

"Does This Vehicle Belong to You?" Processing the Language of Policing for Improving Police-Community Relations

DAN JURAFSKY

STANFORD UNIVERSITY

UCSD LINGUISTICS

MAY 2018

Problems in Police-Community relations in the United States

- A flood of recent viral videos show inappropriate police officer use of force



- Black Americans report more negative experiences in their interactions with the police than other groups (Epp et al. 2014)
- 18% of blacks said they had been unfairly stopped in the past, while only 3% of whites felt the same way (Pew Survey 2016)



Unfair treatment reduces trust between police officers and local communities

By contrast, a person who is treated with **respect**

- Has more trust in the officer's fairness
- And in procedural fairness of the institution
- And is more willing to support the police
(Tyler, 1990; Tyler & Ho, 2001; Tyler & Sunshine, 2003 Mazerolle et al., 2013)



Can Computational Linguistics help?

Measure problems in police-community interactions?

Detect the potential for escalation?


And hopefully reduce the chances of violence?

Our idea: Use body-camera footage as data

Data from the Oakland Police Department

Their officers have been wearing body cameras since 2010





Look at common, everyday interactions with police

(Langton and Durose, Department of Justice, 2013)

¼ of US adults have contact with the police each year

Most police-initiated encounters are traffic stops

Our team:



Rob
Voigt



Nick
Camp



Camilla
Griffiths



Will
Hamilton



Rebecca
Hetey



David
Jurgens



Vinod
Prabhakaran

PI:



Jennifer
Eberhardt



Outline

Study 1: Respect in Police
Language in Traffic Stops

Study 2: Dialog Structure in
Traffic Stops

Study 1:

Rob Voigt, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt. 2017. Language from police body camera footage shows racial disparities in officer respect. PNAS.

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

- Police departments care about showing respect and building trust

Body-Cameras as Research Data

Every interactions between OPD and community for April 2014: 17,304 Videos

A window onto everyday behavior:

- only **vehicle** stops which resulted in a warning or citation (**no arrests**).

Look at the subset of 981 videos that are

- Vehicle stops
- With white or black community members.
 - 245 different officers

Study 1: **text**

The transcribed dataset

Professional transcribers

- Fingerprinted and background checked by the police department
- Watched videos
- Transcribed words
- Diarized (who is talking to whom)

Resulting data set:

- 36,738 utterances, 324,506 words
- by police to community members in traffic stops



Methodological Aside

All the faculty and grad students were also fingerprinted and background checked by the police department

Sample transcription

0:00:00 0:00:09 OFFICER [to dispatch]: Unknown occupant and it's going to be for registration. It should be code four.

0:00:20 0:00:20 OFFICER: Hi.

0:00:20 0:00:20 FEMALE: Hi.

0:00:21 0:00:23 OFFICER: I pulled you over because your registration is expired by almost a year.

0:00:25 0:00:28 FEMALE: Okay, I have the paperwork for it, a moving permit?

0:00:28 0:00:28 OFFICER: I'm sorry?

0:00:29 0:00:30 FEMALE: I have the paperwork for it.

0:00:30 0:00:31 OFFICER: Okay.



Part A: Perceptions of Officer Treatment from Language

Can human raters judge respect from officers' language?

Are there differences in officer respect towards black versus white community members?



First have humans label respect

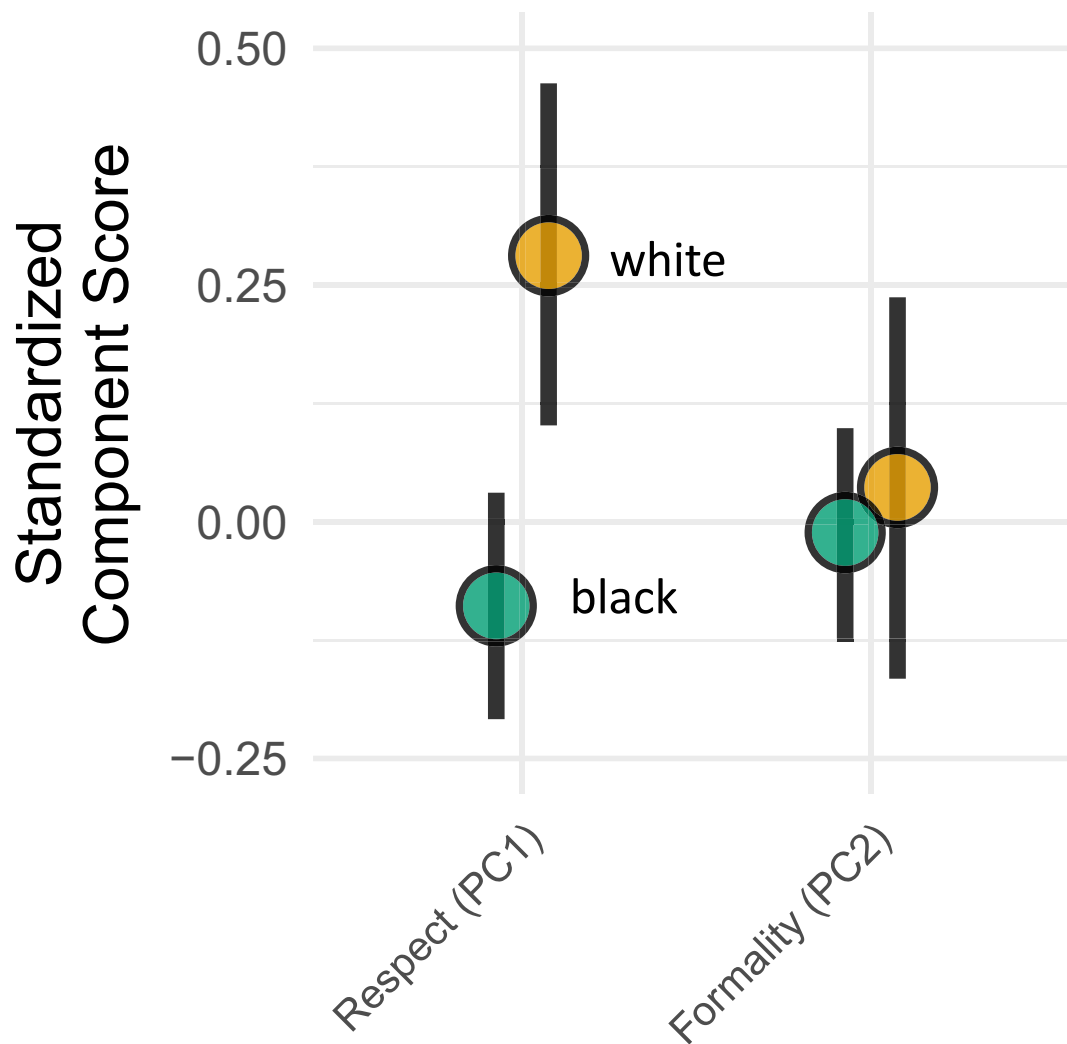
Participant Coders (N=70) blind to citizen race labeled 414 unique officer utterances

- 4-point Likert scales (high rater agreement $\alpha s = .73-.91$)
- Respectful, Polite, Friendly, Formal, and Impartial
- Two principle components

Race on two principle dimensions

Police are more respectful to whites

Police are equally formal/distant (close/familiar) with whites & blacks



But can we compute respect automatically?

1. Engineering: 26 million stops a year
2. Science: Could control for confounds





NLP for respect!

1. Use linguistic theories of respect to develop features
2. Use the human labeled data to learn feature weights
3. Build a classifier to label the Respect/Formality of every sentence

Drawing on prior work on 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.

- Wikipedia editors talk pages
- Stack Exchange questions

Politeness correlates with social role

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 are more polite than Python programmers

Gender:

- Women are more polite

Status:

- Wikipedia editors become ruder after they are elected to admin positions

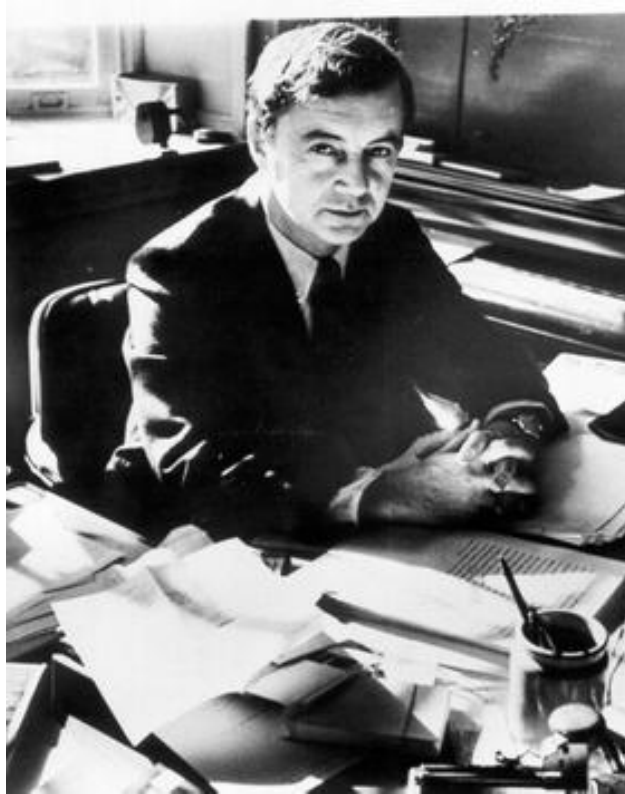


Let's apply this to traffic stops

But first, some methodology

Respect or politeness is more than "please" and "thank you"

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



Erving Goffman



Robin Lakoff



Penelope Brown



Stephen Levinson

Respect or politeness is more than "please" and "thank you"

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

Politeness addresses two basic desires

(1) Negative Politeness: Desire not to be told what to do

- Requests impose on the addressee
- Social peril for failing to comply
- Negative politeness mitigates these social threats.
- **Minimize the request**
- **Put on record that it's an imposition**

"Sorry, I know you're busy, but could you just review this one paper?"

Cues for Negative Politeness

Brown and Levinson (1978), Culpepper (1976), Pennebaker et al. (2007), Danescu-Niculescu-Mizil et al (2013)

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”

What is politeness?

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

Politeness addresses two basic desires

(1) Negative Politeness: Desire not to be told what to do

(2) Positive Politeness: Desire to be paid respect

- Emphasize your value and our relationship

"Hey, that was a really beautifully written review, you must have spent a lot of time on it!!"

Cues for Positive Politeness

Brown and Ford (1961), Culpepper (1976), Brown and Levison (1978), Pennebaker et al. (2007), Danescu-Niculescu-Mizil et al. (2013)

Formal titles

“ma'am”, “sir”, “Mr.”

Introductions

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

Sympathy or concern: mentioning safety

“Drive safely”

Train a simple supervised classifier

- Linear regression
- Using these linguistic features
- 414 human labeled utterances training
- Result:
 - Cross-validated R^2 of .258
 - RMSE .840 compared to human-to-human RMSE .842



What kind of utterances are high in respect?

Apologies

Gratitude

Reassurance

Safety

Formal titles



Now run the classifier on
36,738 utterances

For each utterance we have:

- Automatically assigned respect score
- Automatically assigned formality score

Sample classified utterances

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



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

Linear mixed-effects model coding for many
factors

Random intercepts for interactions nested
within officers



Results

(1) Officers are more respectful to older drivers



Results

(2) Officers are more respectful to white drivers

- No significant disrespect to black drivers
- Just extra respect to 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, **sir**, I'm just going to give you a citation for the cell phone use, okay?"

"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?"

"All right. You **have a safe night,** okay?"

"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.**

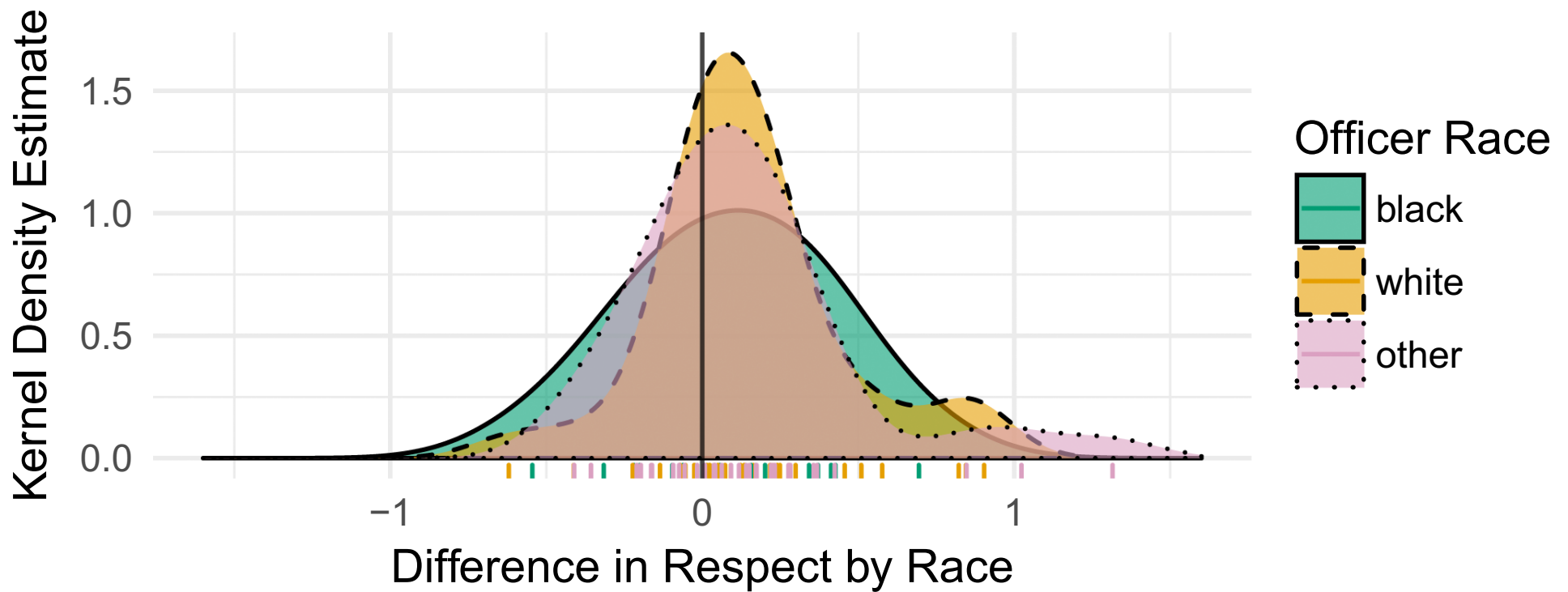
"**No big deal,** just make sure you get those things fixed.

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



Could this be an artifact of
some confounding variable?

The racial disparity in respect is not an effect of officer race



Maybe the disparity is caused by police being less respectful in high-crime neighborhoods


Nope





Or just being less respectful to
men

Nope



Maybe the racial difference is caused by police being less respectful to criminals?

To test this hypothesis:

- Remove all stops where a driver was searched
- Criminals on probation or parole can be freely searched (and therefore are searched)

Police are still more respectful to white drivers

Maybe the racial difference is because the raters are college students

Replicated the lab study with large, racially diverse sample



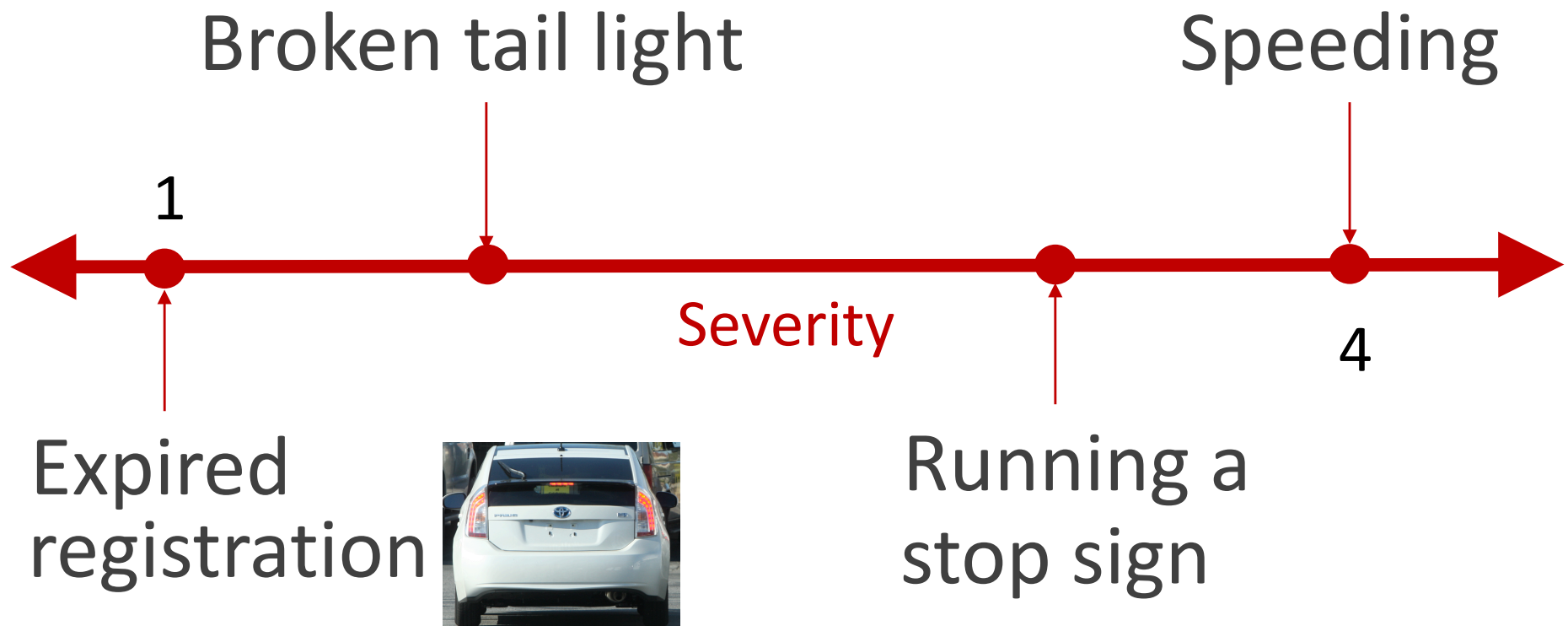
Police still more respectful to white drivers

- Participant race doesn't matter

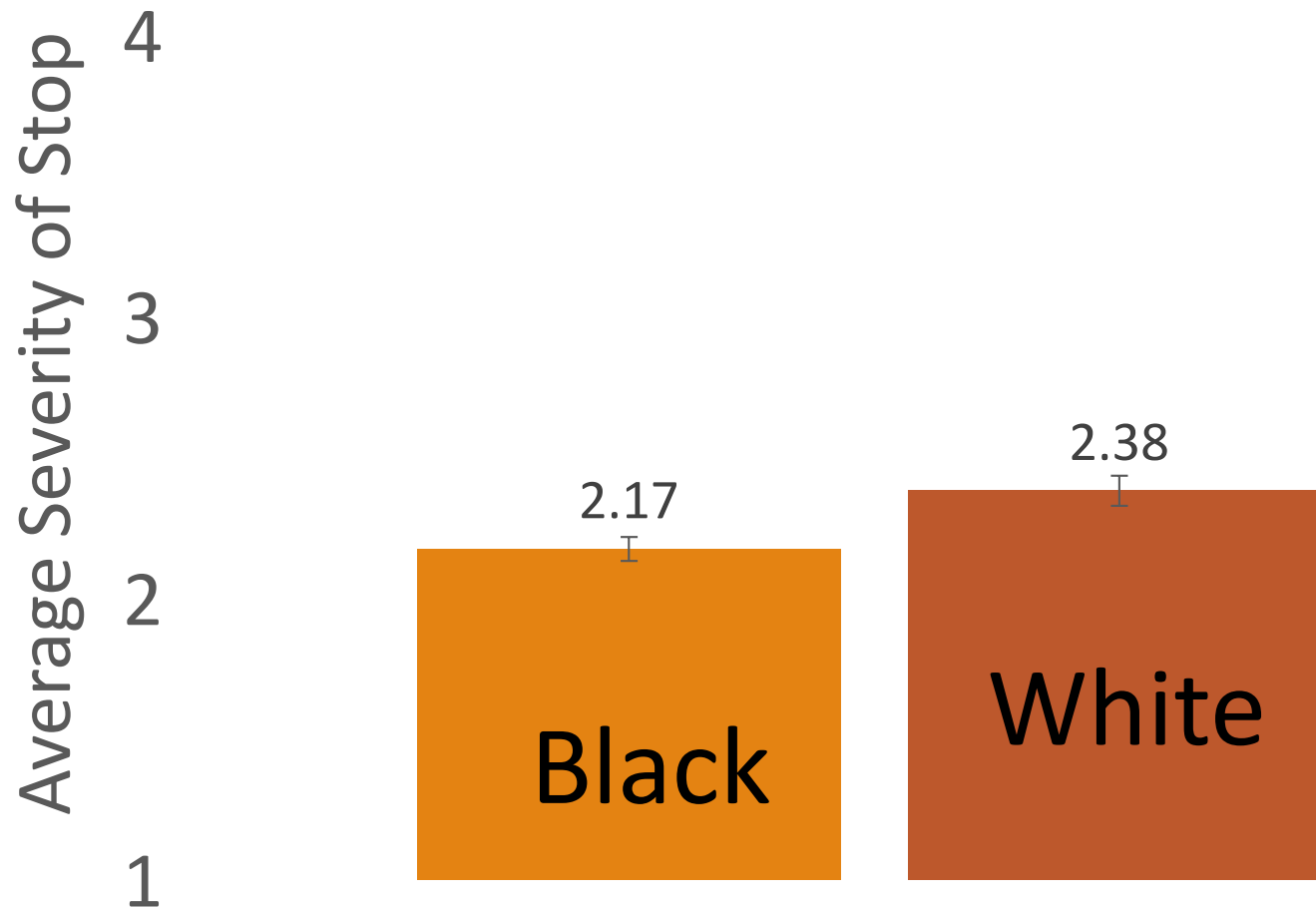
Maybe police are more polite to white people because they are stopped for more minor offenses?



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



Black motorists are stopped for less severe violations than whites



Maybe racial disparity in respect caused by racial differences in outcome?

(i.e. whether the driver got a citation or was let off with a warning)?

Nope.



Maybe the racial disparity is only in words, not tone of voice

- Prosody is an important cue to social meaning!



- Are there racial differences in police prosody?

We replicated the lab study with prosody only

Preliminary results

~15 second tiles of officer speech

- Mundane traffic stops, male drivers
- Low-pass filtered

Get humans to label:

- Respectful/Talking Down
- Tense/At Ease
- Friendly/Cold

Police prosody

Talking down:



Respectful:





Do people detect prosodic differences associated with driver race?

Small but significant effects of race in our pilot

Officers are more *respectful, warm, at ease* when talking to white drivers

Preliminary results

We can't be certain yet what causes these racial disparities

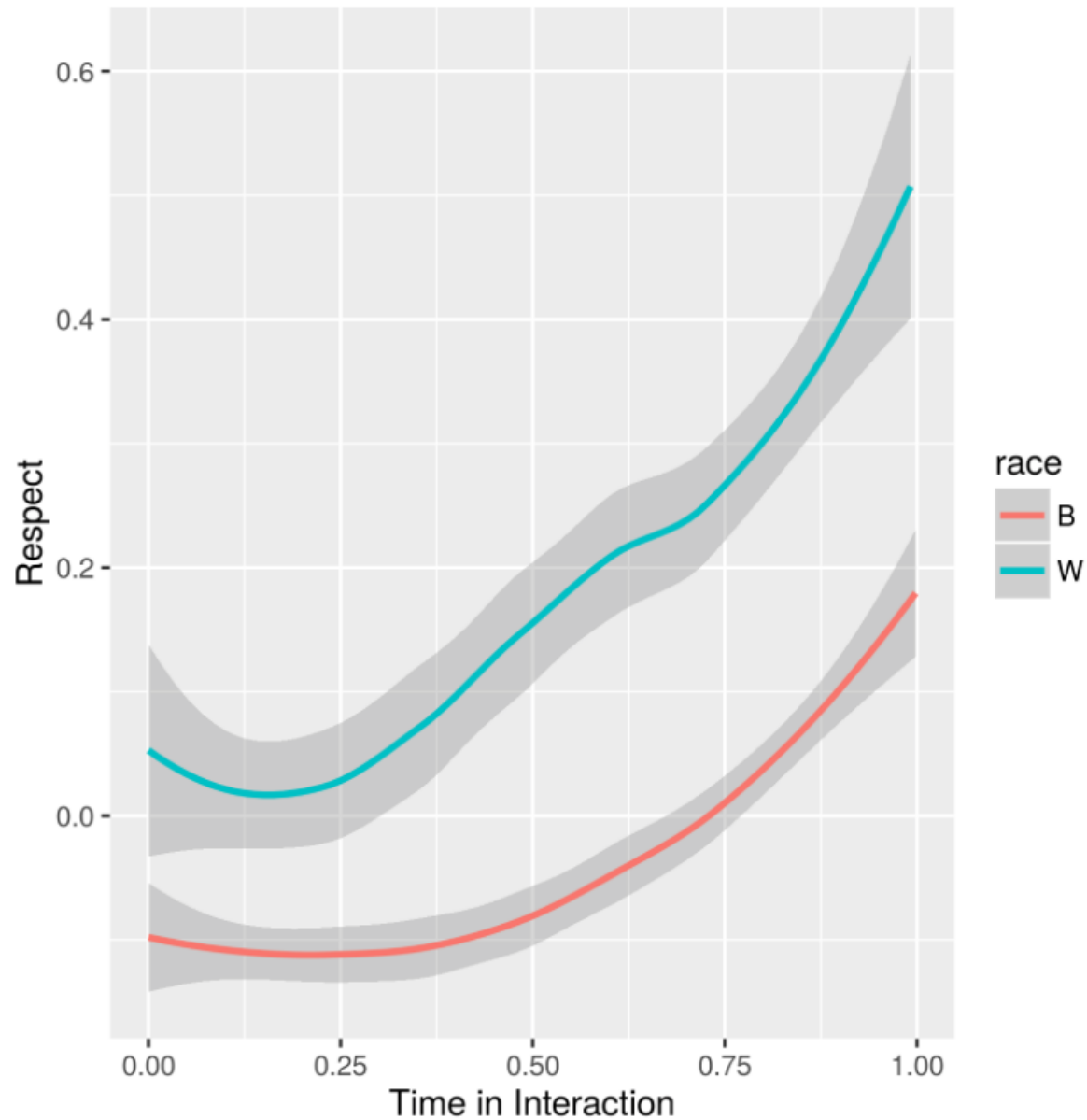
(1) Disparities don't necessarily mean racial bias on the part of officers

(2) Disparities might also be partially caused by driver language

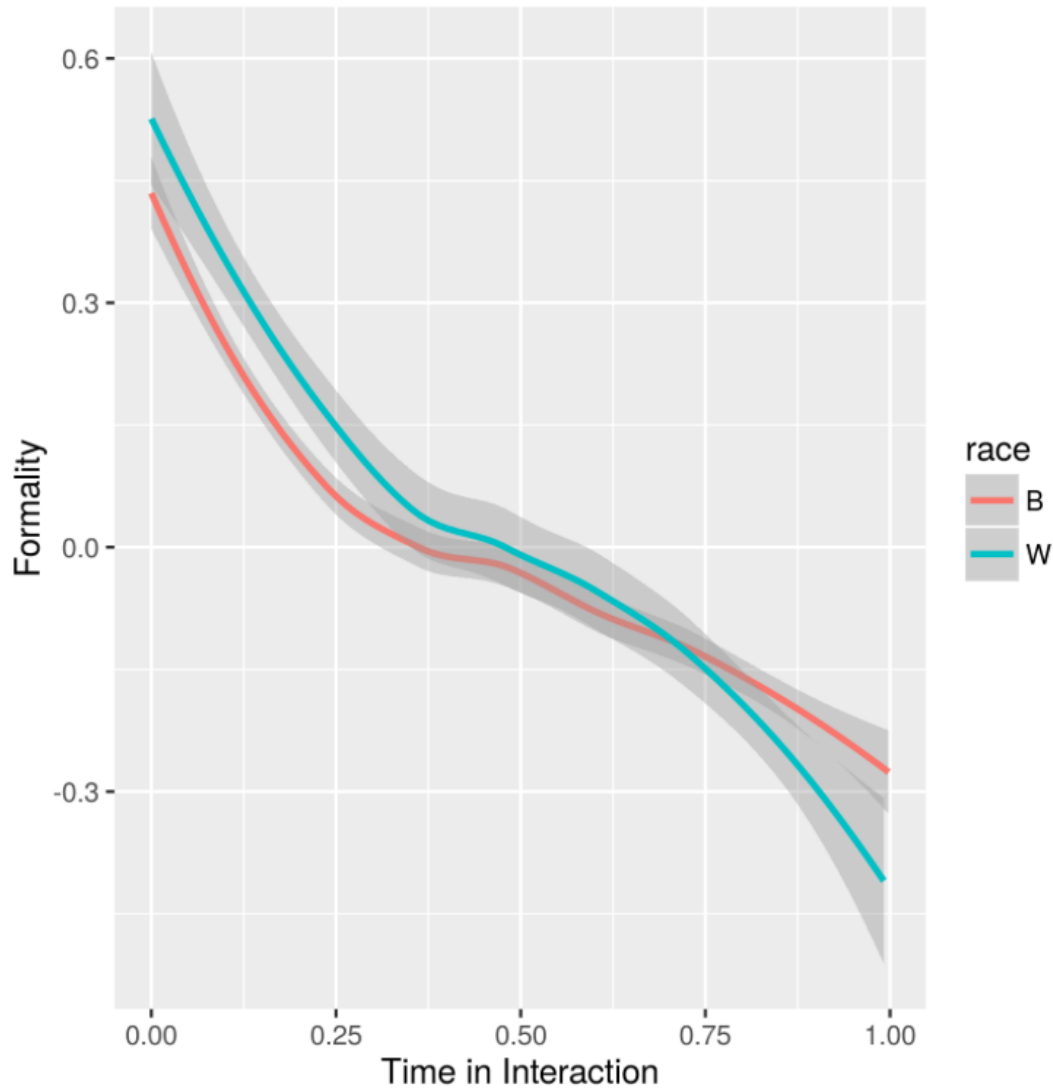
But driver language can't be sole or major cause

- Lab study found racial disparities even when rated in context of driver's utterances
- We found racial disparities at the very initial seconds of the interaction

Racial disparity in respect at every stage of the interaction



Officers become less formal across the conversation



No effect of race.

Could respect have implications for escalation?

Work in progress

We looked at Anger and Swear words in driver language

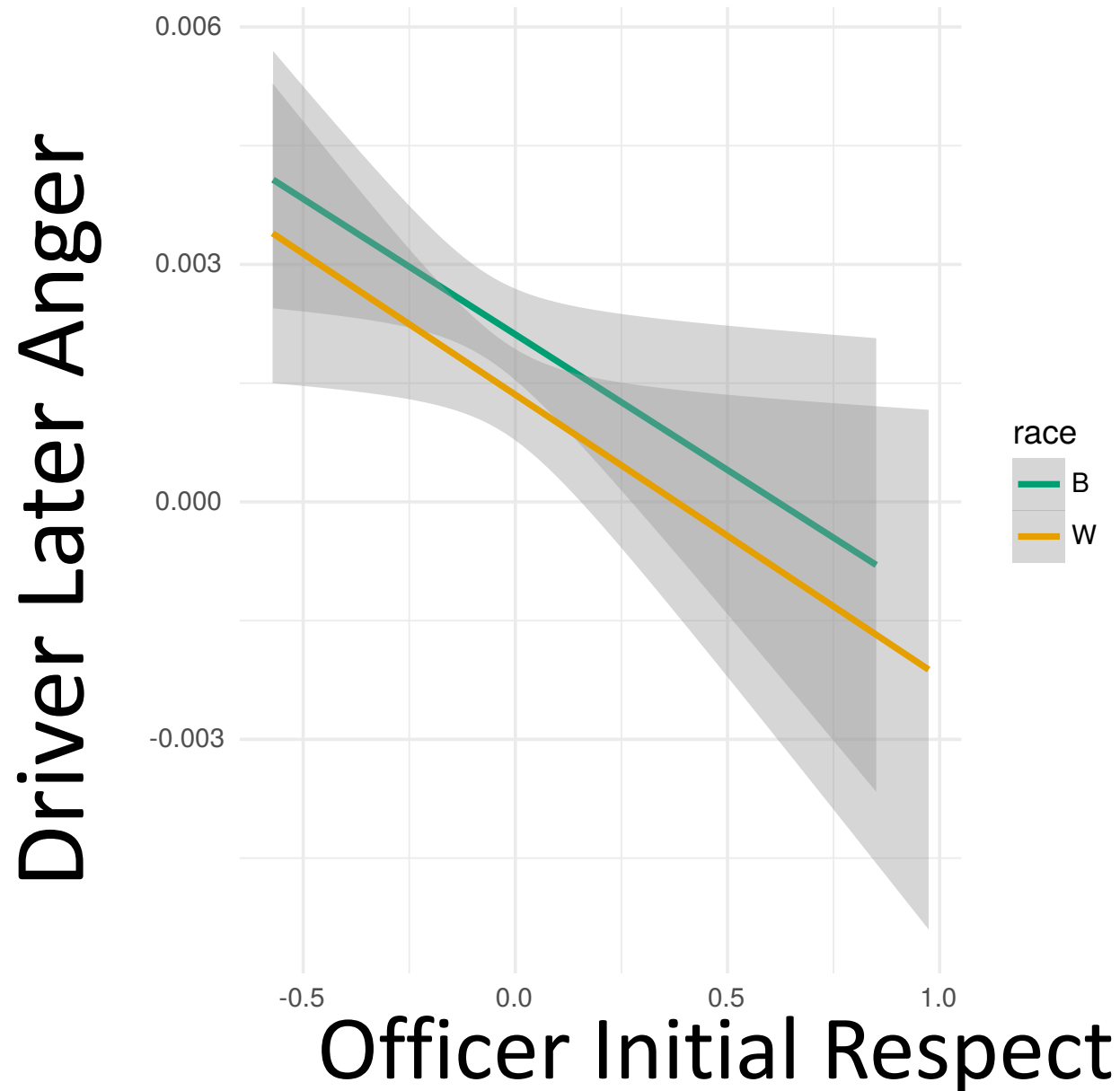
When officers are more respectful, are drivers less angry?

Initial study:

- Officer respect in first half of interaction
- Motorist anger in second half

Yes! Office respect correlates with lower driver anger

Preliminary results



Conclusions

Police officers in our study are *less respectful* to black community members

- Even in everyday encounters with no arrests, searches, uses of force
- Even though black community members are stopped for less severe offenses
- In words and tone of voice

Respect matters for fairness but also for everyone's safety

- Officer respect is associated with lower driver anger

Study 2 (Work in Progress)

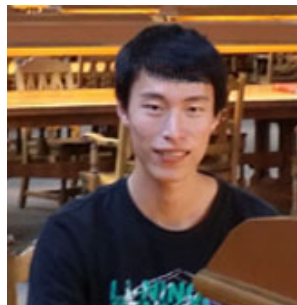
Can we model the richer dialog structure of police interactions with community members?



Vinod
Prabhakaran



Camilla
Griffiths



Hang Su



Prateek
Verma



Nelson
Morgan



Jennifer
Eberhardt

"task-oriented dialogs have a structure that closely parallels the structure of the task"

Barbara Grosz



OFFICER: Sir, hello, my name's Officer [NAME] of the Oakland Police Department.

Greeting

Giving Reason

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.

Asking Details

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.

Issuing Sanction



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.

Good Bye

OFFICER: Thank you.

Dialog structure has policy implications!

1. 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."

2. 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 complain they get asked intrusive and investigatory questions, especially in certain neighborhoods.

Are there differences in who gets asked these questions?

What kind of dialogue structure?

Institutional Dialog Acts

Institutional Talk (Heritage 2005)

1. Speech Acts
2. Dialog Acts: conversational analytic structure (Schegloff etc.)
 - Greetings
 - Farewells
3. Adds task-related structure related to scripts
 - Asking for documents
 - Issuing a citation

Some of the Dialog Acts

Greetings (“Hey, how are you?”)

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.”)

Classifying 21 Dialog Acts

Data

- Total number of annotated vehicle stops: 113
- Total number of turns: 4245

Classifying 21 Dialog Acts

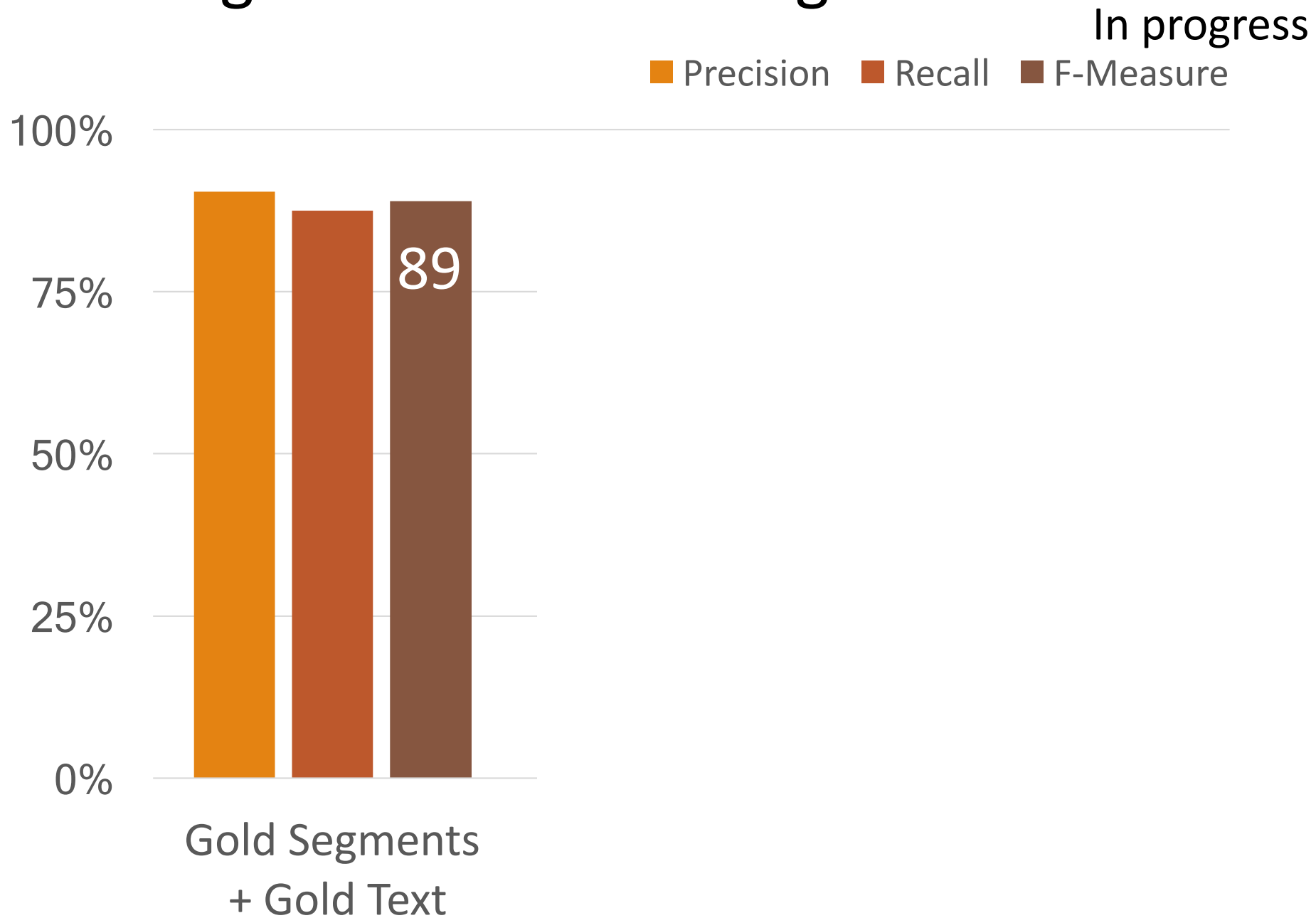
Features:


- position in discourse, length of utterance
- neural embeddings, n-grams, lexicons, regular expressions
- dependency parse features
- unsupervised topic assignments

Classification Algorithm:

- Linear SVM one-vs-rest multi-class classification
 - (Convolutional nets, MLPs, CRFs, all about the same)

Detecting Institutional Dialog Acts






Now run this classifier on all 900
vehicle stops

Are conversations with black drivers
structured differently than with
white drivers?

Preliminary results



Blacks are told the **reason** for
the stop later than whites

After asking for a license instead of
right away

Preliminary results



Lenience: Blacks are twice as likely to be told they are getting off easy

"I'm doing you a favor."

"Oh, I'm giving you a big break today."

"I'm going to let you off with a, a warning today."

"I'll tell you what, we'll let you, uh, slide on that, all right?"

Implying they actually deserve more serious punishments


Preliminary results



Even though police are not actually more lenient to blacks:

Blacks and whites are equally likely to be let off with a warning

And recall that blacks are stopped for less serious offenses



Blacks more likely to be asked
if this is their car

"Is this your car, boss?"

"Does the car belong to you?"

"And uh, is this your car registered to
you, sir?"

"So who does this- does this car
belong to you?"

Preliminary results

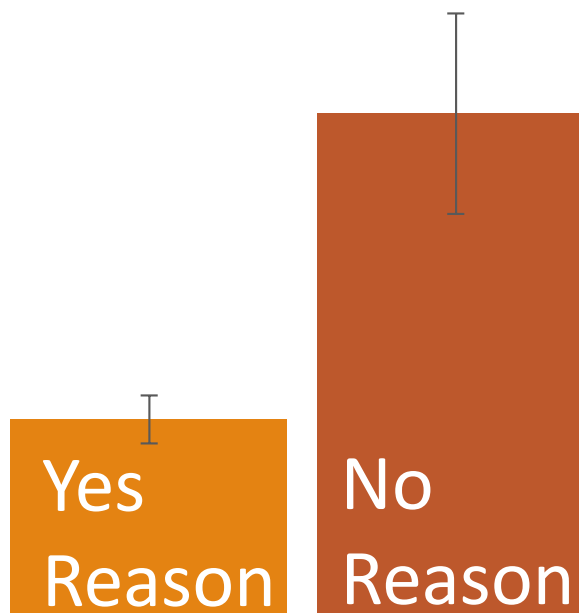
Could presence/absence of events have implications for escalation?

0.2

Driver **Anger** words

0.1

0



Do police give REASON in first 10 turns?

Preliminary results

Conclusion

Black community members experience very different police conversations than whites

Whites are more likely:

- To be told the reason for the stop
- To have the officer express concern for their safety

Blacks are more likely

- To be asked if this is their car
- To be told that the officer is "cutting them a break"

These differences may lead to more driver anger and negative emotion

- Important implications for escalation and safety

Moving toward scalability

Can we do this from raw speech?



Hang
Su



Prateek
Verma



Vinod
Prabhakaran



Nelson
Morgan



Jennifer
Eberhardt

Collaboration with ICSI Berkeley

Can we do this task without hiring human transcribers?

Required if there is any chance of doing such studies across the country

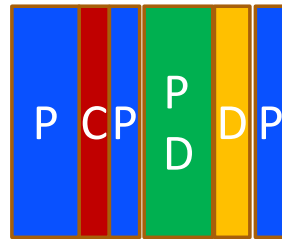
Dialog Acts from raw speech (work in progress)

Vinodkumar Prabhakaran, Camilla Griffiths, Hang Su, Prateek Verma, Nelson Morgan, Jennifer Eberhardt, and Dan Jurafsky. 2018. Detecting Institutional Dialog Acts in Police Traffic Stops. Proceedings of the Transactions of Association for Computational Linguistics (TACL)

We built modern deep bi-directional neural speech recognition systems

- Trained on police data
 - Plus renoised Switchboard
 - Plus data augmentation (vocal tract and frame shift)

Conversational Event Detection

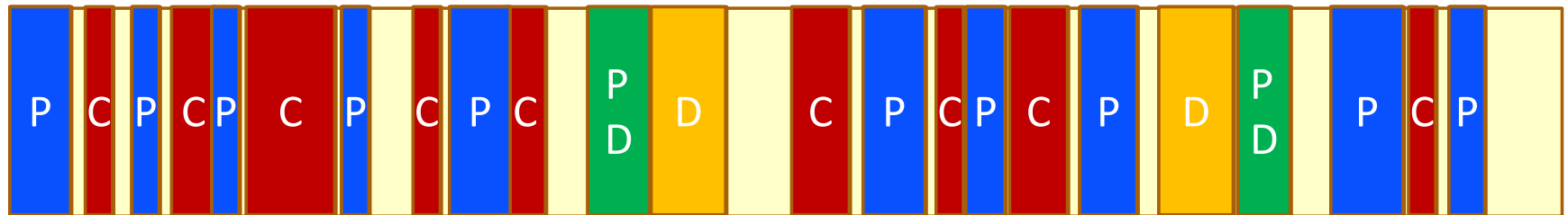


- N-Best List**
- Drive safely
 - Drive safe
 - Drive a safe
 - Drive safety



SAFETY

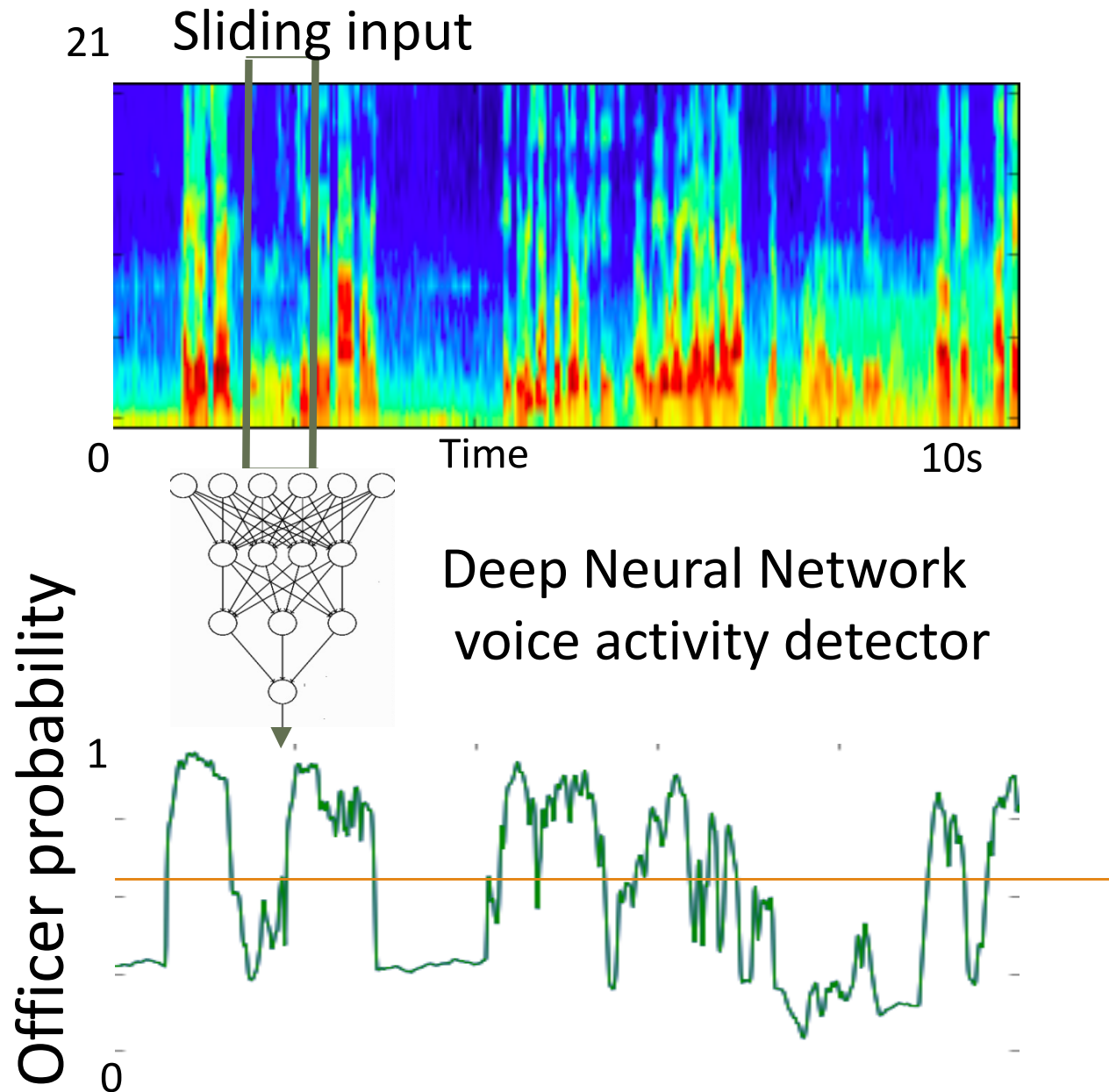
Diarization: Who's talking?



Various detectors for each kind of speech

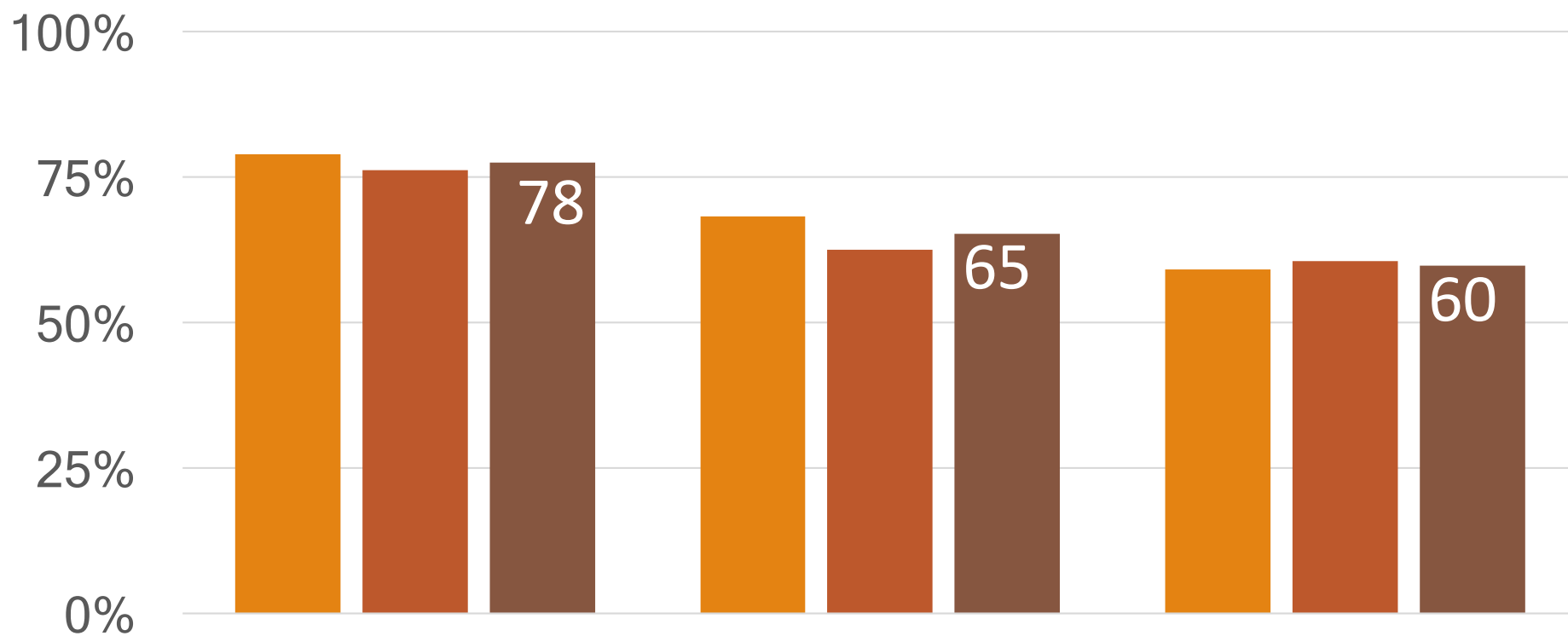
- Noise
- Police versus community member
- Dispatch
- Police speech to dispatch

Diarization: Police voice activity



Which Dialog Act occurs in this single turn? (work in progress)

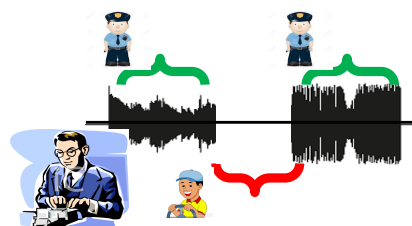
■ Precision ■ Recall ■ F-Measure



Gold Segments
+ Gold Text



Gold Segments
+ ASR Text



Predicted
Segments
+ ASR Text





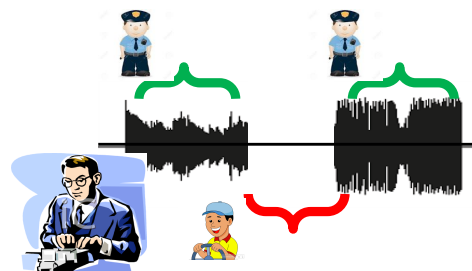
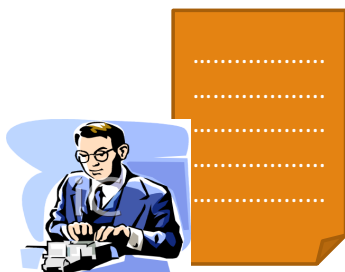
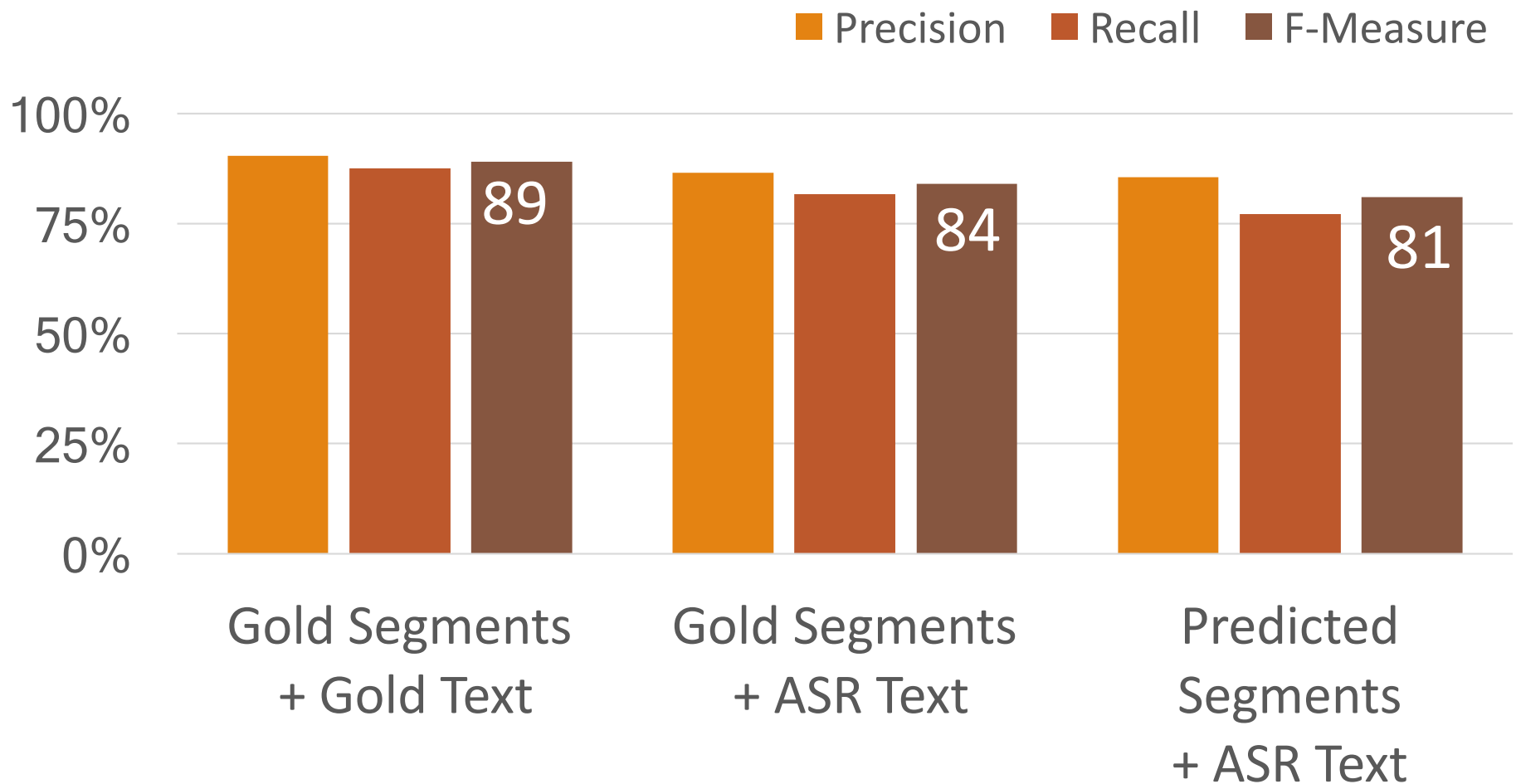
We may not need to detect **all** the institutional acts in a stop.

We may just want to know if a specific act happened or not

- Did the officer give a reason for a stop?
- Was the community member asked intrusive questions?

Does a Dialog Act occur anywhere in a stop?

Binary classification (work in progress)



Work in progress going forward

Extending our work on detecting anger and negative emotion in the **speech of the community members**

- Disrespect
- Compliance
- Anxiety
- Escalation
- For this task we have video!!!!

Work in progress

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

- Together with Oakland police staff
 - We developed training materials based on procedural justice
 - Using examples of "good" and "bad" stop interactions
- This summer: look at results before and after training
- Does training improve officer-community interaction?





Work in progress

- Adding data from more police departments
- Effect on police-community interaction of traumatic shootings
- Linguistics of prosody
- Better speech recognition and diarization

Conclusions

The first automated NLP-based analysis of police body camera footage

- Confirms reports about disparate treatment of black Americans
- Will (we hope) allow us to measure and improve officer training

Extracting Social Meaning from Language

Police language is one kind of social meaning

Other kinds of social meaning we work on:

- Political language
 - Framing of immigrants or minorities
 - Agenda-setting in government-controlled media
- Toxic speech
 - Reddit communities attacking each other
 - Gendered condescension in comments
- Schizophrenia diagnosis from interviews



A short taste of our work on
another social language area

Computational linguistics applied to
linguistic and cultural change

Relational models of word meaning in linguistics and cognitive science

Ferdinand de Saussure

Signs are defined by their **relationship** with each other

Ludwig Wittgenstein (PI #43):

"The meaning of a word is its **use** in the language"

Zellig Harris (1954):

"If A and B have almost identical **environments** we say that they are synonyms"

Osgood, Suci, Tanenbaum (1957)

Meaning as a continuous, dimensioned Euclidean semantic space defined by orthogonal dimensions

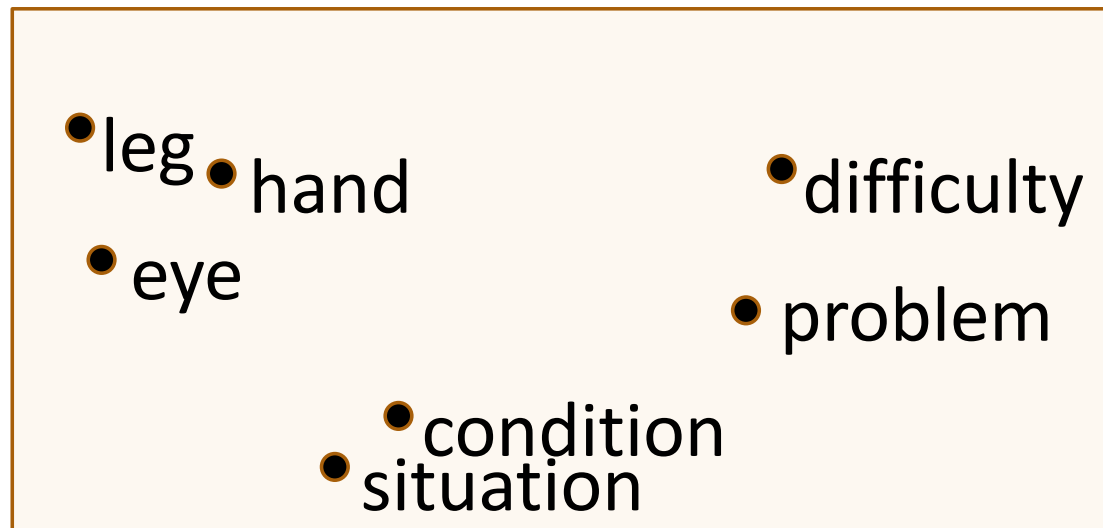
Modern relational model of meaning

Focus on similarity

Each word represented by a vector

- a list of numbers =
- a point in space

Similar words are "nearby in space"



Distributional semantics: Define a word as a vector

A vector is called an "embedding" because it's embedded into a space

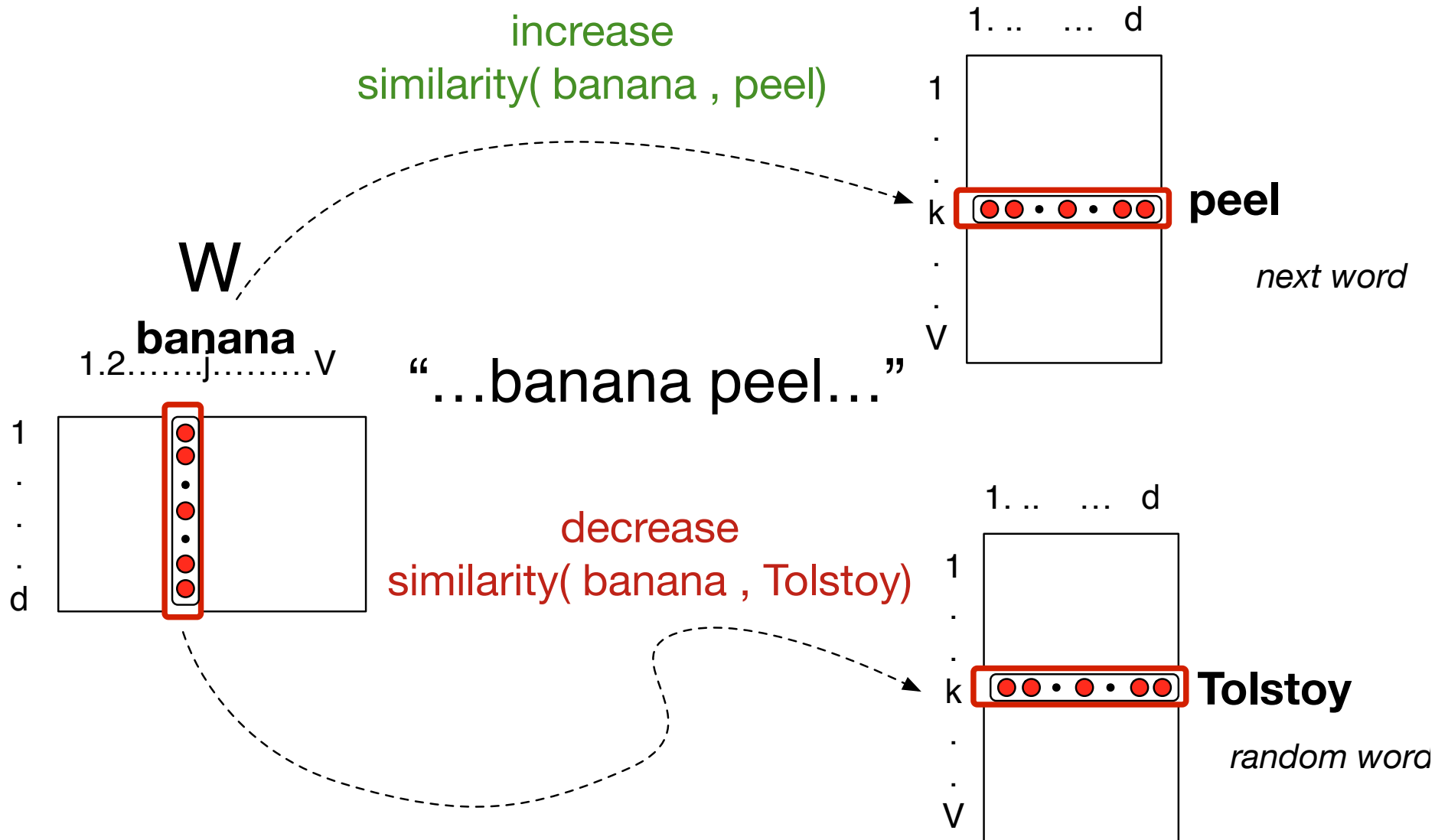
Common:

- Word2Vec (Mikolov et al. 2013)
- 300-dimensional vector

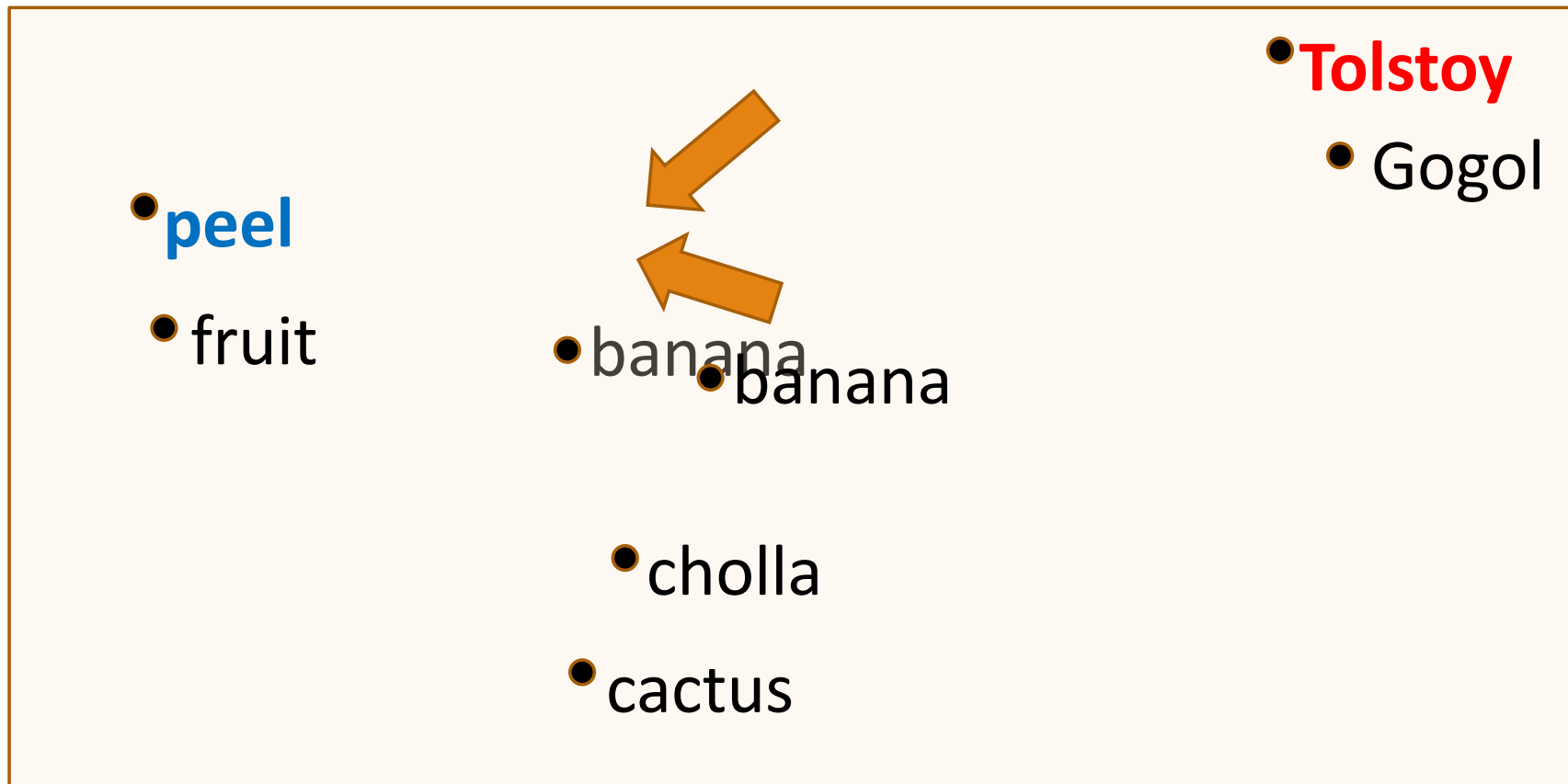
Vectors are learned iteratively


- Make vectors for a word look **like** the vectors for its neighbor
- Make vectors for a word look **different** than vectors for other words

Learning embeddings: seeing "banana peel"



Iteratively learning embeddings





Embeddings are the fundamental way to represent words in computational linguistics

- Parsing
- Thematic role labeling
- Coreference Resolution
- Word sense disambiguation
- Machine Translation
- Question Answering
- Summarization



Can embeddings help us test theories of linguistic change?

Towards a Computational Historical Semantics

William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. Proceedings of ACL 2016.

William L. Hamilton, Jure Leskovec, Dan Jurafsky. 2016. Cultural Shift or Linguistic Drift? Comparing Two Computational Models of Semantic Change. Proceedings of EMNLP 2016.

William L. Hamilton, Kevin Clark, Jure Leskovec, Dan Jurafsky. 2016. Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. Proceedings of EMNLP 2016.



Will Hamilton



Jure Leskovec



Testing theories of semantic change

The role of frequency and polysemy

Semantic bleaching

Increasing subjectification over time
(Traugott and Dasher 1992)

Semantic differentiation (Bréal 1897)

Role of frequency in change?

Frequent words change **faster**

- Lenition (phonetic reduction) happens in frequent words

Frequent words change **slower**

- High frequency words are more resistant to morphological regularization
 - Bybee, 2007; Pagel et al., 2007; etc.)

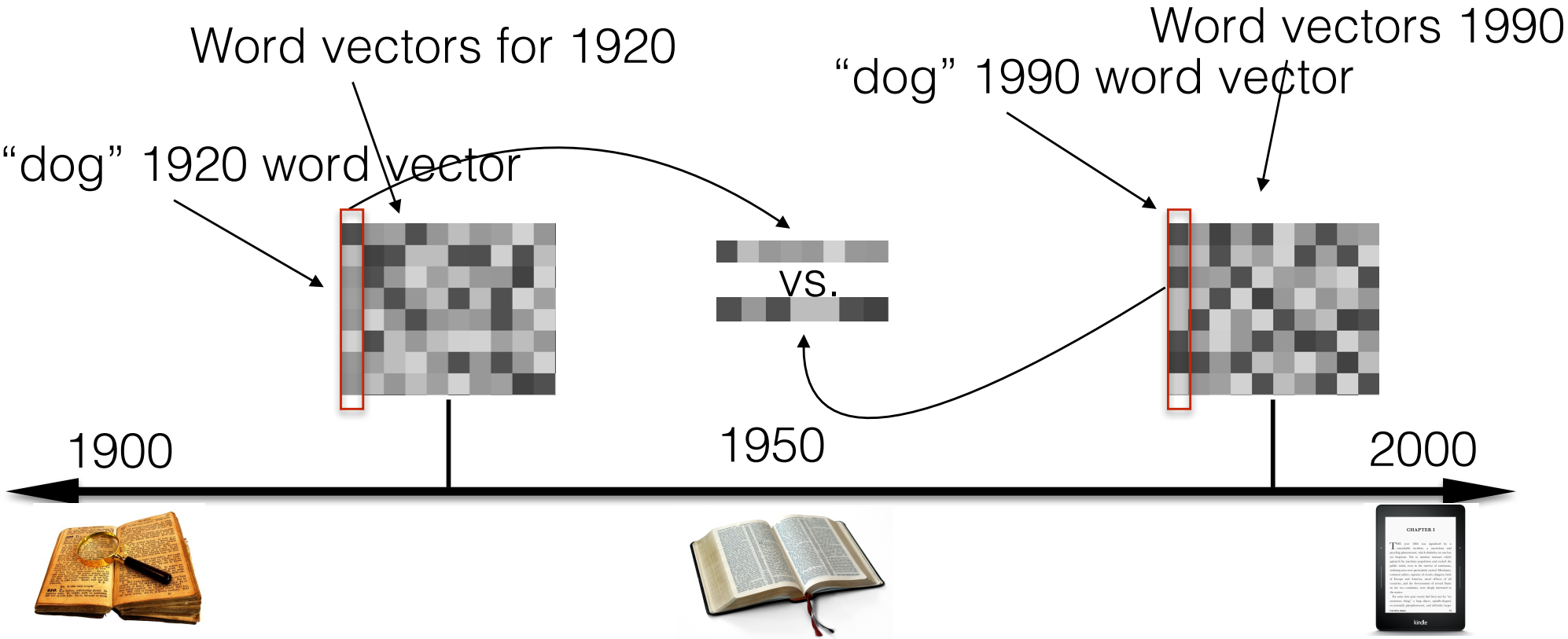
Role of polysemy/homonymy in change

The number of senses a word has

- Bank (1) sloping land (2) financial institution
- Words gain senses as they drift (Bréal, 1897; Wilkins, 1993; Hopper and Traugott, 2003)
- Polysemous words occur in more diverse contexts, affecting lexical access speed (Adelman et al., 2006) and rates of L2 learning (Crossley et al., 2010).

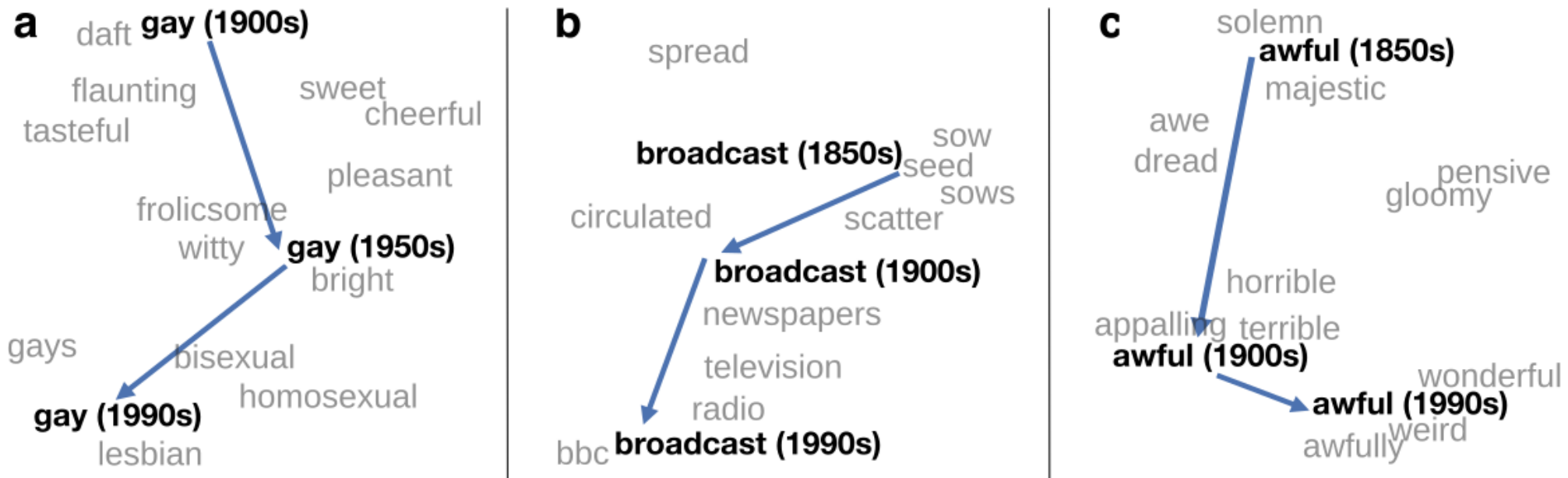
But does that make them faster or slower to change?

Diachronic word embeddings for studying language change!



Visualizing changes in meaning

Project 300 dimensions down into 2



~30 million books, 1850-1990, Google Books data

Statistical laws of semantic change

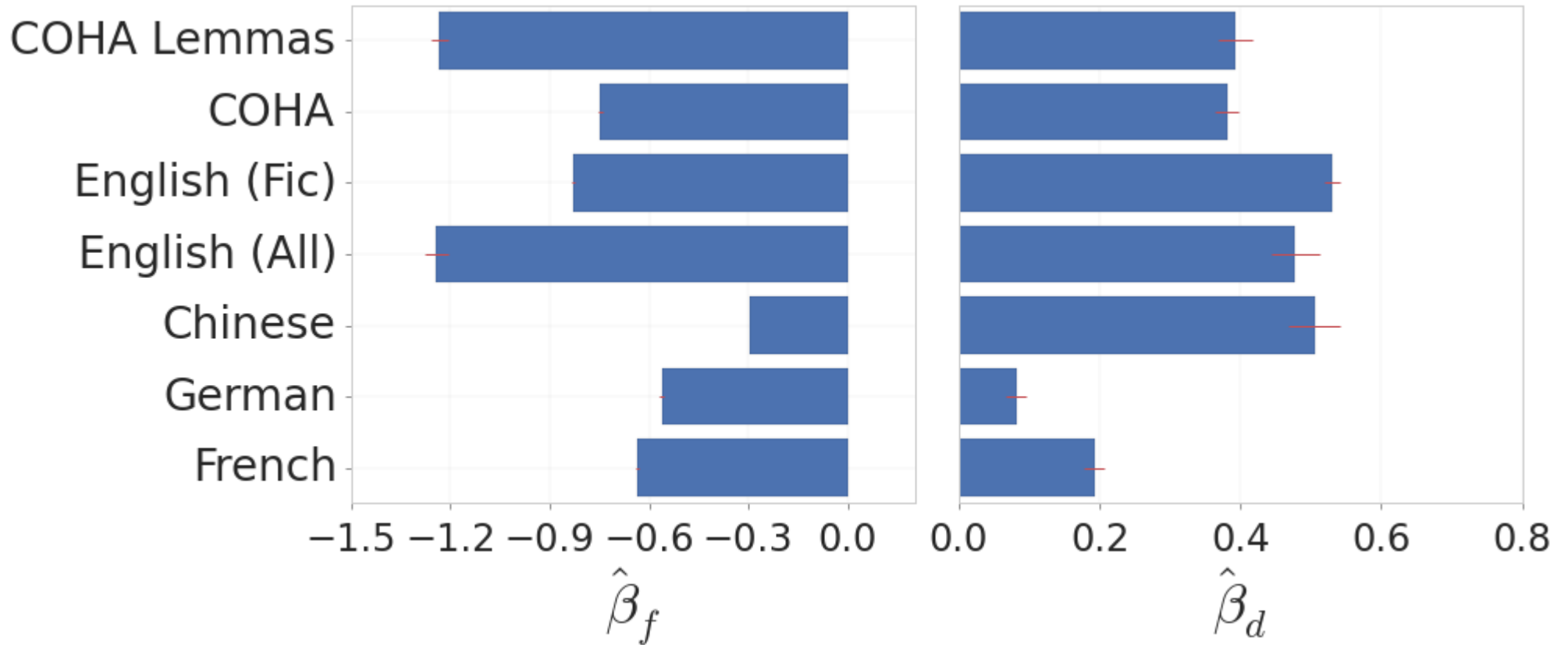
$$\Delta(w_i) \propto f(w_i)^{\beta_f^{<0}} \times d(w_i)^{\beta_d^{>0}}$$

Rate of semantic
change

Frequency

Polysemy score

Results across languages

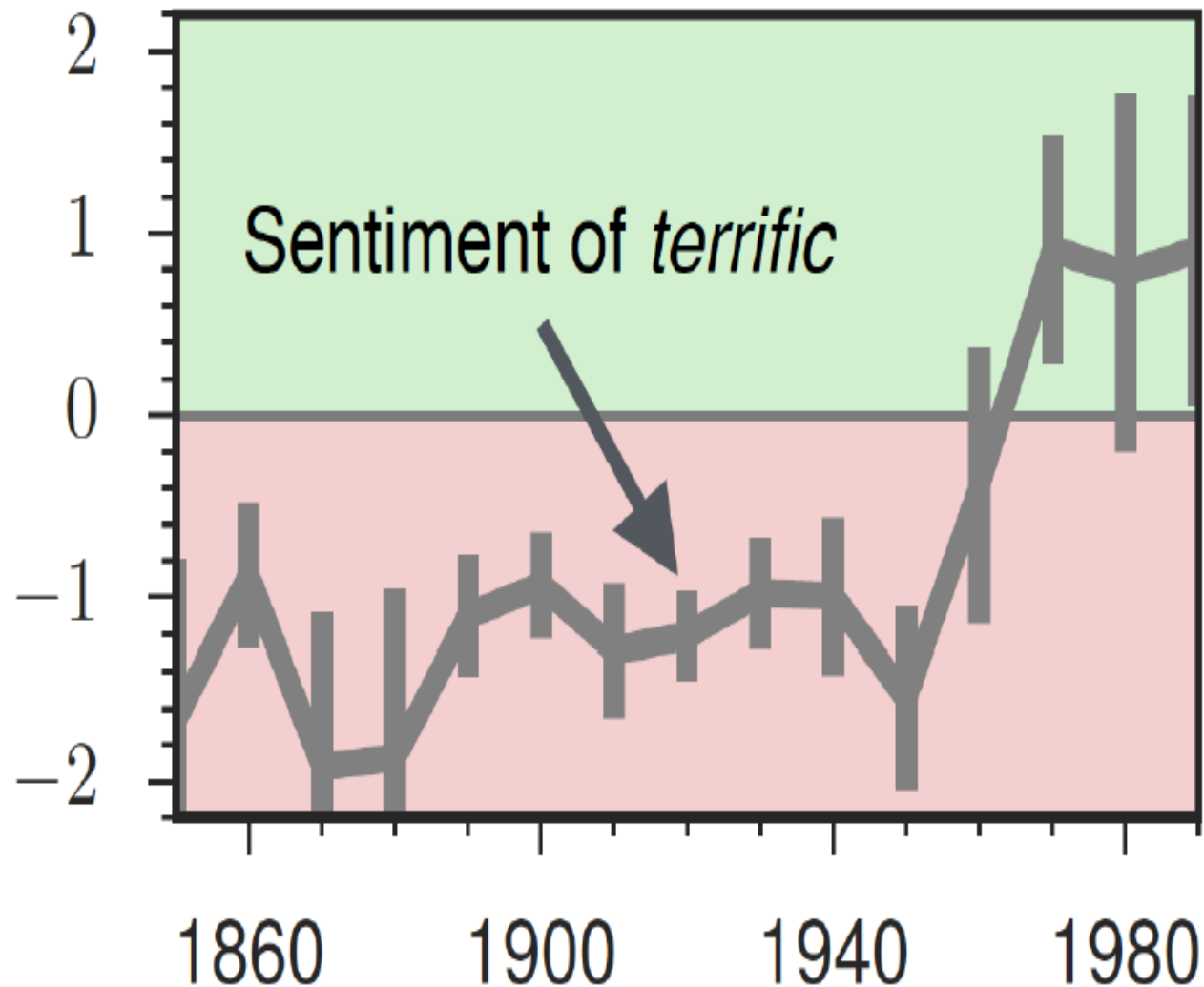


Effect of frequency
(consistently negative)

Effect of polysemy
(consistently positive)

The evolution of connotation

Negative sentiment words change faster than positive words



Embeddings reflect cultural bias

Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. 2016. "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings." In *NIPS*. 4349-4357.

Ask "Paris : France :: Tokyo : x"

- x = Japan

Ask "father : doctor :: mother : x"

- x = nurse

Ask "man : computer programmer :: woman : x"

- x = homemaker

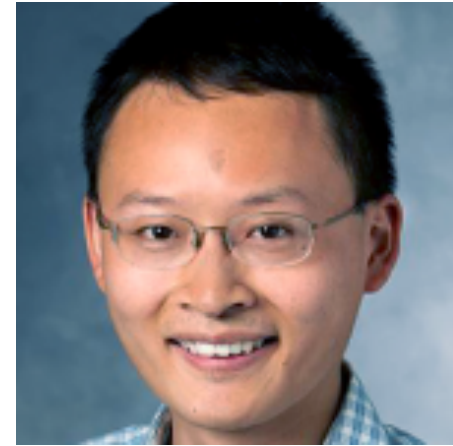
Historical embedding: a tool to investigate history of cultural biases



Nikhil Garg



Londa Schiebinger



James Zou

Nikhil Garg, Londa Schiebinger, Dan Jurafsky, James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. Proceedings of the National Academy of Sciences 2018.

Historical embedding: a tool to investigate cultural biases

Take the historical embeddings from the previous paper

Compute historical biases of words:

- Gender bias: how much closer a word is to "woman" synonyms than "man" synonyms.
- Ethnic bias: how much closer a word is to last names of a given ethnicity than to names of Anglo ethnicity
- Correlate with occupational data from historical census

Look at how all these change over time

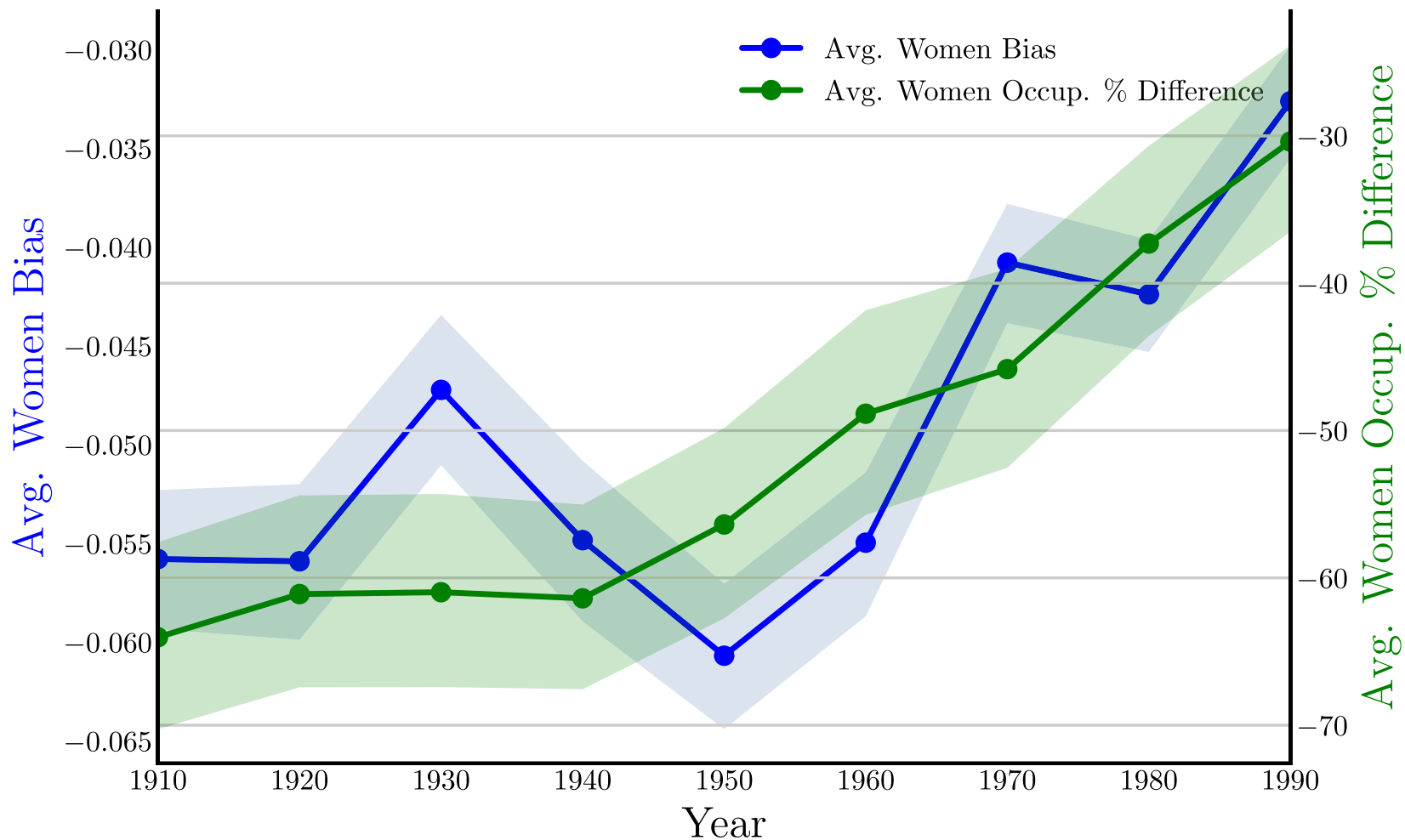
Historical embedding: a tool to investigate cultural biases

Is the word "nurse" or "carpenter" closer to the word "man" or "woman"?

Embedding bias reflects actual gender differences in occupations



Embeddings reflects gender bias in occupations across time (1910-1990)



Embeddings reflect framings of women over time

Embeddings for competence adjectives are biased toward men

- *Smart, wise, brilliant, intelligent, resourceful, thoughtful, logical, etc.*

This bias is slowly decreasing 1960-1990

If rate continues, should be equally associated with women in 10 years.

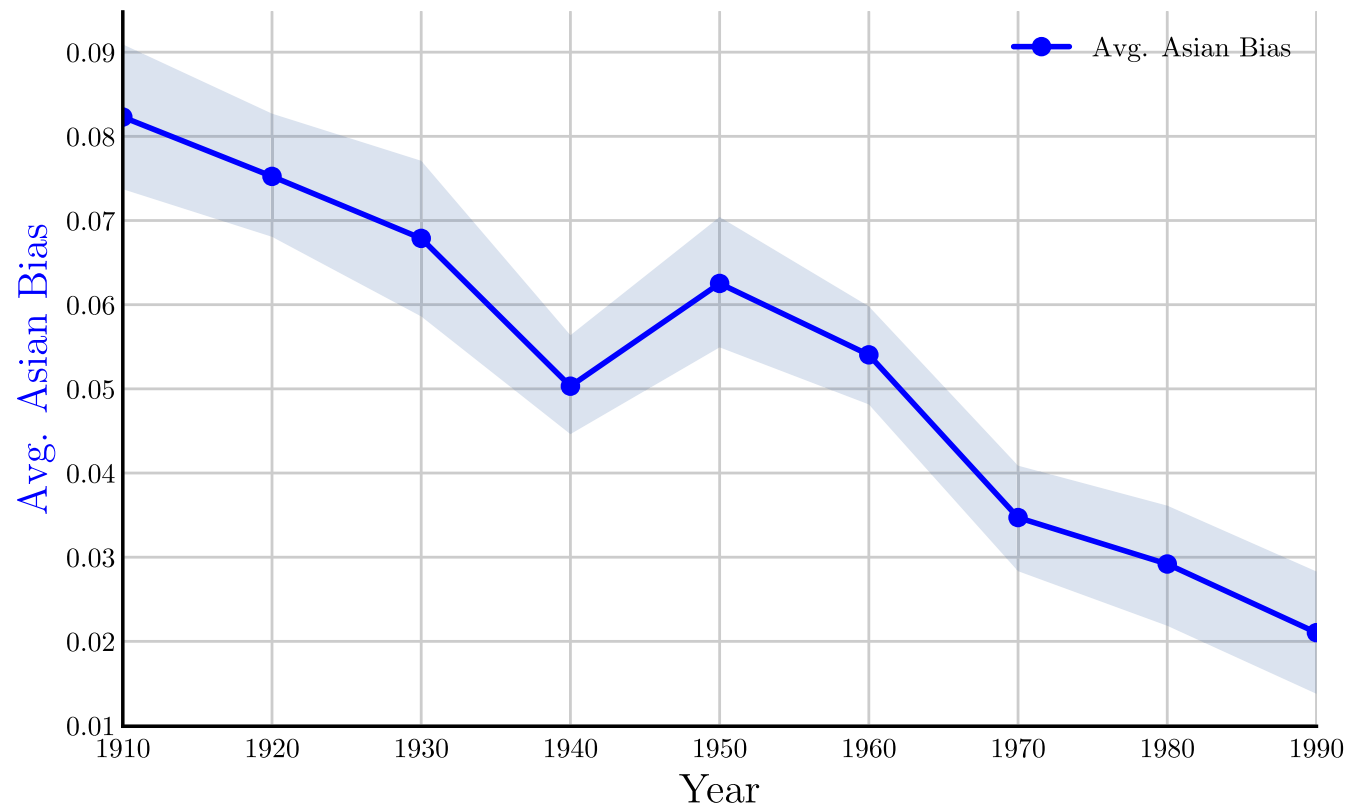
Just one aspect of framing

Embeddings reflect ethnic stereotypes over time

- Princeton trilogy experiments
- Attitudes toward ethnic groups (1933, 1951, 1969) scores for adjectives
 - *industrious, superstitious, nationalistic, etc*
- Embedding bias (Chinese vs White) correlates with adjective scores and with the change 1933-1979

Change in linguistic framing 1910-1990

Change in association of Asian names with adjectives framed as "othering" (*barbaric, monstrous, bizarre*)



The most biased Asian (vs. White) adjectives over time

1910	1950	1990
Irresponsible	Disorganized	Inhibited
Envious	Outrageous	Passive
Barbaric	Pompous	Dissolute
Aggressive	Unstable	Haughty
Transparent	Effeminate	Complacent
Monstrous	Unprincipled	Forceful
Hateful	Venomous	Fixed
Cruel	Disobedient	Active
Greedy	Predatory	Sensitive
Bizarre	Boisterous	Hearty



Work in progress: Computationally induce framings over time

- Looking at news media 1850-2000
- How ethnic groups are viewed
- How immigration is viewed

Conclusion

Embeddings are fine-grained relational models of (some aspects of) word meaning

Allow us to test linguistic theories of semantic change

Also a tool for quantifying cultural biases and framings and their changes over time

"Running experiments in the past"

Social variables are important for computational linguistics!

- Demographic characteristics:
 - Race, gender, ethnicity
- Social relations
 - Power, respect
- Affect
- Diachrony and temporal context

The common misconception is that language has to do with **words** and what they mean.

It doesn't.

It has to do with **people** and what *they* mean.



Herbert H. Clark & Michael F. Schober, 1992