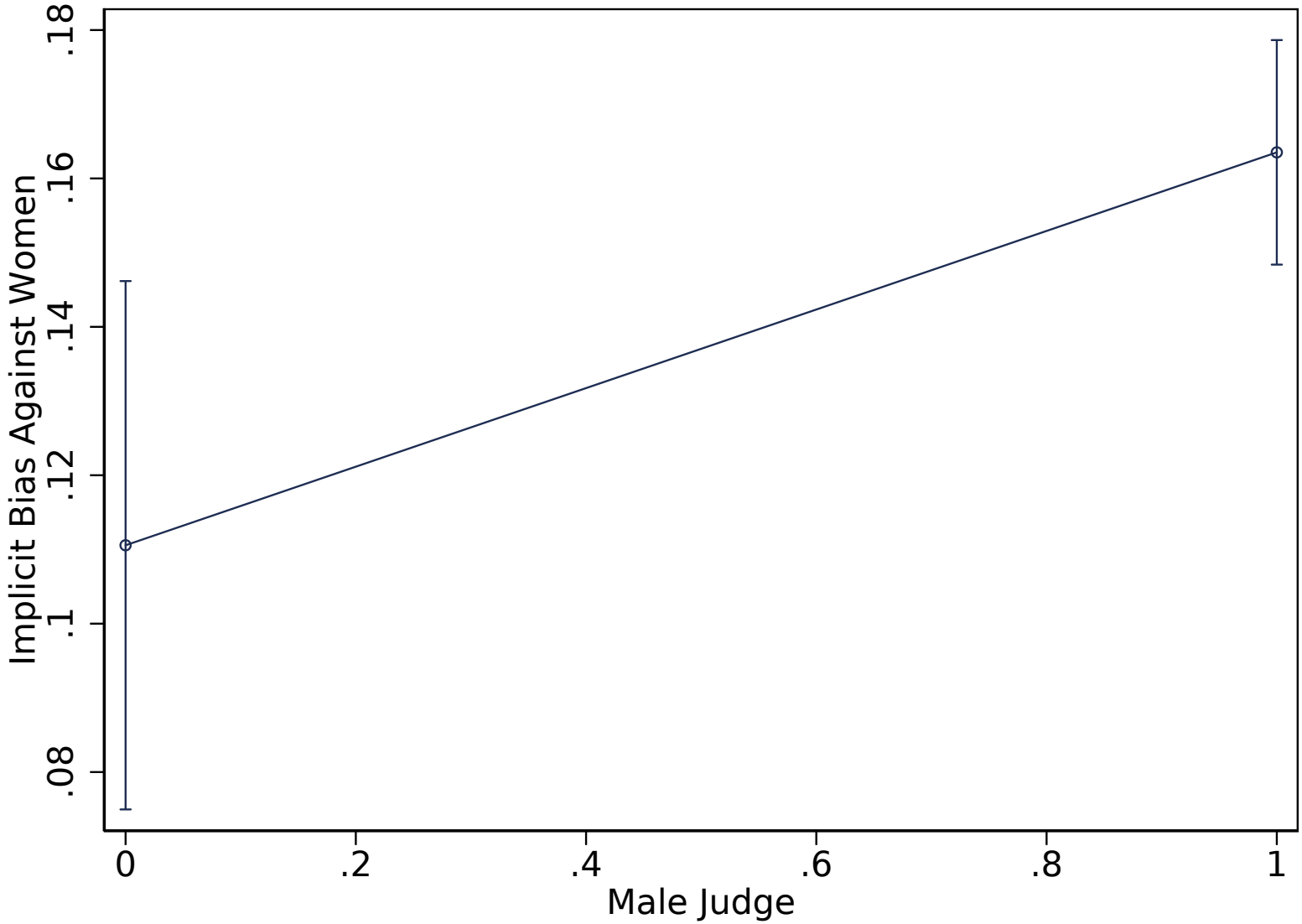
("Measuring the Consequences of Implicit Bias Using Semantics Derived Automatically from Judicial Corpora", Ash, Chen, Ornaghi 2018)

Republican judges have higher gender bias and race bias

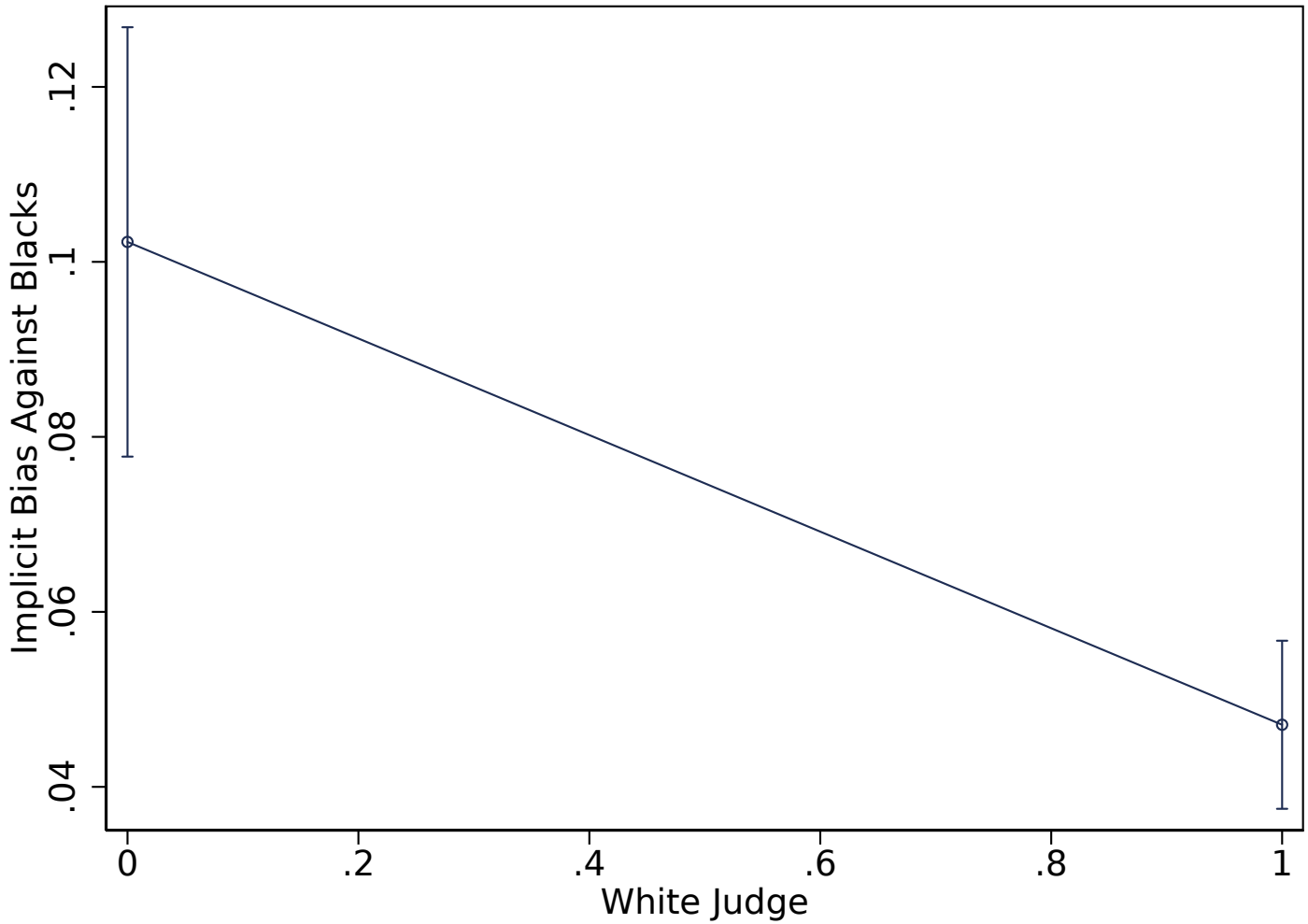




Male judges have higher gender bias than female judges



White judges have *lower* race bias than black judges



Both the words and the IAT work at an unconscious level, in contrast to the decisions which are more conscious.

Trump nominees have high race and gender 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

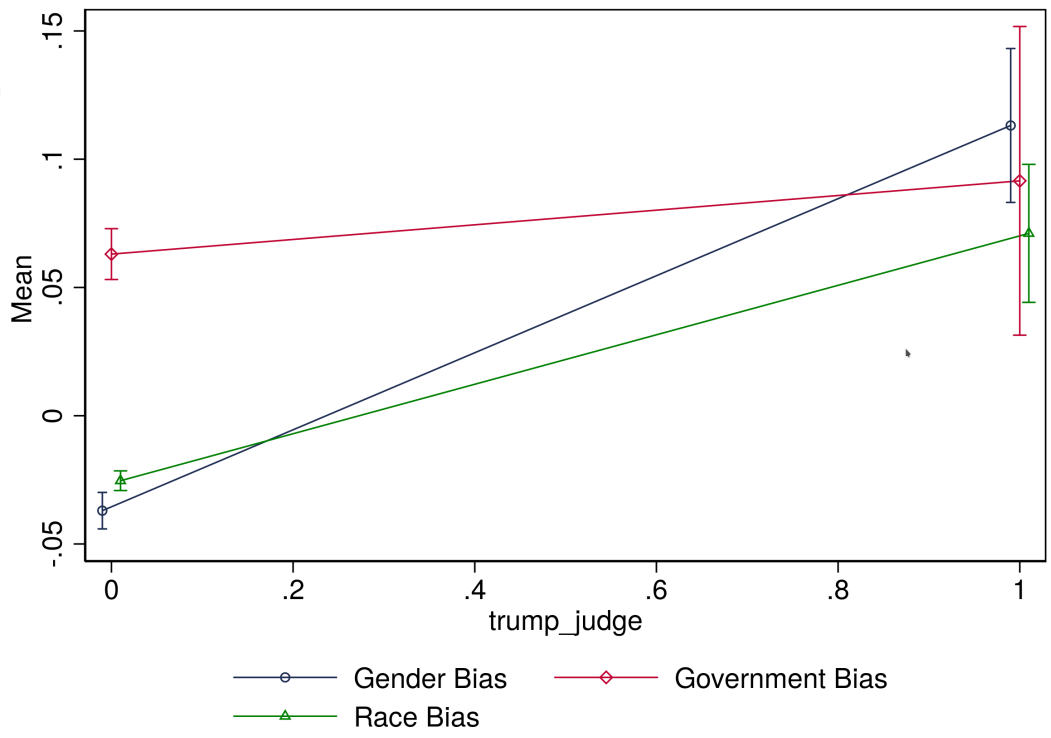
Amul Thapar of Kentucky, U.S. Court of Appeals for the Sixth Circuit

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



Caveats: Predictions Work on Easy Cases

- Katz et al. (2014) doesn't predict close cases well (so do error analysis)
  - Also 70% accuracy $\approx$ 70% reversal in Supreme Court
- Voice masculinity basically orthogonal to random forest prediction

	Judge Votes for Lawyer					
Predicted Vote	0.257***		0.258***	0.250***		0.248***
from Random Forest	(0.0486)		(0.0487)	(0.0485)		(0.0489)
Masculine		-0.0223**	-0.0207**		-0.0852**	-0.0780**
		(0.0101)	(0.0101)		(0.0359)	(0.0361)
Cluster			Lawyer and Judge			
Collapsed	No	No	No	Yes	Yes	Yes
Observations	26447	26391	26391	1229	1229	1229
R-squared	0.061	0.002	0.063	0.058	0.008	0.064
Sample: Male Petitioners, Democrat Judges						

Figure 1: Best Prediction and Perceived Masculinity

- Additional 3-10% of variance explained with voice masculinity
- Random forest also selects masculinity, improves accuracy by 2%

("Covering: Mutable Characteristics and Perceptions of Voice in the U.S. Supreme Court", Chen, Halberstam, Yu)

Caveats: Predictions Work on Easy Cases

- Katz et al. (2014) doesn't predict close cases well (so do error analysis)
  - Also 70% accuracy $\approx$ 70% reversal in Supreme Court
- Voice masculinity basically orthogonal to random forest prediction

	Judge Votes for Lawyer					
Predicted Vote	0.257***		0.258***	0.250***		0.248***
from Random Forest	(0.0486)		(0.0487)	(0.0485)		(0.0489)
Masculine		-0.0223**	-0.0207**		-0.0852**	-0.0780**
		(0.0101)	(0.0101)		(0.0359)	(0.0361)
Cluster			Lawyer and Judge			
Collapsed	No	No	No	Yes	Yes	Yes
Observations	26447	26391	26391	1229	1229	1229
R-squared	0.061	0.002	0.063	0.058	0.008	0.064
Sample: Male Petitioners, Democrat Judges						

Figure 1: Best Prediction and Perceived Masculinity

- Additional 3-10% of variance explained with voice masculinity
- Random forest also selects masculinity, improves accuracy by 2%

("Covering: Mutable Characteristics and Perceptions of Voice in the U.S. Supreme Court", Chen, Halberstam, Yu)

Caveats: Predictions Work on Easy Cases

- Katz et al. (2014) doesn't predict close cases well (so do error analysis)
  - Also 70% accuracy $\approx$ 70% reversal in Supreme Court
- Voice masculinity basically orthogonal to random forest prediction

	Judge Votes for Lawyer					
Predicted Vote	0.257***		0.258***	0.250***		0.248***
from Random Forest	(0.0486)		(0.0487)	(0.0485)		(0.0489)
Masculine		-0.0223**	-0.0207**		-0.0852**	-0.0780**
		(0.0101)	(0.0101)		(0.0359)	(0.0361)
Cluster			Lawyer and Judge			
Collapsed	No	No	No	Yes	Yes	Yes
Observations	26447	26391	26391	1229	1229	1229
R-squared	0.061	0.002	0.063	0.058	0.008	0.064
Sample: Male Petitioners, Democrat Judges						

Figure 1: Best Prediction and Perceived Masculinity

- Additional 3-10% of variance explained with voice masculinity
- Random forest also selects masculinity, improves accuracy by 2%

("Covering: Mutable Characteristics and Perceptions of Voice in the U.S. Supreme Court", Chen, Halberstam, Yu)

Caveats: Predictions Work on Easy Cases

- Katz et al. (2014) doesn't predict close cases well (so do error analysis)
  - Also 70% accuracy $\approx$ 70% reversal in Supreme Court
- Voice masculinity basically orthogonal to random forest prediction

	Judge Votes for Lawyer					
Predicted Vote	0.257***		0.258***	0.250***		0.248***
from Random Forest	(0.0486)		(0.0487)	(0.0485)		(0.0489)
Masculine		-0.0223**	-0.0207**		-0.0852**	-0.0780**
		(0.0101)	(0.0101)		(0.0359)	(0.0361)
Cluster			Lawyer and Judge			
Collapsed	No	No	No	Yes	Yes	Yes
Observations	26447	26391	26391	1229	1229	1229
R-squared	0.061	0.002	0.063	0.058	0.008	0.064
Sample: Male Petitioners, Democrat Judges						

Figure 1: Best Prediction and Perceived Masculinity

- Additional 3-10% of variance explained with voice masculinity
- Random forest also selects masculinity, improves accuracy by 2%

("Covering: Mutable Characteristics and Perceptions of Voice in the U.S. Supreme Court", Chen, Halberstam, Yu)

Summary

N-grams to predict decisions

- Average accuracy of 75% in predicting violations (on average)
- Average accuracy of 70% in predicting future violations

But higher accuracy achieved by guessing "violation"

article	violation: **80%**	non-violation: **20%**

What is the social science or policy application?

- Predicting reversals
  - should a lawyer appeal? should a judge revise?
- Predicting sentencing harshness (and disparities)
  - can we score nominees prior to appointment? after appointment?
- Predicting rearrest
  - can algorithms increase efficiency *and* equity?
- Early predictability
  - can we score "revealed preference" indifference?
- Predicting ideology
  - does phonology matter?
- Personalized nudges of judges
  - rather than checklists, to increase justice? and legitimacy of law?

Summary

N-grams to predict decisions

- Average accuracy of 75% in predicting violations (on average)
- Average accuracy of 70% in predicting future violations

But higher accuracy achieved by guessing "violation"

article	violation: **80%**	non-violation: **20%**

What is the social science or policy application?

- Predicting reversals
  - should a lawyer appeal? should a judge revise?
- Predicting sentencing harshness (and disparities)
  - can we score nominees prior to appointment? after appointment?
- Predicting rearrest
  - can algorithms increase efficiency *and* equity?
- Early predictability
  - can we score "revealed preference" indifference?
- Predicting ideology
  - does phonology matter?
- Personalized nudges of judges
  - rather than checklists, to increase justice? and legitimacy of law?

Summary

N-grams to predict decisions

- Average accuracy of 75% in predicting violations (on average)
- Average accuracy of 70% in predicting future violations

But higher accuracy achieved by guessing "violation"

article	violation: **80%**	non-violation: **20%**

What is the social science or policy application?

- Predicting reversals
  - should a lawyer appeal? should a judge revise?
- Predicting sentencing harshness (and disparities)
  - can we score nominees prior to appointment? after appointment?
- Predicting rearrest
  - can algorithms increase efficiency *and* equity?
- Early predictability
  - can we score "revealed preference" indifference?
- Predicting ideology
  - does phonology matter?
- Personalized nudges of judges
  - rather than checklists, to increase justice? and legitimacy of law?

Summary

N-grams to predict decisions

- Average accuracy of 75% in predicting violations (on average)
- Average accuracy of 70% in predicting future violations

But higher accuracy achieved by guessing "violation"

article	violation: **80%**	non-violation: **20%**

What is the social science or policy application?

- Predicting reversals
  - should a lawyer appeal? should a judge revise?
- Predicting sentencing harshness (and disparities)
  - can we score nominees prior to appointment? after appointment?
- Predicting rearrest
  - can algorithms increase efficiency *and* equity?
- Early predictability
  - can we score "revealed preference" indifference?
- Predicting ideology
  - does phonology matter?
- Personalized nudges of judges
  - rather than checklists, to increase justice? and legitimacy of law?

Summary

N-grams to predict decisions

- Average accuracy of 75% in predicting violations (on average)
- Average accuracy of 70% in predicting future violations

But higher accuracy achieved by guessing "violation"

article	violation: **80%**	non-violation: **20%**

What is the social science or policy application?

- Predicting reversals
  - should a lawyer appeal? should a judge revise?
- Predicting sentencing harshness (and disparities)
  - can we score nominees prior to appointment? after appointment?
- Predicting rearrest
  - can algorithms increase efficiency *and* equity?
- Early predictability
  - can we score "revealed preference" indifference?
- Predicting ideology
  - does phonology matter?
- Personalized nudges of judges
  - rather than checklists, to increase justice? and legitimacy of law?

Summary

N-grams to predict decisions

- Average accuracy of 75% in predicting violations (on average)
- Average accuracy of 70% in predicting future violations

But higher accuracy achieved by guessing "violation"

article	violation: **80%**	non-violation: **20%**

What is the social science or policy application?

- Predicting reversals
  - should a lawyer appeal? should a judge revise?
- Predicting sentencing harshness (and disparities)
  - can we score nominees prior to appointment? after appointment?
- Predicting rearrest
  - can algorithms increase efficiency *and* equity?
- Early predictability
  - can we score "revealed preference" indifference?
- Predicting ideology
  - does phonology matter?
- Personalized nudges of judges
  - rather than checklists, to increase justice? and legitimacy of law?

Summary

N-grams to predict decisions

- Average accuracy of 75% in predicting violations (on average)
- Average accuracy of 70% in predicting future violations

But higher accuracy achieved by guessing "violation"

article	violation: **80%**	non-violation: **20%**

What is the social science or policy application?

- Predicting reversals
  - should a lawyer appeal? should a judge revise?
- Predicting sentencing harshness (and disparities)
  - can we score nominees prior to appointment? after appointment?
- Predicting rearrest
  - can algorithms increase efficiency *and* equity?
- Early predictability
  - can we score "revealed preference" indifference?
- Predicting ideology
  - does phonology matter?
- Personalized nudges of judges
  - rather than checklists, to increase justice? and legitimacy of law?

Summary

N-grams to predict decisions

- Average accuracy of 75% in predicting violations (on average)
- Average accuracy of 70% in predicting future violations

But higher accuracy achieved by guessing "violation"

article	violation: **80%**	non-violation: **20%**

What is the social science or policy application?

- Predicting reversals
  - should a lawyer appeal? should a judge revise?
- Predicting sentencing harshness (and disparities)
  - can we score nominees prior to appointment? after appointment?
- Predicting rearrest
  - can algorithms increase efficiency *and* equity?
- Early predictability
  - can we score "revealed preference" indifference?
- Predicting ideology
  - does phonology matter?
- Personalized nudges of judges
  - rather than checklists, to increase justice? and legitimacy of law?

Data Science Justice Collaboratory (Toulouse Institute for Advanced Study, NYU Courant Institute of Mathematics, ETH Zurich, Harvard Medical School)