Language from Police Body Camera Footage Shows Racial Disparities In Officer Respect

Voigt et al.

CS 224C Presentation:

Rhea Kapur, Chijioke Mgbahurike, Anna Saraiva

An Untapped Resource

- Despite the proliferation of body-worn cameras in law enforcement

An Untapped Resource

- Despite the proliferation of body-worn cameras in law enforcement
- **Lack of analyzation by law enforcement**

An Untapped Resource

- Despite the proliferation of body-worn cameras in law enforcement
- **Lack of analyzation by law enforcement**
- **Concentration on high profile engagements**

How do everyday interactions between police and community members differ in respect to black vs white community members?

Data

- Transcribed body camera footage
- Routine vehicle stops (N = 981)
- White (N=299) Black (N=682)
- Oakland Police Department
- April 2014

Novelty

- Datasource

Novelty

- Datasource
- Systematic analysis

Novelty

- Datasource
- Systematic analysis
    - 3 tiered study
    -

Novelty

- Datasource
- Systematic analysis
  - 3 tiered study
  - Tied to theories in sociolinguistics

Can humans **reliably** judge officers' respect from language

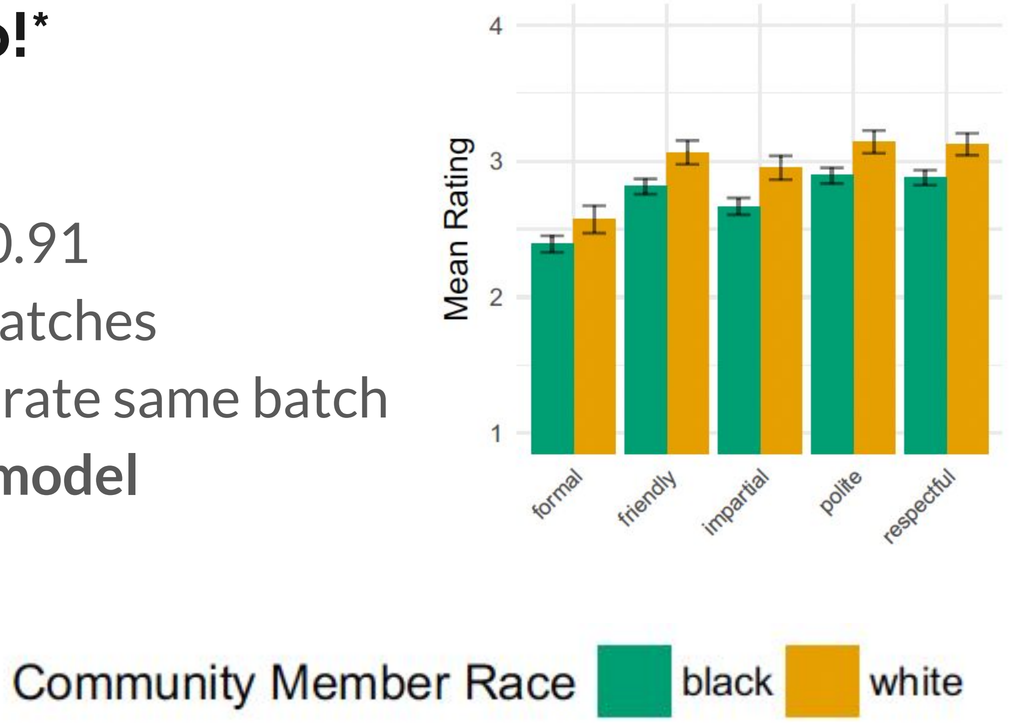
Does judged respect **differ** against white vs black comm. members

They Can! And They Do!

- 414 unique randomly sampled officer utterances:
  - 4 point likert scale:
    - Respectful
    - Polite
    - Friendly
    - Formal
    - Impartial

They Can! And They Do!*

- Annotator Consistency
  - Cronbach's ***α***= 0.73 - 0.91
  - Group utterances in batches
  - Same 10+ annotators rate same batch
  - **Linear mixed effects model**

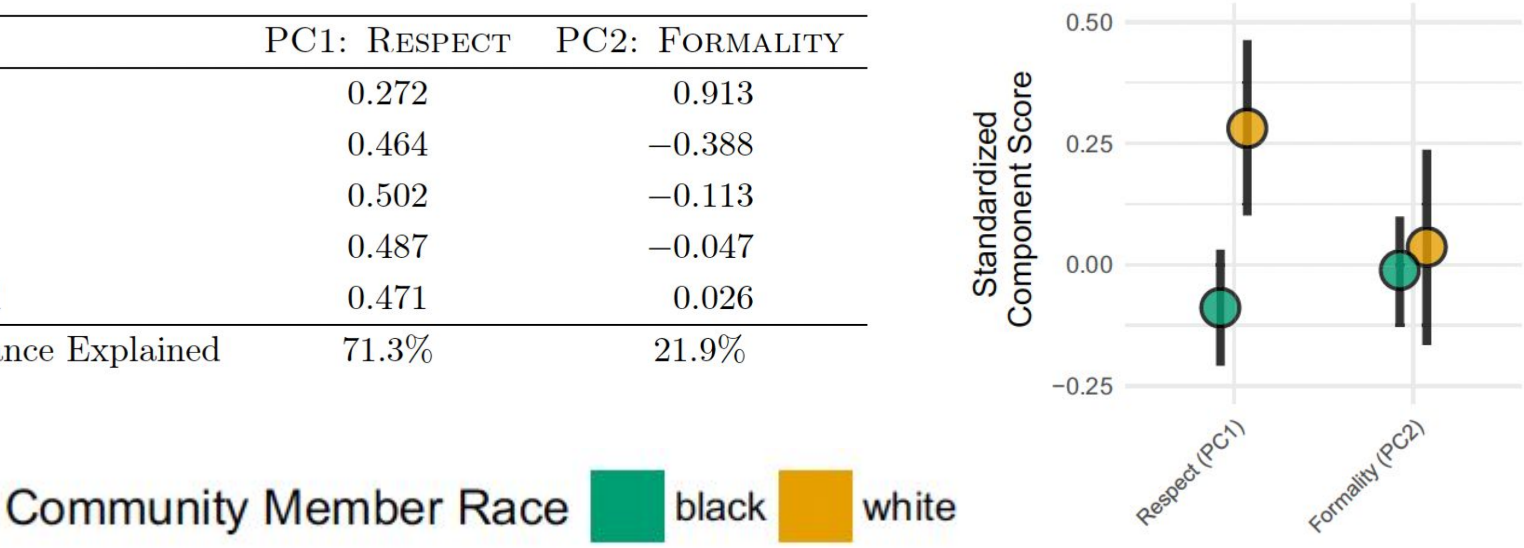


	Respectful			*Polite*			*Impartial*			*Friendly*			*Formal*		
	b	CI	p	b	CI	p	b	CI	p	b	CI	p	b	CI	p
Fixed Parts															
Intercept	2.94	2.83 – 3.04	**<.001**	2.95	2.85 – 3.06	**<.001**	2.69	2.57 – 2.80	**<.001**	2.85	2.74 – 2.96	**<.001**	2.49	2.37 – 2.61	**<.001**
Driver Age	0.03	-0.02 – 0.08	.22	0.01	-0.04 – 0.07	.59	0.01	-0.05 – 0.07	.75	0.00	-0.05 – 0.05	1.00	0.08	0.02 – 0.14	**.01**
Driver Gender (F)	0.04	-0.07 – 0.16	.42	0.05	-0.07 – 0.16	.42	-0.0!	-0.13 – 0.12	.92	0.02	-0.10 – 0.14	.72	0.09	-0.04 – 0.22	.18
Driver Race (B)	-0.22	-0.33 – 0.10	**<.001**	-0.22	-0.34 – -0.11	**<.001**	-0.26	-0.39 – -0.13	**<.001**	-0.23	-0.36 – -0.11	**<.001**	-0.14	-0.28 – 0.01	**.04**
Random Parts															
σ^2		0.17			0.19			0.21			0.22			0.25	
$\tau_{00,\text{Stop}}$		0.05			0.04			0.07			0.05			0.06	
N_{Stop}		251			251			251			251			251	
ICC_{Stop}		0.22			0.19			0.24			0.17			0.18	
Observations		414			414			414			414			414	
R^2 / Ω_0^2		.52 / .39			.48 / .35			.56 / .42			.47 / .33			.47 / .34	

Table 5: Linear mixed-effects models results for judgements in Study 1.

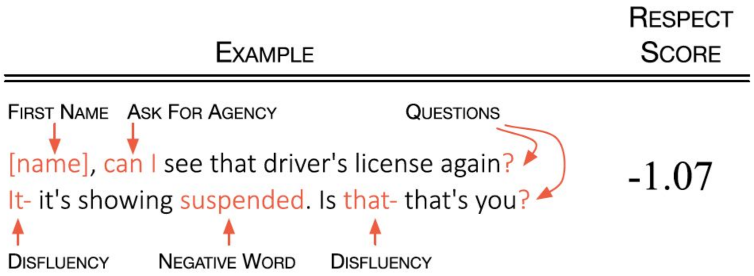
PCA: Please Confirm Assumptions

	PC1: Respect	PC2: Formality
Formal	0.272	0.913
Friendly	0.464	−0.388
Impartial	0.502	−0.113
Polite	0.487	−0.047
Respectful	0.471	0.026
% of Variance Explained	71.3%	21.9%



Can we **model** it?

First Names	Top 1000 most common first names from the 1990 US Census, where first letter is capitalized in transcript
Formal Titles	Lexicon: "sir", "ma'am", "maam", "mister", "mr*", "ms*", "madam", "miss", "gentleman", "lady"
For Me	Lexicon: "for me"
For You	Lexicon: "for you"
Give Agency	Lexicon: "let you", "allow you", "you can", "you may", "you could"
Gratitude	Lexicon: "thank", "thanks", "appreciate"
Goodbye	Lexicon: "goodbye", "bye", "see you later"
Hands on the Wheel	Regular expression capturing cases like "keep your hands on the wheel" and "leave your hands where I can see them": "hands? ([^.,?!:;]+)?(wheel|see)"
Hedges	All words in the "Tentat" LIWC lexicon
Impersonal Pronoun	All words in the "Imppron" LIWC lexicon
Informal Titles	Lexicon: "dude*", "bro*", "boss", "bud", "buddy", "champ", "man", "guy*", "guy", "brotha", "sista", "son", "sonny", "chief"
Introductions	Regular expression capturing cases like "I'm Officer [name] from the OPD" and "How's it going?": "((i|my name).+officer | officer.+(oakland|opd))|((hi|hello|hey|good afternoon|good morning|good evening|how are you doing|how 's it going))"
Last Names	Top 5000 most common last names from the 1990 US Census, where first letter is capitalized in transcript
Linguistic Negation	All words in the "Negate" LIWC lexicon
Negative Words	All words in the "Negativ" category in the Harvard General Inquierer, matching on word lemmas
Positive Words	All words in the "Positiv" category in the Harvard General Inquierer, matching on word lemmas



A Reliable Model...

Root Mean Squared Error:

- Respect - Model: 0.840, Human: 0.842
- Formality - Model: 0.882, Human: 0.764
- 414 sampled utterance

What doesn't **matter?**

Showed No Difference

Showed No Difference

Officer
Race

Showed No Difference

Officer Race

Geographic Info

Showed No Difference

Officer Race

Geographic Info

Number of Officers

Showed No Difference

Officer Race

Geographic Info

Race & Formality

Number of Officers

Showed No Difference

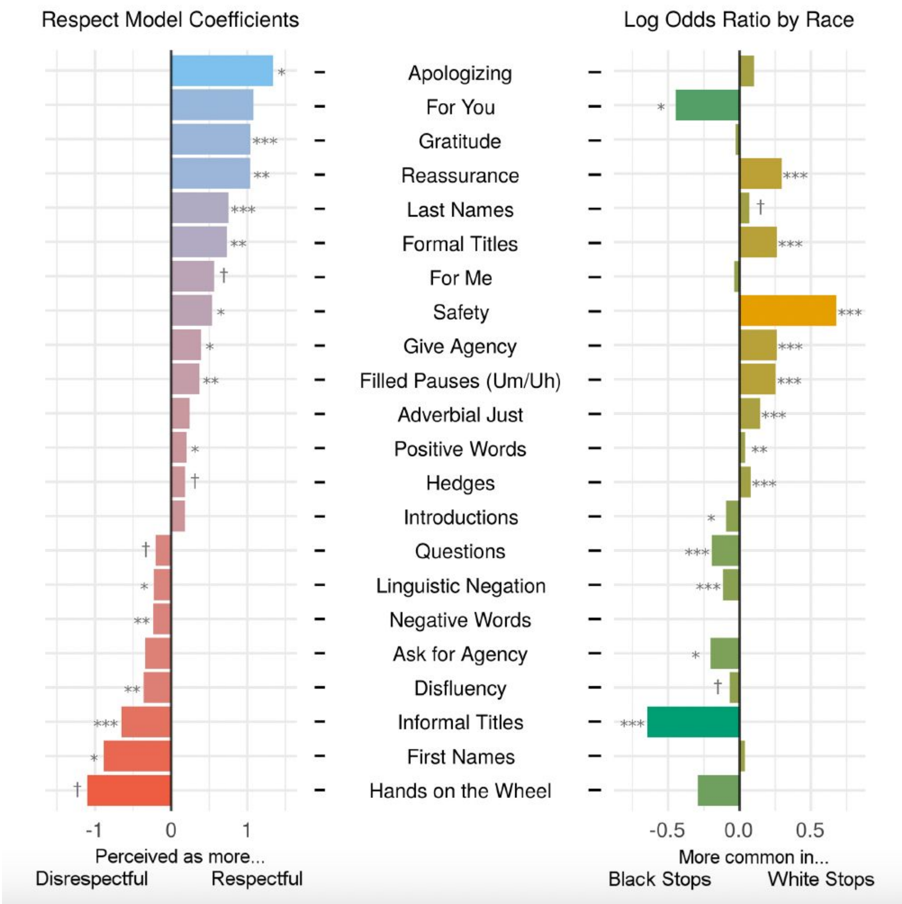
Offense Severity

Officer Race

Geographic Info

Race & Formality

Number of Officers



Respect Model Coefficients
Log Odds Ratio by Race
Apologizing
For You
Gratitude
Reassurance
Last Names
Formal Titles
For Me
Safety
Give Agency
Filled Pauses (Um/Uh)
Adverbial Just
Positive Words
Hedges
Introductions
Questions
Linguistic Negation
Negative Words
Ask for Agency
Disfluency
Informal Titles
First Names
Hands on the Wheel
-1
0
1
Perceived as more...
Disrespectful
Respectful
-0.5
0.0
0.5
More common in...
Black Stops
White Stops

To **recap...**

To recap...

- Scalable and sensitive technique for body-cam interactions
- Consistent judgement from police-community interactions
- Respect & Formality Axis
- Positive & negative strategies for politeness
- Racial disparities in respect for Black community members

Peer Reviewer

Review & Expand

Strengths:

- Use of human participants as well as language models
- Test “thin slices” approach before implementing it in scale
- Analysis of evolution of disparities as interaction time passed

Review & Expand

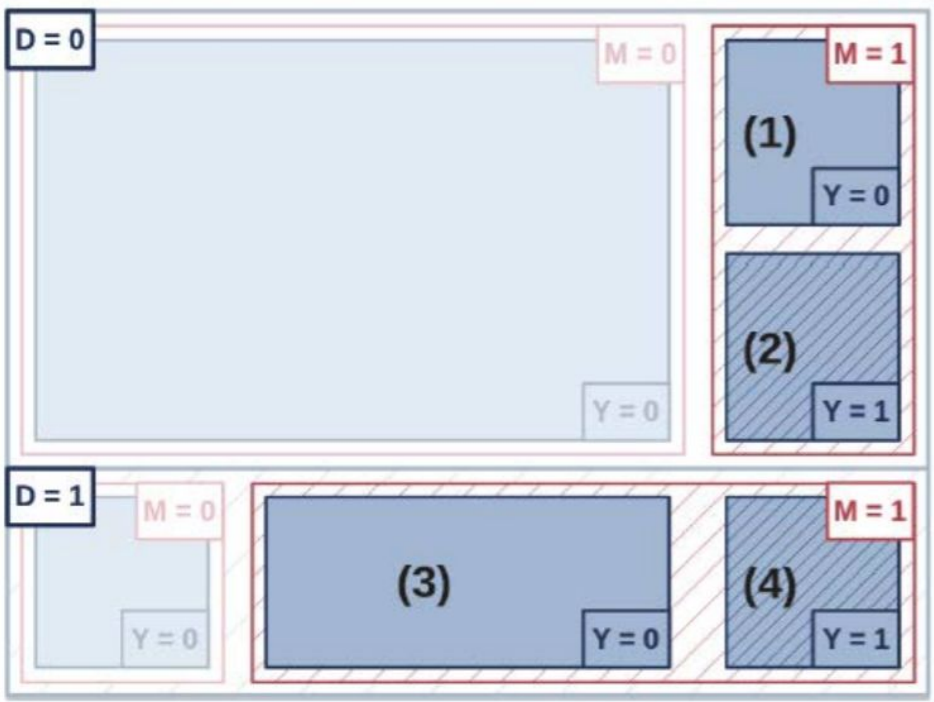
Strengths:

- Use of human participants as well as language models
- Test "thin slices" approach before implementing it in scale
- Analysis of evolution of disparities as interaction time passed

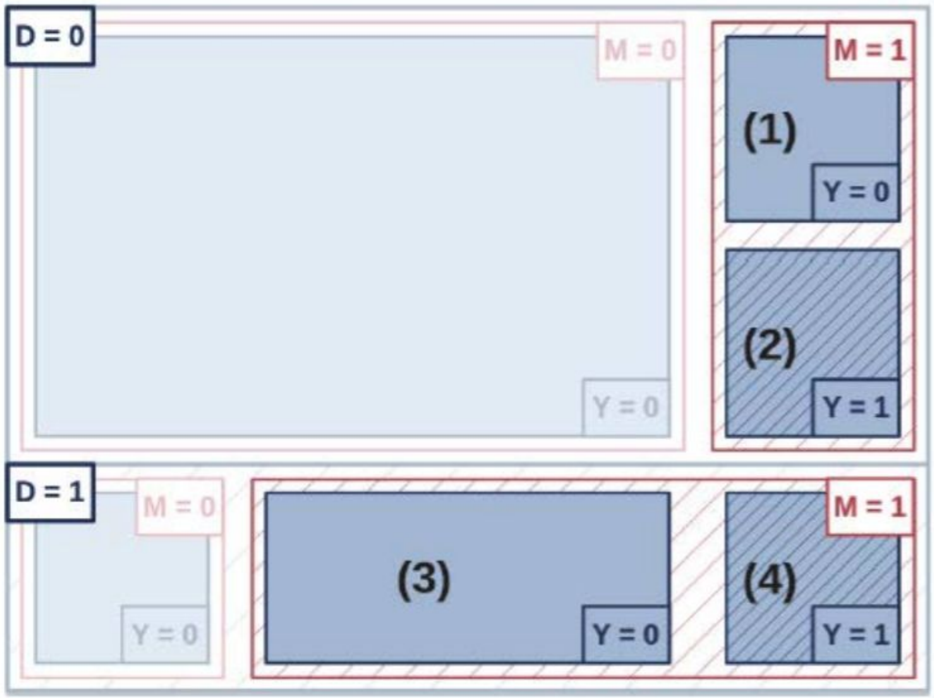
Weaknesses:

- The problem with analyzing stops: **search *vs* hit rates** (similar problems with Fryer (2019))
- Camera footage doesn't give the full picture
- Controlling for potentially **endogenous** variables
- Very limited **external validity**, focusing on one city in one month

Search Rate *vs* Hit Rate



Search Rate *vs* Hit Rate



Racial bias is present if

$$\frac{(4)}{(3)+(4)} > \frac{(2)}{(1)+(2)}$$

But denominators themselves might be endogenous and subject to bias

Review & Expand

Questions:

- Were there checks for accuracy of transcriptions?
- Besides Respect and Formality, what were the differential rates in specific phrases, such as "*hands on wheel*"?

Review & Expand

Questions:

- Were there checks for accuracy of transcriptions?
- Besides Respect and Formality, what were the differential rates in specific phrases, such as "*hands on wheel*"?

Future Research:

- Transcription done at scale
- Expand to other cities and time periods
- Additional behavior derived from camera footage

NYPD
Industry Practitioner

Developing an Internal Officer Feedback System at the NYPD

- Let’s quantify officer respectfulness *within our city and police department*

Developing an Internal Officer Feedback System at the NYPD

- Let’s quantify officer respectfulness *within our city and police department*
  - Replicate aspects of the study to analyze **respectfulness on an overall and per-officer basis**

Developing an Internal Officer Feedback System at the NYPD

- Let's quantify officer respectfulness *within our city and police department*
    - Replicate aspects of the study to analyze **respectfulness on an overall and per-officer basis**
    - **Extend to include speech data and facial expression analysis (Field et al. 2023)**

Developing an Internal Officer Feedback System at the NYPD

- Let's quantify officer respectfulness *within our city and police department*
  - Replicate aspects of the study to analyze **respectfulness on an overall and per-officer basis**
  - **Extend to include speech data and facial expression analysis (Field et al. 2023)**
- **Online dashboard analyzing each bodycam interaction as it comes in live**, with overall respectfulness evaluation and factor breakdown (following the paper)

Developing an Internal Officer Feedback System at the NYPD

- Let's quantify officer respectfulness *within our city and police department*
  - Replicate aspects of the study to analyze **respectfulness on an overall and per-officer basis**
  - **Extend to include speech data and facial expression analysis (Field et al. 2023)**
- **Online dashboard analyzing each bodycam interaction as it comes in live**, with overall respectfulness evaluation and factor breakdown (following the paper)
  - Officers can then review this on their own time

Developing an Internal Officer Feedback System at the NYPD

- Let's quantify officer respectfulness *within our city and police department*
  - Replicate aspects of the study to analyze **respectfulness on an overall and per-officer basis**
  - **Extend to include speech data and facial expression analysis (Field et al. 2023)**
- **Online dashboard analyzing each bodycam interaction as it comes in live**, with overall respectfulness evaluation and factor breakdown (following the paper)
  - Officers can then review this on their own time
- Use this information in personalized officer training and review sessions

NYPD Academic Researcher

Extended Meta-Study Using Internal System Data

- Are general takeaways from Voigt et al. (ex. that interactions are less respectful when blacks are pulled over vs. whites) **replicated in *our specific police department***, in the city of New York?

Extended Meta-Study Using Internal System Data

- Are general takeaways from Voigt et al. (ex. that interactions are less respectful when blacks are pulled over vs. whites) **replicated in *our specific police department***, in the city of New York?
- How does the inclusion of **speech data, facial expression analysis, and prosodic information** change or strengthen the takeaways both in the broader context of Voigt et al. and also for our department?

Extended Meta-Study Using Internal System Data

- Are general takeaways from Voigt et al. (ex. that interactions are less respectful when blacks are pulled over vs. whites) **replicated in *our specific police department***, in the city of New York?
- How does the inclusion of **speech data, facial expression analysis, and prosodic information** change or strengthen the takeaways both in the broader context of Voigt et al. and also for our department?
- **If we look at officer characteristics as well** (race, age, gender, etc.) are there any patterns re. respectfulness?

Social Impact Assessor

Implications of the Paper

Concrete:

- Police-community interactions can be quantitatively assessed and categorized
- Robust benchmarks for police accountability
- Tools for police training

Abstract:

- What other sociolinguistic effect could be analyzed?
- Applicability to parallel industries, i.e. healthcare, education
- Feasible real world impact