CS 124/LINGUIST 180
From Languages to Information

WINTER 2024

TWO IMPORTANT APPLICATIONS OF NLP:

NLP FOR PUBLIC GOOD

AND FOR COMPUTATIONAL SOCIAL SCIENCE

NLP for Public Good and Computational Social Science

1. NLP for Public Good: Analyzing Police Body-worn Camera Conversations

Can we improve police-community relations?

2. NLP for Computational Social Science: Measuring US political discourse about immigration

Can we learn about polarization, develop new ways to understand and measure toxic speech like dehumanization?

1. NLP + Social Psychology for Improving Police-Community Relations

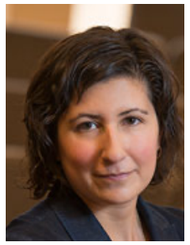
Rob Voigt

Nick Camp

Eugenia Rho

Vinod Prabhakaran

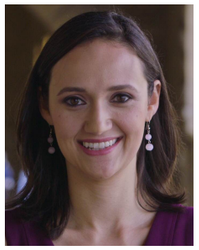
Camilla Griffiths



MarYam Hamedani

Will Hamilton

Maggie Harrington

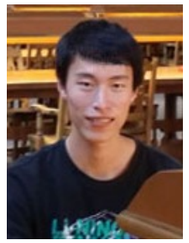


Rebecca Hetey

David Jurgens

Nelson Morgan

Reid Pryzant



Hang Su

Prateek Verma

Yuyang Zhong

PI: Prof. Jennifer Eberhardt

Problems in Police-Community relations in the United States

Black Americans have long reported more unfair stops, negative interactions with the police (Epp et al. 2014, Pew Survey 2016)

Inappropriate police officer use of force, widely captured on viral videos

George Floyd killing

Can Natural Language Processing help?

Measure problems in police-community interactions?

Detect potential for escalation and violence?

Help design interventions?

Our idea:
Use body-camera footage as data

Data first from one police department (now + other departments)

Look at common, everyday interactions

Procedural Justice

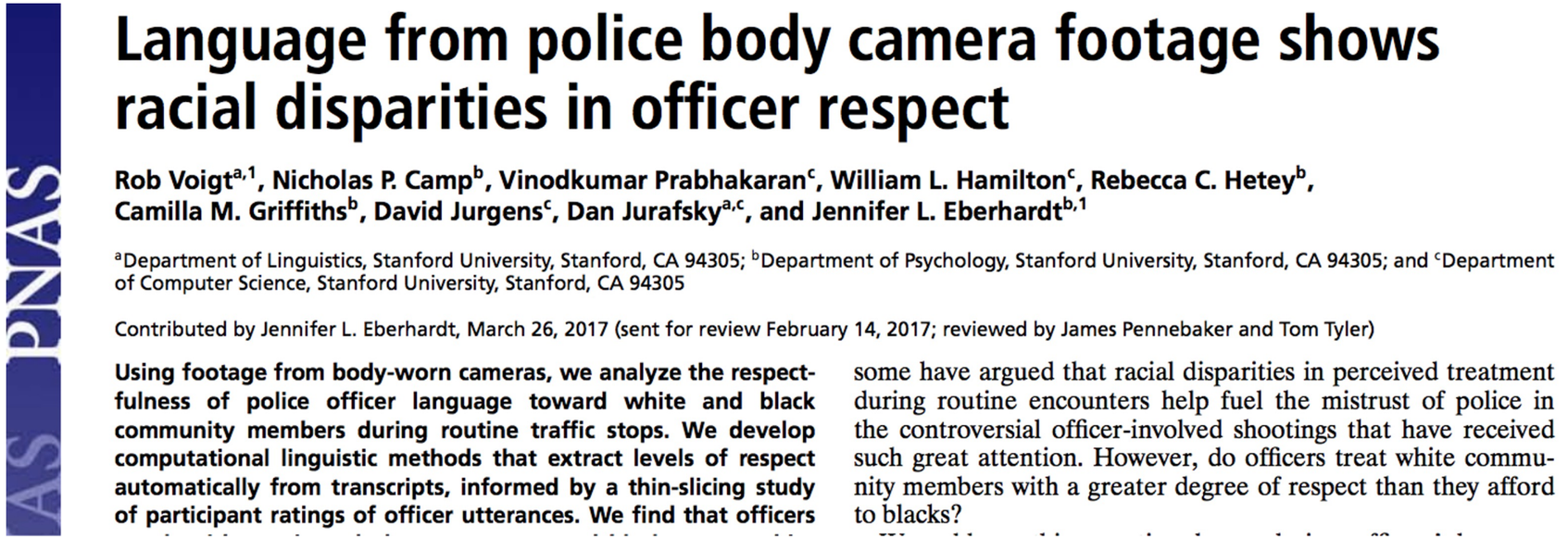
The idea that the justice system should strive for not only *equitable outcomes*, but also an ***equitable process*** (including interpersonal treatment)

Respect

A person who is treated with respect has more trust in the fairness of the officer and the institution

(Tyler, 1990; Tyler & Ho, 2001; Tyler & Sunshine, 2003 Mazerolle et al., 2013)

Study 1: Do police officers treat black community members with a different degree of respect than white?



PNAS

Language from police body camera footage shows racial disparities in officer respect

Rob Voigt[a,1], Nicholas P. Camp[b], Vinodkumar Prabhakaran[c], William L. Hamilton[c], Rebecca C. Hetey[b], Camilla M. Griffiths[b], David Jurgens[c], Dan Jurafsky[a,c], and Jennifer L. Eberhardt[b,1]

[a]Department of Linguistics, Stanford University, Stanford, CA 94305; [b]Department of Psychology, Stanford University, Stanford, CA 94305; and [c]Department of Computer Science, Stanford University, Stanford, CA 94305

Contributed by Jennifer L. Eberhardt, March 26, 2017 (sent for review February 14, 2017; reviewed by James Pennebaker and Tom Tyler)

Using footage from body-worn cameras, we analyze the respectfulness of police officer language toward white and black community members during routine traffic stops. We develop computational linguistic methods that extract levels of respect automatically from transcripts, informed by a thin-slicing study of participant ratings of officer utterances. We find that officers

some have argued that racial disparities in perceived treatment during routine encounters help fuel the mistrust of police in the controversial officer-involved shootings that have received such great attention. However, do officers treat white community members with a greater degree of respect than they afford to blacks?

Body-Cameras as Research Data

About 1000 "everyday interactions" from one month

Vehicle stops with warning/citation; **no arrests**

245 different officers

Classifiers for linguistic properties of interaction

Trained on human labels

Humans labeled sentences for respect levels

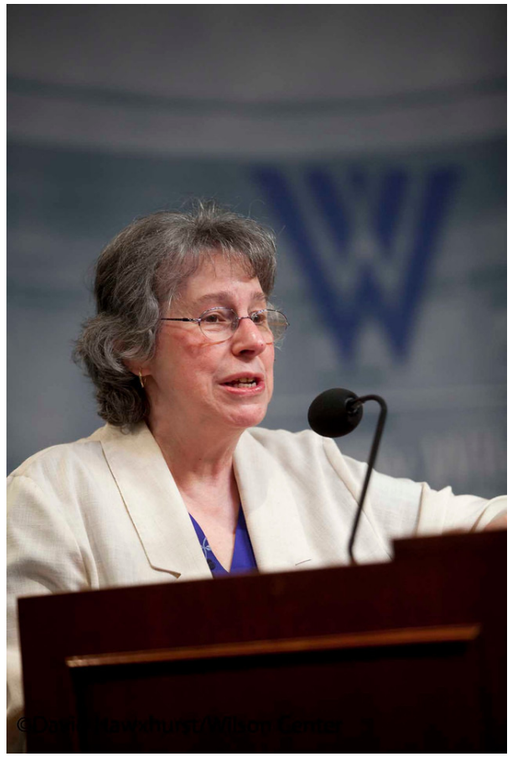
Humans labeled sentences for dialogue structure

1. Classifier labels sentences for *respect*
2. Classifier labels sentences for *dialogue structure*

Face in modeling politeness and respect

[Erving Goffman 1967; Lakoff 1973; Brown and Levinson, 1978]

Erving Goffman



Robin Lakoff

Penelope Brown
Stephen Levinson

"Face" in modeling respect

[Erving Goffman 1967; Lakoff 1973; Brown and Levinson, 1978]

POSITIVE POLITENESS
(HEARER'S SELF-IMAGE)

Emphasize your value

Emphasize my good relationship with you

NEGATIVE POLITENESS
(HEARER'S FREEDOM OF ACTION)

Minimize my request

Put the imposition on record

Cues for Positive Politeness

Formal titles

“ma'am”, “sir”, “Mr."

Introductions

“Hello”, “My name is”, “I’m Officer X”

Sympathy or concern: mentioning safety

“Drive safely”, "Be safe now"

Cues for Negative Politeness

Apologizing

"sorry", "oops", "my fault", "excuse me"

Gratitude

"thanks", "appreciate"

Imposition minimizers

"it's ok", "don't worry", "no big deal", "you're good"

Hedges

"just", "a little", "kind of", "sort of"

Prior applications of computational politeness!

Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. ACL 2013.

Community:

- Midwesterners are more polite
- Ruby programmers more polite than Python programmers

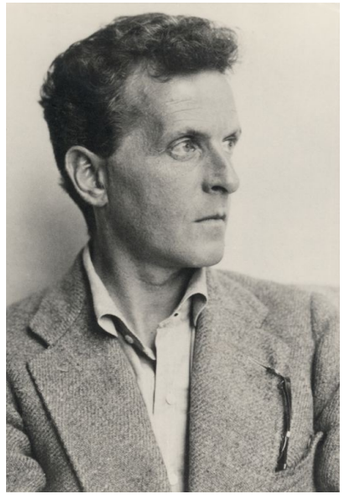
Gender:

- Women are more polite than men

Power and Status:

- Wikipedia editors get ruder after elected to admin

Dialogue: Two linguistic insights:



Each turn in a dialogue is a kind of action

Wittgenstein (1953) and Austin (1962)

Dialog structure mirrors task structure

Barbara Grosz (1977)

OFFICER:	Sir, hello, my name's Officer [NAME] of the Oakland Police Department.
MALE:	Hi.
OFFICER:	The reason why I pulled you over is when you passed me back there you were texting or talking on your cell phone.
MALE:	I was looking at a text, yes.
OFFICER:	Okay. Do you have um, what year is the car you're driving?
MALE:	It's a 2010.
OFFICER:	2010. And do you still live in [ADDRESS]?
MALE:	Yes.
OFFICER:	All right, sir. This is a citation for having your cell phone in your hand [...] It's not a moving violation. [...]You actually have two months ... to take care of the citation, okay? Please drive carefully.
MALE:	Okay.
OFFICER:	Thank you.

Greeting

Giving Reason

Asking Details

Issuing Sanction

Good Bye

Human labeled Dialog structure

Dialog structure has policy implications!

Procedural Justice:

People are more likely to obey the law when they believe authorities use **procedures** that are **just** and **fair**.

Departments require officers to give the driver the **reason** for the stop:

"The reason why I pulled you over is when you passed me back there you were texting or talking on your cell phone."

Could delaying these explanations lead to problematic or escalating encounters?

Is this your car? Do you live here?

Epp, Charles R., Steven Maynard-Moody, and Donald P. Haider-Markel. 2014. *Pulled over: How police stops define race and citizenship*. University of Chicago.

Black community members experience intrusive and investigatory questions, especially in certain neighborhoods.

Can we quantify these differences in who gets asked these kinds of questions?

1. Classifier for respect

APOLOGY INTRODUCTION LAST NAME Sorry to stop you. My name's Officer [name] with the Police Department.	0.84
FORMAL TITLE SAFETY PLEASE There you go, ma'am. Drive safe, please.	1.21

2. Classifiers for 23 Dialog Acts

Vinodkumar Prabhakaran, Camilla Griffiths, Hang Su, Prateek Verma, Nelson Morgan, Jennifer Eberhardt, and Dan Jurafsky. 2018. Detecting Institutional Dialog Acts in Police Traffic Stops. Transactions of the Association for Computational Linguistics 6: 467--481

Giving Reason (“The reason I stopped you is …”)

Asking for Documents (Insurance/License/ etc.)

Issuing Sanction (Citation/Warning/Fix-it Ticket)

Drive Safe (“Drive safely now”)

Offering Help (“Do you need help?”)

Inquiring Ownership (“Is this your car?”)

Mentioning Lenience (“I'll give you a break.”)

Is there an effect of race across all 36,738 utterances?

Yes.

Officers are more respectful to white drivers

Officers use more appropriate dialogs with white drivers

Some examples?

More positive politeness to white drivers: Formal titles

"All right, **sir**, take care."

"Okay, **ma'am**. Do you have your insurance and registration, **ma'am**?"

"All right **Mr. X**, listen. I'm going to let you, uh, go with a verbal warning tonight"

More positive politeness to white drivers: Concern for driver safety

"Okay. All right. **Drive safely.** All right?"

"So I'm just glad you're **safe**. You're cool. Right? It just take a little bit of, like, distraction to, to get someone hurt. You know? And **I just want you and your baby to be safe**."

More negative politeness to white drivers: Reassurance and Downplayers

"**No problem.** I understand. Just your license, please."

"Yeah**. Don't worry** about that. **It's all good.**

"**Just** have uh, anybody sign the back of, the back of that, to **just** uh, **just** prove that it's been taken care of."

Black drivers more likely to be asked legitimacy questions

"Is this your car, boss?"

"Does the car belong to you?"

"Why are you here?"

What about escalated stops?

PNAS RESEARCH ARTICLE | SOCIAL SCIENCES OI

Escalated police stops of Black men are linguistically and psychologically distinct in their earliest moments

Eugenia H. Rho, Maggie Harrington, Yuyang Zhong, Reid Pryzant, Nicholas P. Camp, Dan Jurafsky, Jennifer L. Eberhardt

Prof. Eugenia H. Rho
CS, Virginia Tech

Language in Escalated Stops

Many calls for police to de-escalate encounters

Killing of George Floyd led to the largest racial justice movement of the 21st century

These concerns about escalation arise even when no force is used

Language in Escalated Stops

- How does escalation unfold?
- What kinds of language characterize escalated stops?
  - handcuffing
  - searching
  - arrests
- These are normal traffic stops, no warrants or car chases
- At the very beginning of the stop
  - first 45 words/30 seconds

Which dialog acts are associated with start of escalated stops?

Officers in escalated stops in first 45 words:

- **Less likely to give a reason for the stop**
  - 15% versus 38%
- **More likely to start with an order**
  - 22% vs. 8%

How predictable is escalation from officer's first words?

Use large language model (DeBERTaV3) to predict escalated-or-not from first 45 words

71% accuracy

What features is it discovering? *Order* and *reason* dialogue acts!

"Escalated stops often begin in escalation"

Could all these results be artifact of some confounding variable?

No. We controlled in our regressions for:

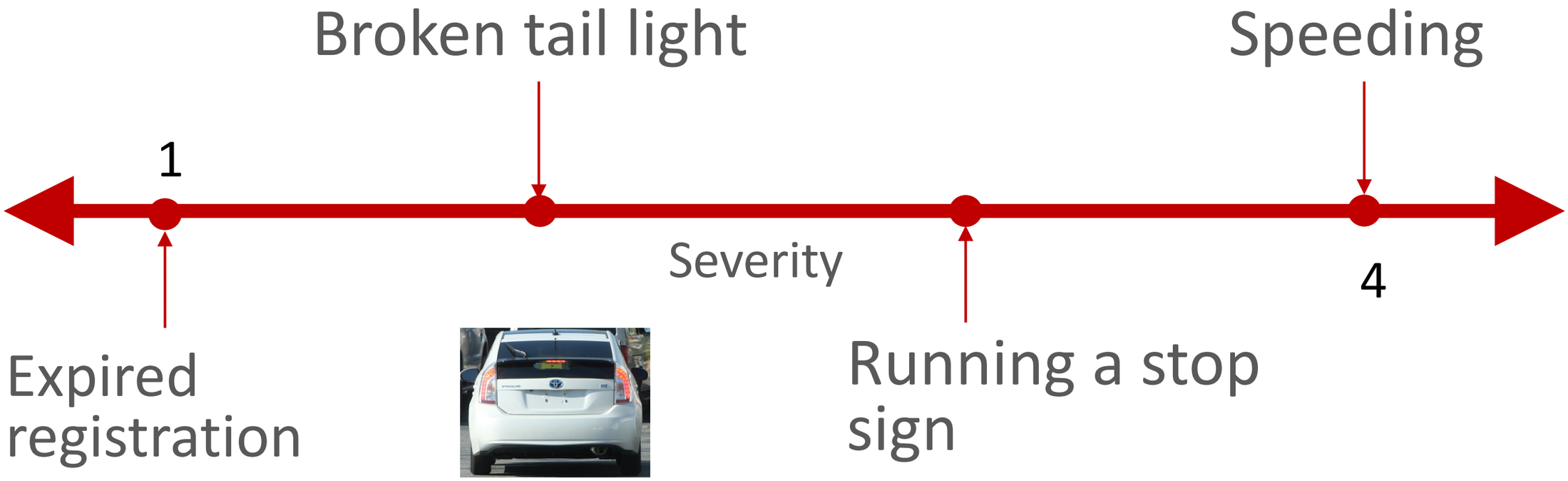
- Officer Race
- Driver Gender
- Crime rate in neighborhood
- Arrest status of driver
- What the driver was stopped for
- Replicates with data from DMV

Could the disparity be caused by police being less respectful in high-crime neighborhoods

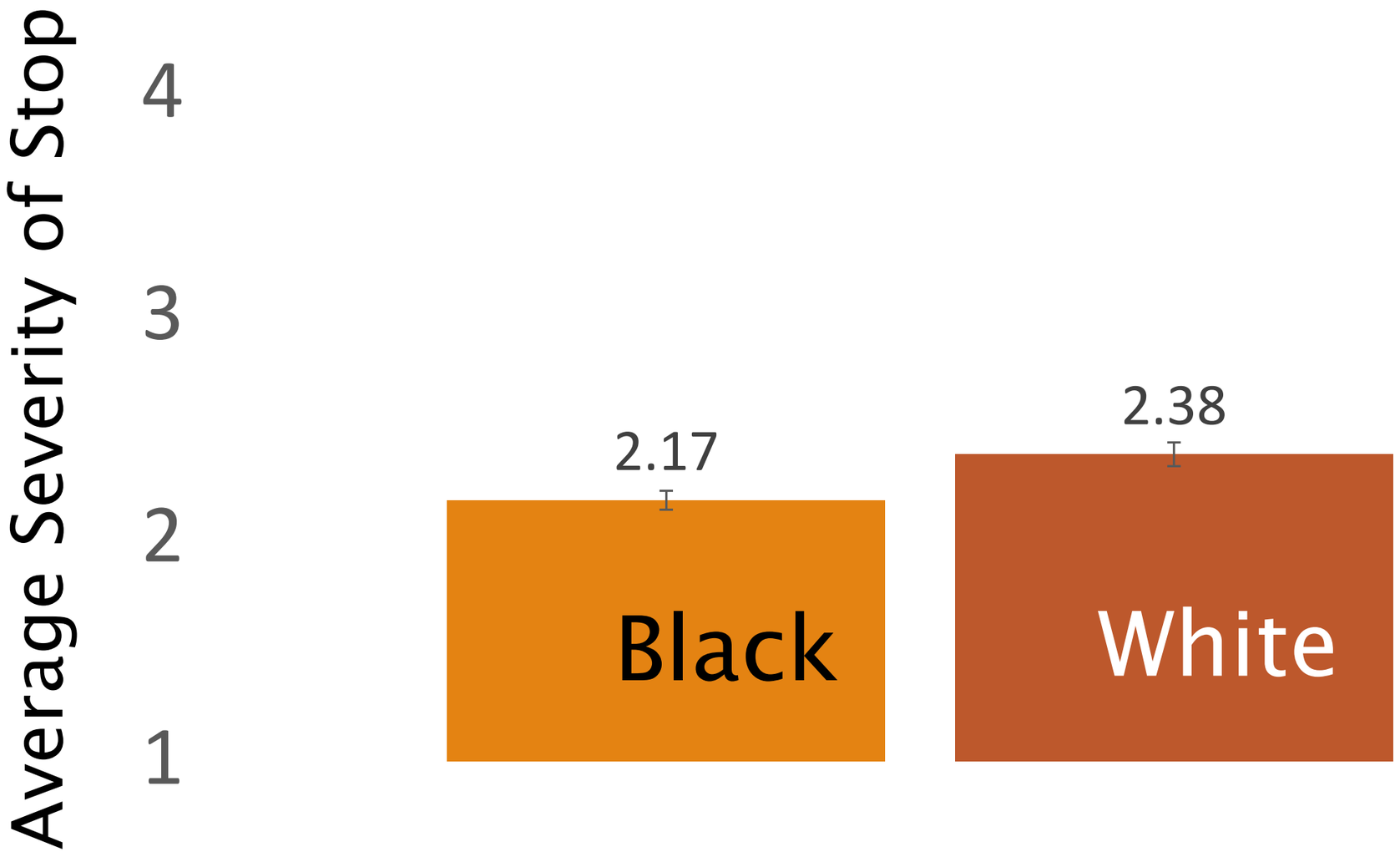
Nope

Another theory for disparity: Could police be more respectful to white people because they are stopped for more minor offenses?

We asked police officers to code every stop for severity of the infraction



Black motorists are stopped for less severe violations than whites



Could the racial difference be because the raters are college students

Replicated the lab study with large, racially diverse sample

Police still rated as more respectful to white drivers

- Participant race doesn't matter

NLP quantifies how black and white drivers have very different interactions with police

1. Escalated stops of black drivers begin in escalation
   - With **orders** rather than **reasons**
2. More generally:

   White drivers tend:
   - To be spoken to with more respect
   - To have concern for their safety expressed

   Black drivers tend
   - To be spoken to with less respect
   - To have their legitimacy challenged

Can we use insights from our study to improve officer training?

- Can NLP be a tool for policy-makers?
- Use NLP to help develop training materials
  - Incorporated into a procedural justice training.
  - All officers are trained
- Does training improve officer-community interaction? Examine recordings of 122 officers pre- and post-training
  - Yes! (paper under review)

Summary: The first NLP analysis of police body camera footage

- Quantifies reports about disparate treatment of black Americans
- Allows us to measure and improve officer training
- NLP can help us both understand our social world and hopefully help to make it better

Part 2. Can NLP help us see how Immigrants are Framed in US Political Discourse?

PNAS

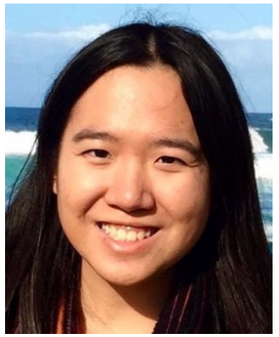
RESEARCH ARTICLE | COMPUTER SCIENCES
POLITICAL SCIENCES

OPEN ACCESS

Computational analysis of 140 years of US political speeches reveals more positive but increasingly polarized framing of immigration

Dallas Card, Serina Chang, Chris Becker, Julia Mendelsohn, Rob Voigt, Leah Boustan, Ran Abramitzky, and Dan Jurafsky,

Dallas Card



Serina Chang

Chris Becker

Julia Mendelsohn

Rob Voigt

Leah Boustan

Ran Abramitzky

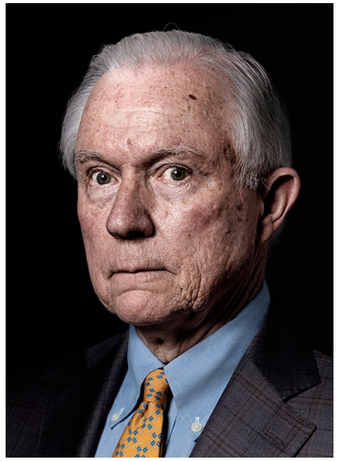
Echoes in anti-immigration rhetoric

"There is an appalling **danger** to the American **wage** earner from the **flood** of **low**, **unskilled**, **ignorant**, **foreign labor** which has **poured** into the country for some years past"

- Senator Henry Cabot Lodge (1896)

"[W]e absolutely must not **flood** the labor market with **foreign workers**—legal or **illegal**—in order to bring **wages** down."

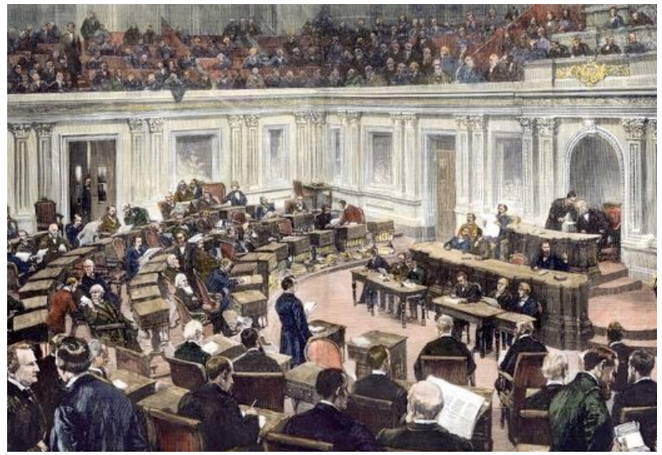
- Attorney General Jeff Sessions (2018)



Studying the History of U.S. Immigration using 200,000 Congressional Speeches

Use the **Congressional Record** (1880-2020), to study the debate about immigration in U.S. politics

1. Changes in attitudes and **polarization**
2. Varying emphases and **frames**
3. Use of dehumanizing **metaphors**

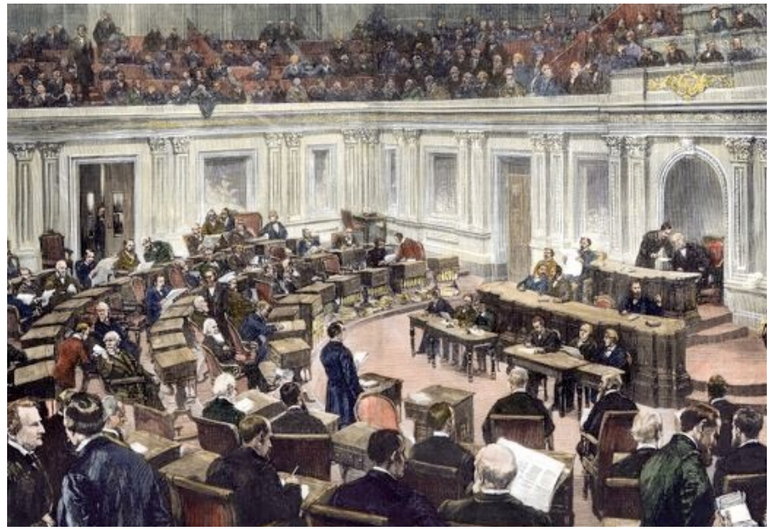


1. Changes in attitude and polarization

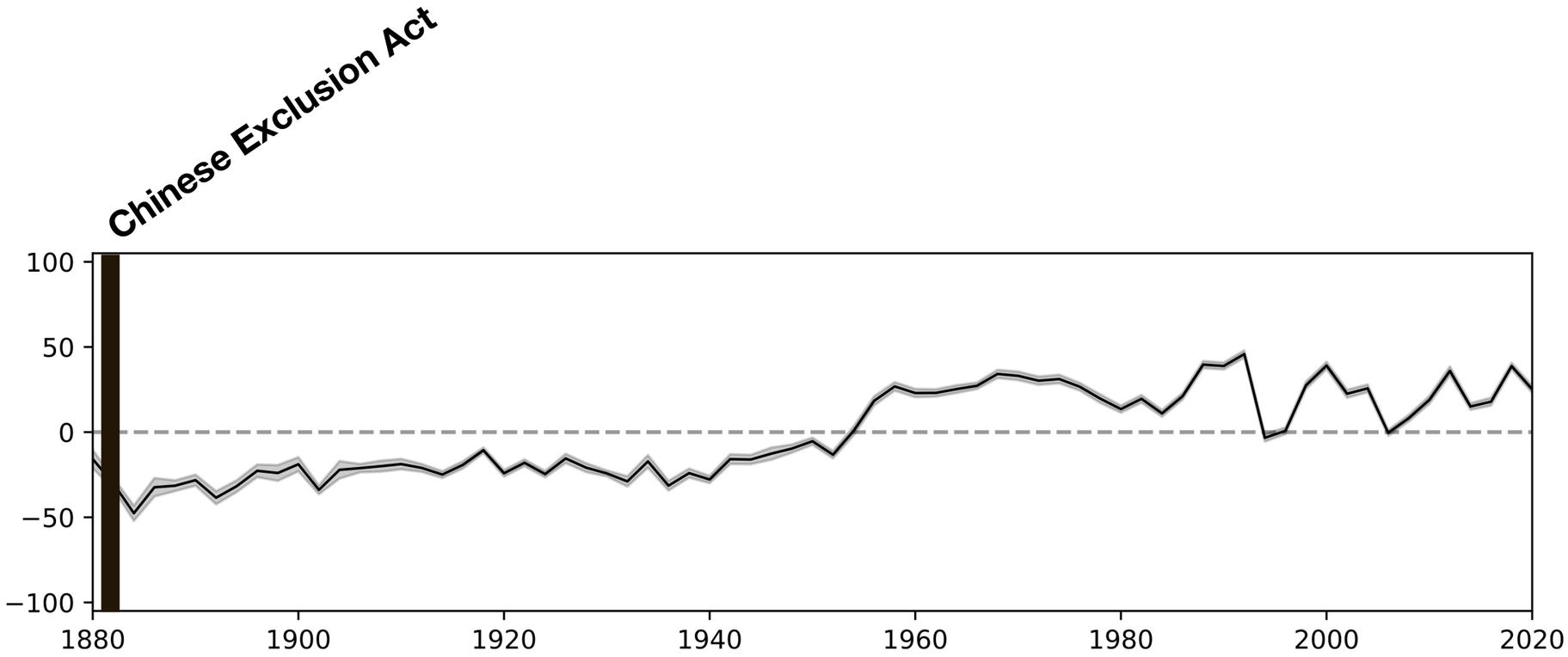
Build NLP classifiers to detect immigration speeches

Build NLP classifiers to detect pro-immigration vs anti-immigration stance

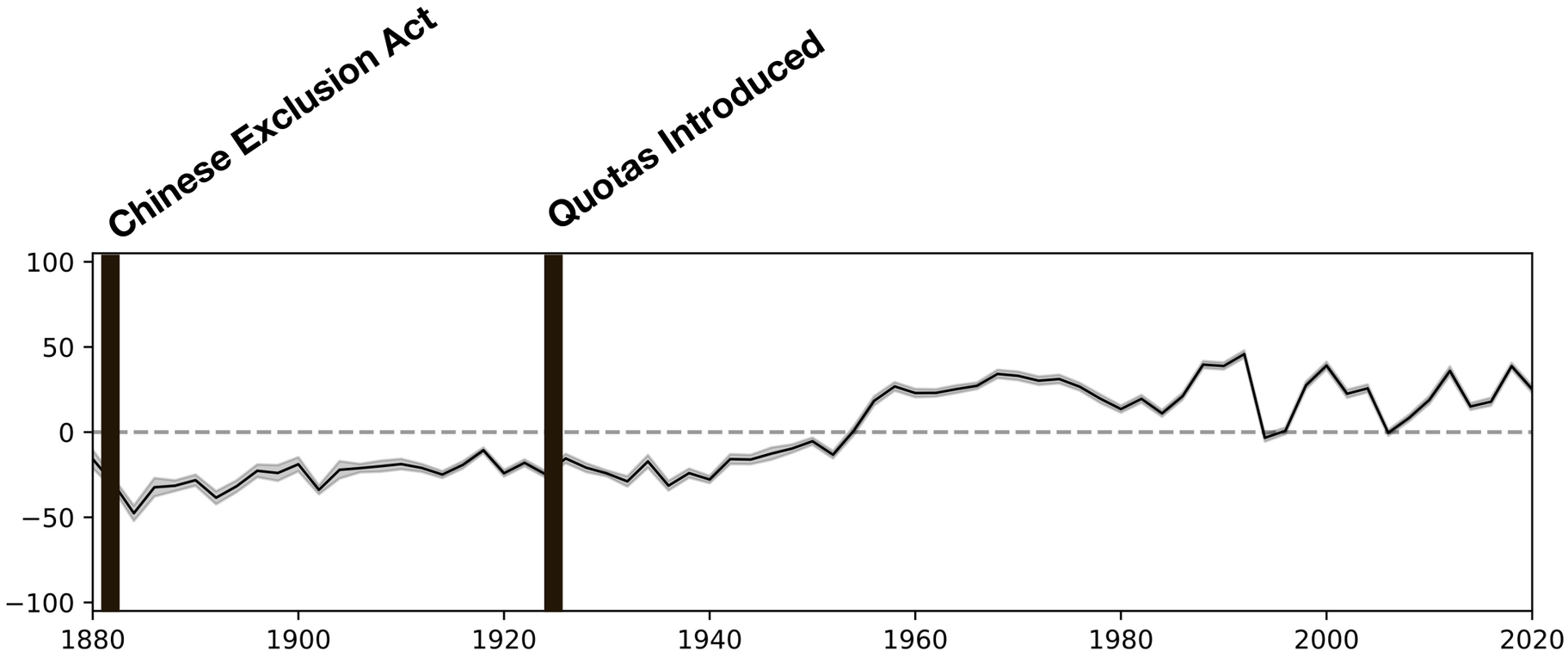
Plot "% Pro - % Anti" over time and over political parties



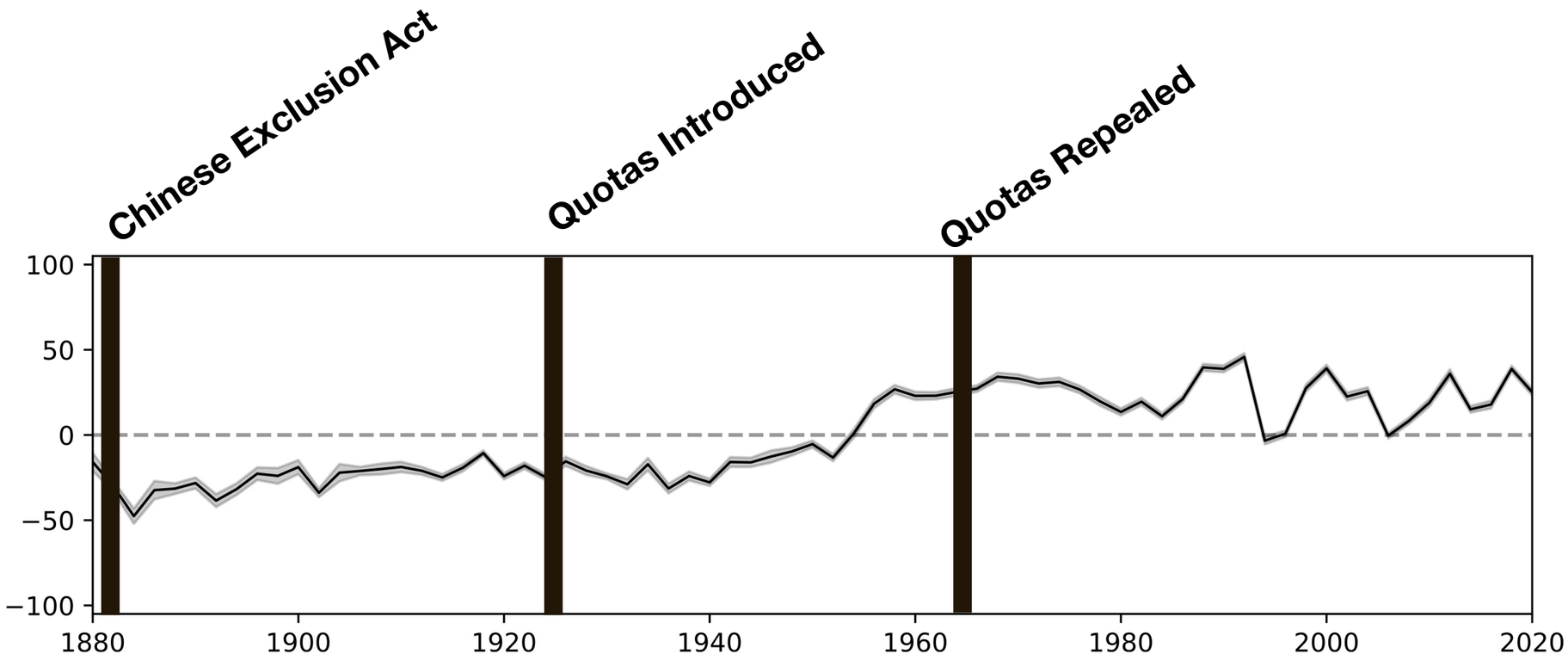
% Pro speeches - % Anti speeches



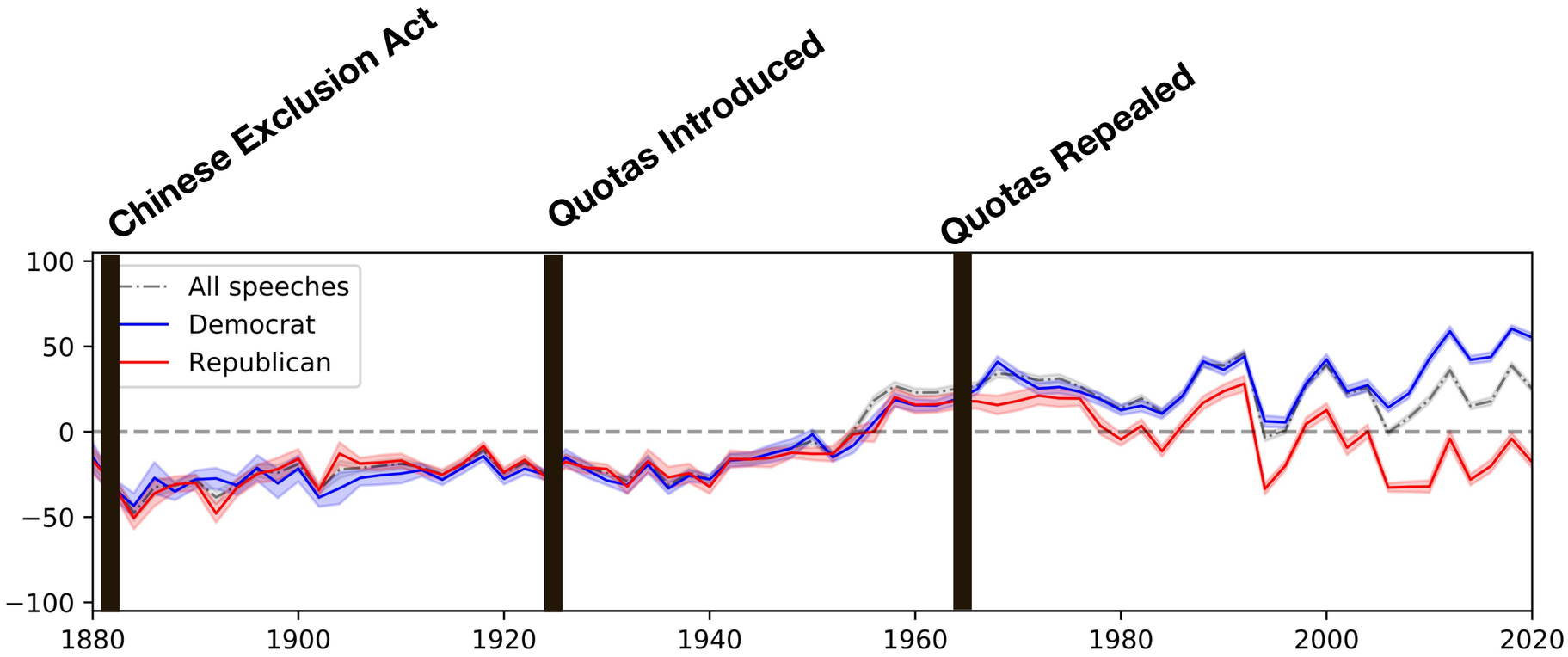
% Pro speeches - % Anti speeches



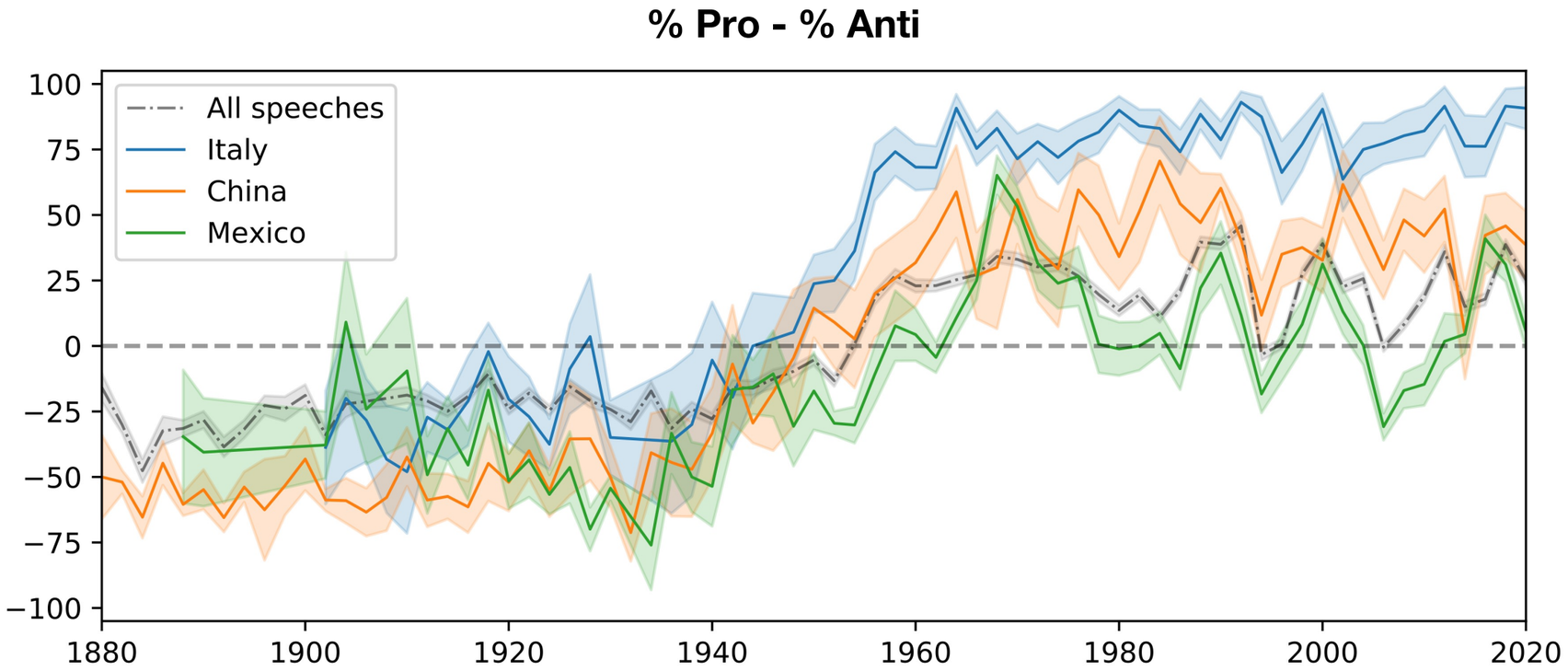
% Pro speeches - % Anti speeches



% Pro speeches - % Anti speeches



Tone and nationality



2. Framing Lexicons

Identify terms that are used disproportionately in reference to immigrants

e.g., **stop illegal** immigrants **pouring**
…

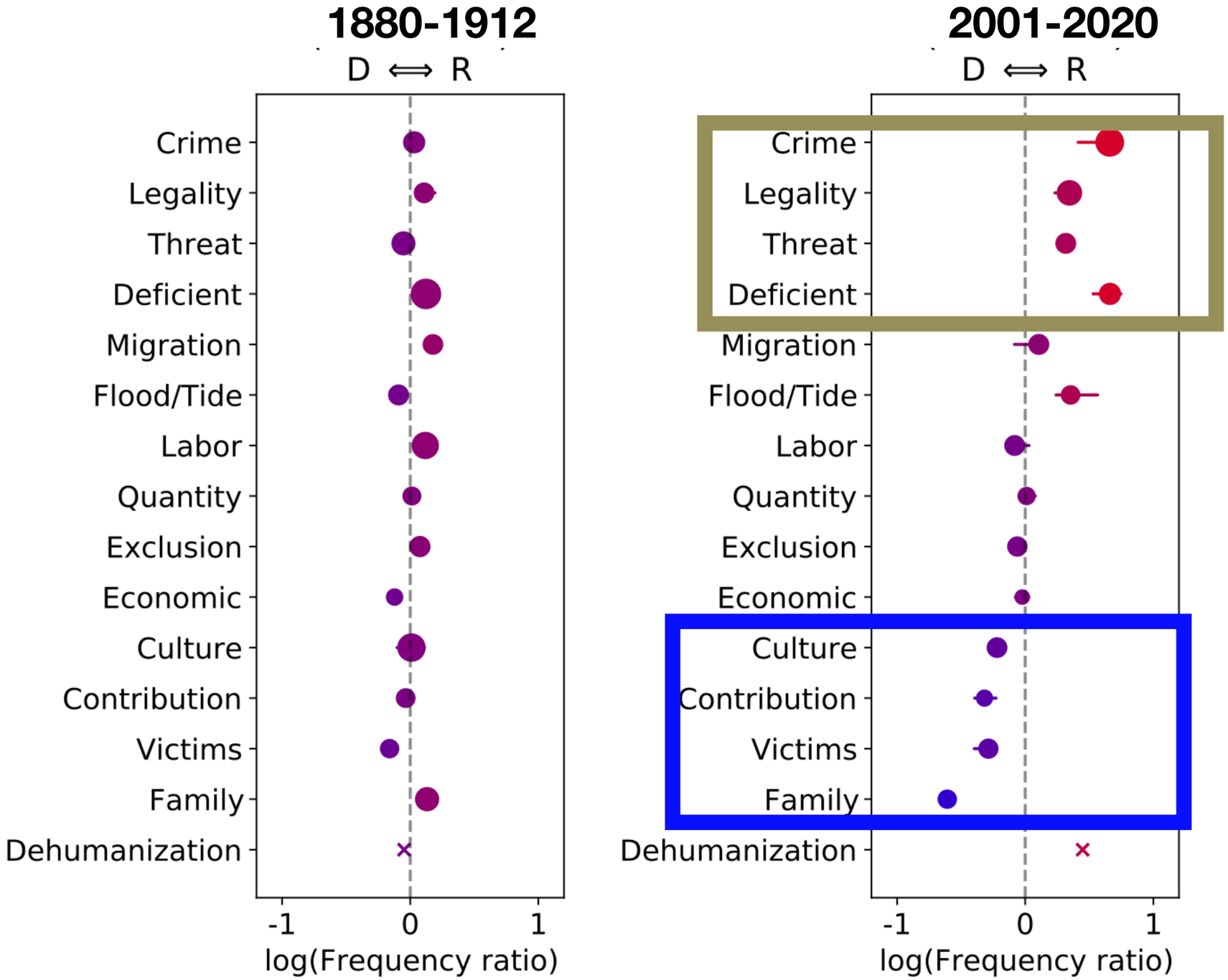
Curated into 14 frames

Use word vectors to expand lexicons

Example: Flood / Tide Frame

absorb (v), absorption (n), drain (v), fill (v), flood (n), flood (v), flow (n), flow (v), inflow (n), influx (n), outflow (n), pour (v), spill (v), stream (n), stream (v), surge (n), tide (n), trickle (n), wave (n)

Frame use by party



3. Metaphorical Dehumanization

Santa Ana, Otto. "Like an animal I was treated': Anti-immigrant metaphor in US public discourse." *Discourse & Society* 10, no. 2 (1999): 191-224.

Words like "flood", "wave", "stream", "pour" are often used to describe immigrants in a dehumanizing way:

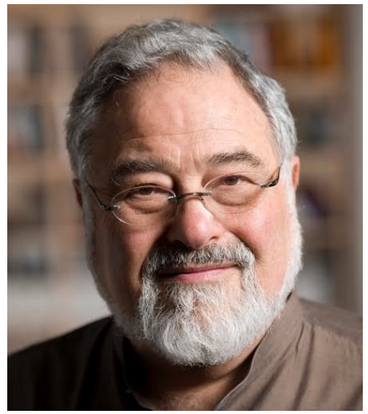
*"the **flood** of illegal immigrants **streaming** into the country"*

These words **implicitly** evoke a metaphor:

Immigrants are a catastrophic force of nature

Metaphor in modeling implicit concepts

Lakoff and Johnson 1980 "Metaphors we live by"



George Lakoff

Argument is War

- Your claims are *indefensible*.
- They attacked *every weak point* in my argument.
- She *demolished* their argument.
- I've never *won* an argument with them.
- They *shot down* all my arguments.

Time is Money

- This app will *save* you hours
- I don't *have* time to *spare*.
- How do you *spend* your time these days?
- That flat tire *cost* me an hour
- I've *invested* a lot of time in this.

Metaphorical Dehumanization

Santa Ana 2002, O'Brien 2003, Haslam 2006, Cunningham-Parmeter 2011, Mendelsohn et al. 2020

Language used	Implicit Metaphor
"the **herding** of these aliens in stockades "	Animal
"…immigrants will **swarm** over our land and **devour** its resources."	Vermin
"illegal immigration **infects**… "	Disease
"the **flood** of illegal immigrants **streaming** into the country"	Disaster
"prevent the **dumping** of undesirable immigrants into the country	Cargo

Implicit Metaphorical Dehumanization

Politicians don't **explicitly** say "rat" or "disease" on the floor of Congress

How to measure *implicit* dehumanization?

- Instead of "animal": "immigrants **feeding** at the trough"
- Instead of "vermin": "immigrants **crawling** across the border"
- Instead of "cargo/objects: "**dumping** of immigrants into the country"

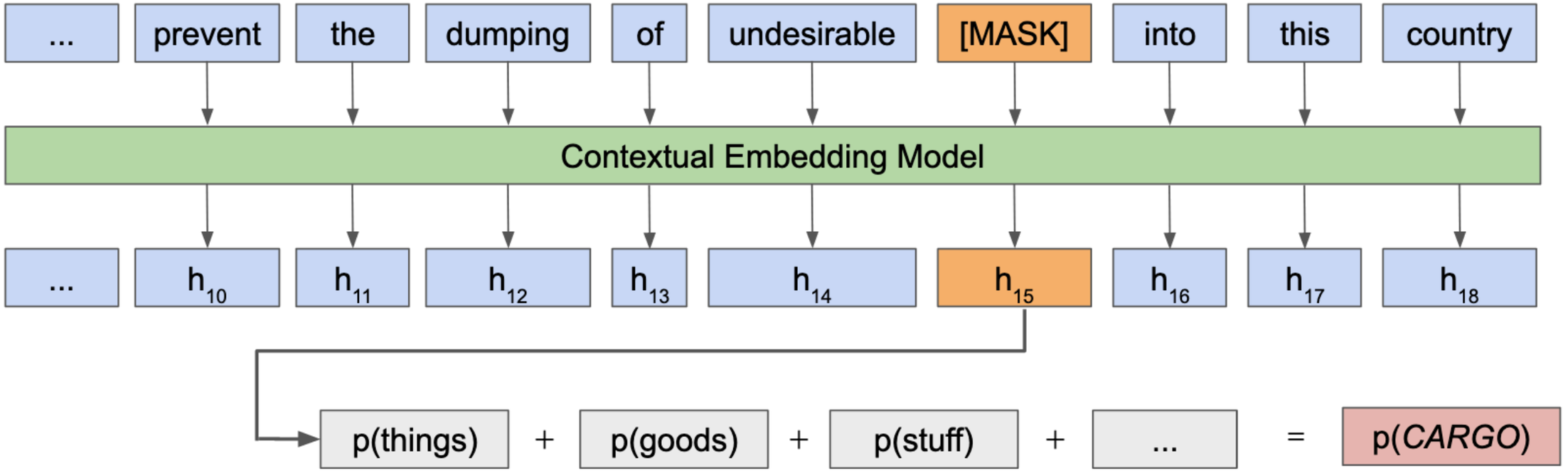
Idea:

- Take texts describing immigrants
- Mask out the immigrant word
- Ask a large language model what word it thinks the speaker is implicitly saying

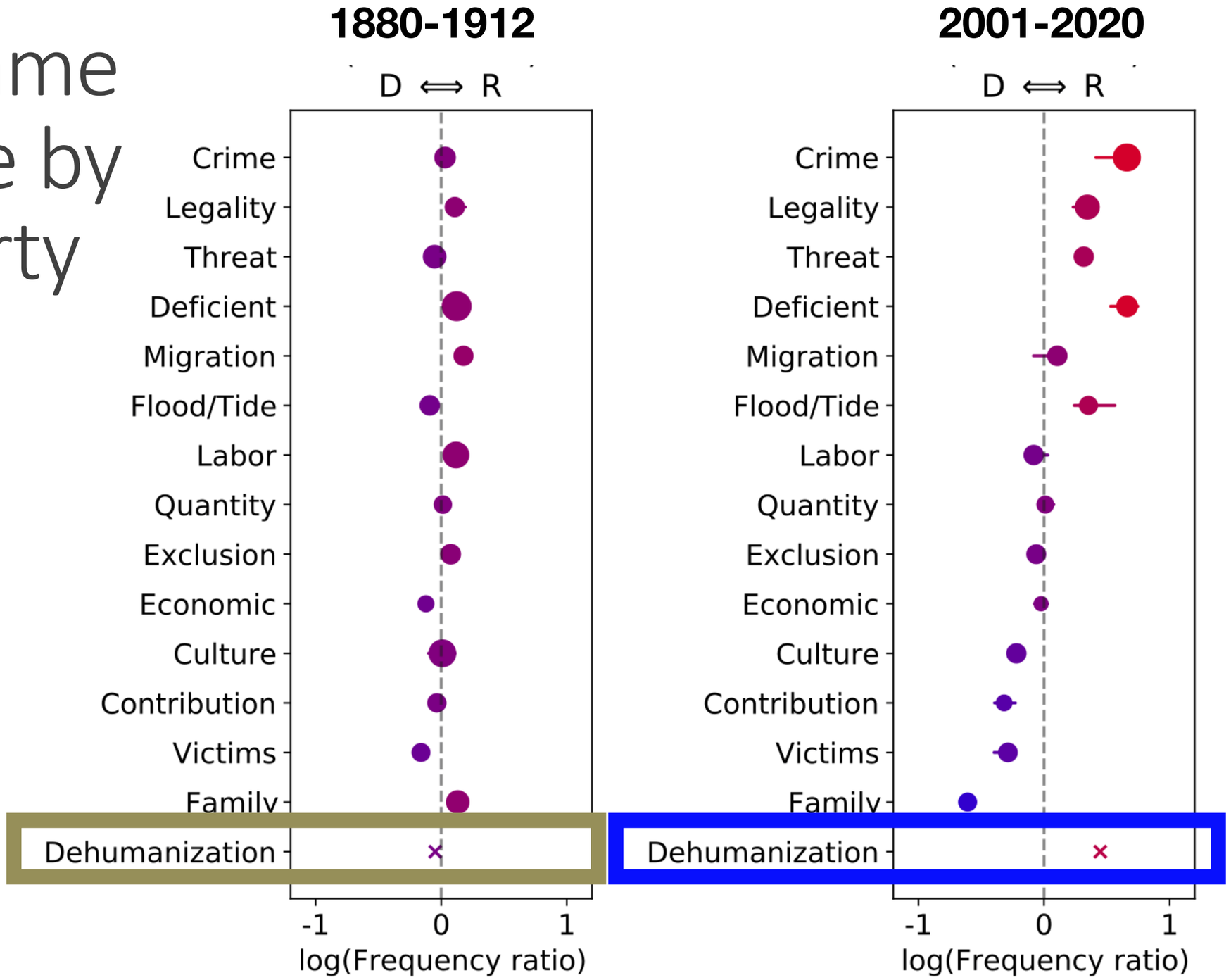
Metaphorical Dehumanization

Computing the probability of the CARGO metaphor from one text

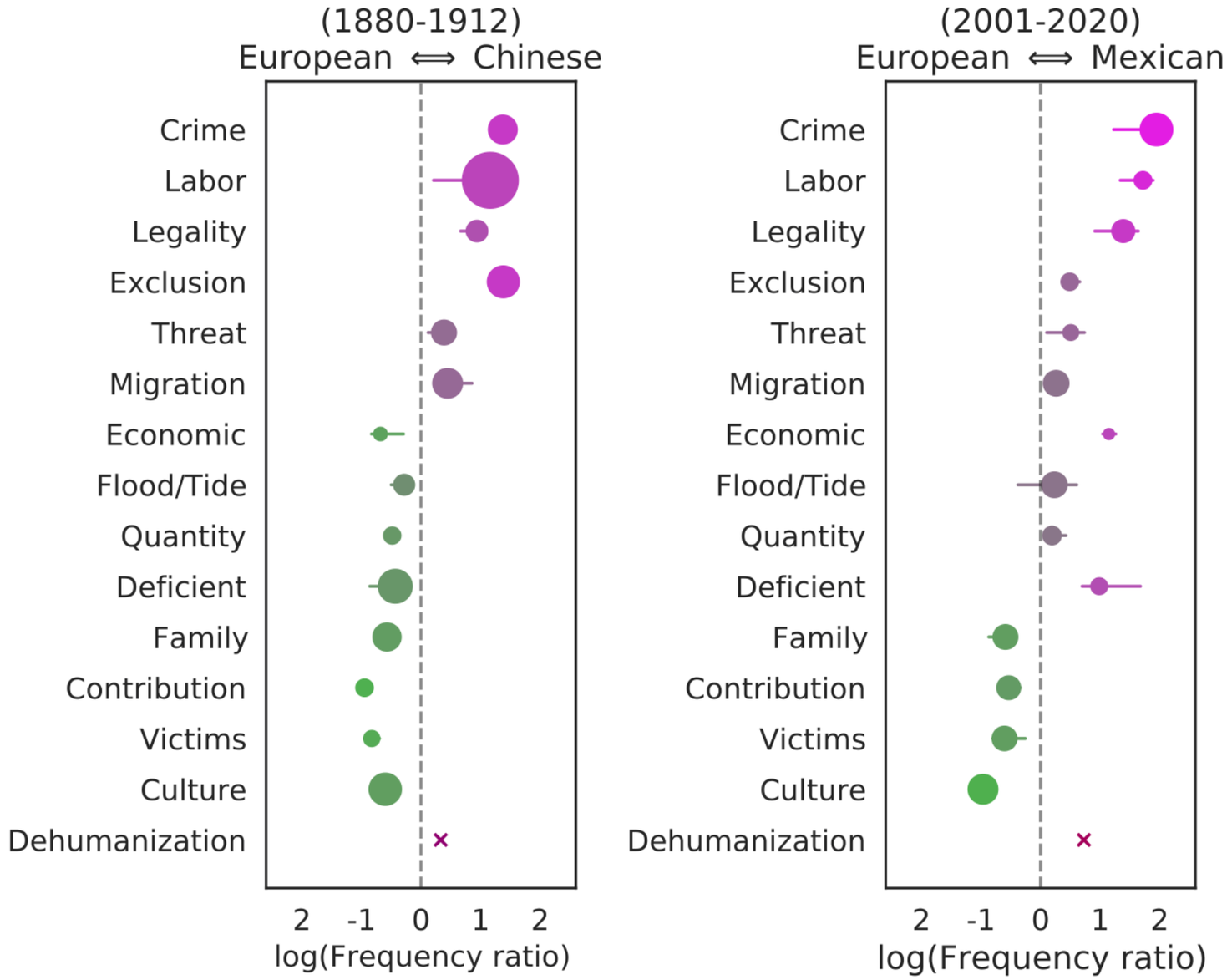
`prevent the dumping of undesirable immigrants into this country`



Frame use by party



Frames and Nationality 1900 vs 2020



Summary: NLP to study Immigration Attitudes

NLP can help us understand US attitudes toward immigration

Dramatic **rise in pro-immigration attitudes** after WWII

- Republicans then decline to 1890 or 1920 standard

Divergent use of **positive** (e.g., families) and **negative** (e.g., crime) **frames**

- Held by Republicans vs. Democrats
- Held toward European vs. Chinese/Mexican immigrants

LLMs can detect **dehumanizing metaphors** (animals, cargo, etc.)

- LLMs as exciting new tool for measuring implicit language!

NLP for Social Good and Social Science

NLP for Analyzing Police Body-worn Camera Conversations

Can we improve police-community relations?

NLP as tool for public good

NLP for studying political discourse about immigration

Can we learn about polarization, develop new ways to measure toxic speech like dehumanization?

NLP as tool for social sciences

I gave you examples from policing and politics.
But there are so many more possible applications!!!

Education (Dora Demszky's talk in two weeks)

Medicine and health

Food and nutrition

…

Any domain where there is text data!!