RACIAL DISPARITIES FROM POLICE OFFICERS

JAIDAN GIGLIO

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

Authors: Rob Voigt, Nicholas P. Camp, Vinodkumar Prabhakaran, William L. Hamilton, Rebecca C. Hetey, Camilla M. Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L. Eberhardt

Published March 26, 2017

Significance

This study demonstrated the usage of body camera footage as a source of useful data when analyzed with computational linguistics. The results could potentially be used to improve the relationship between civilians and police officers.

Background

Data:

- 981 stops total (68% of total 1,440 stops during period)
  - 682 black drivers
  - 299 white drivers
  - stops conducted by 245 different officers
  - 183 hours footage from entire month of April 2014
  - 36,738 officer utterances

Prior Research:

- observed recollection of events, rather than footage
- limited to small number of interactions
- contains biases:
  - citizens/officers recollection may be inaccurate to what actually happened
  - officers might change their behavior if they know study is happening

Key Concepts:

- people often focus on specific instances, rather than all cop cam footage as a whole
- looked at body cam footage during traffic stops in Oakland, CA
- black people tend to report more negative experiences with police
  - less fair and more disrespectful
- officers choice of words have a heavy weight in how people with power are viewed

Research Questions

1. “Do officers treat white community members with a greater degree of respect than they afford to blacks?”

2. “Could participants reliably glean these qualities from such brief exchanges?”

Methodology

Tools used: every utterance was processed with Stanford CoreNLP 3.4.1 in order to generate sentence and word segmentation, part-of-speech tags, and dependency parses for feature extraction and analysis

CoreNLP

Study 1

414 unique officer utterances (1.1% of total)

- directed towards 312 black people and 102 white members
- participants viewed text of officer utterance, with driver utterance that came before it
- names anonymized, not told of race or gender

Participants rated officers speech on several overlapping dimensions of respect on a four-point bipolar scale. The dimensions included Very or Somewhat Impolite/Polite, Disrespectful/Respectful, Judgmental/Impartial, and Informal/Formal.

Annotator Agreement

Cronbach's α, a reliability coefficient, was used for inter-annotator consistency which ranged from a moderate value of 0.73 to a high value of 0.92.

- Each utterance was rated by 10 participants
- Scores averaged across rates to calculate single rating on each dimension

Batch	Formal	Friendly	Impartial	Polite	Respectful
1	0.82	0.86	0.84	0.86	0.83
2	0.88	0.89	0.86	0.86	0.87
3	0.80	0.87	0.73	0.84	0.78
4	0.85	0.91	0.79	0.88	0.87
5	0.77	0.89	0.81	0.87	0.87
6	0.91	0.82	0.81	0.87	0.86
7	0.85	0.86	0.84	0.84	0.84

Model Outputs for Each Rated Dimension

B is fixed effects (estimate average across population), CI is confidence interval (range of values expected where 95% is expected to lie), P is probability of observing data (small values means unlikely to be due to chance)

	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.

Principal Component Analysis (PCA) Loadings

Used to decompose ratings into underlying components

(value represents strength and direction of components, variance shows weight specific component had)

Two categories of the 5 explained 93.2% variance (respect + formality)

Formality equal between white and black drivers

Respect higher among white drivers

	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%

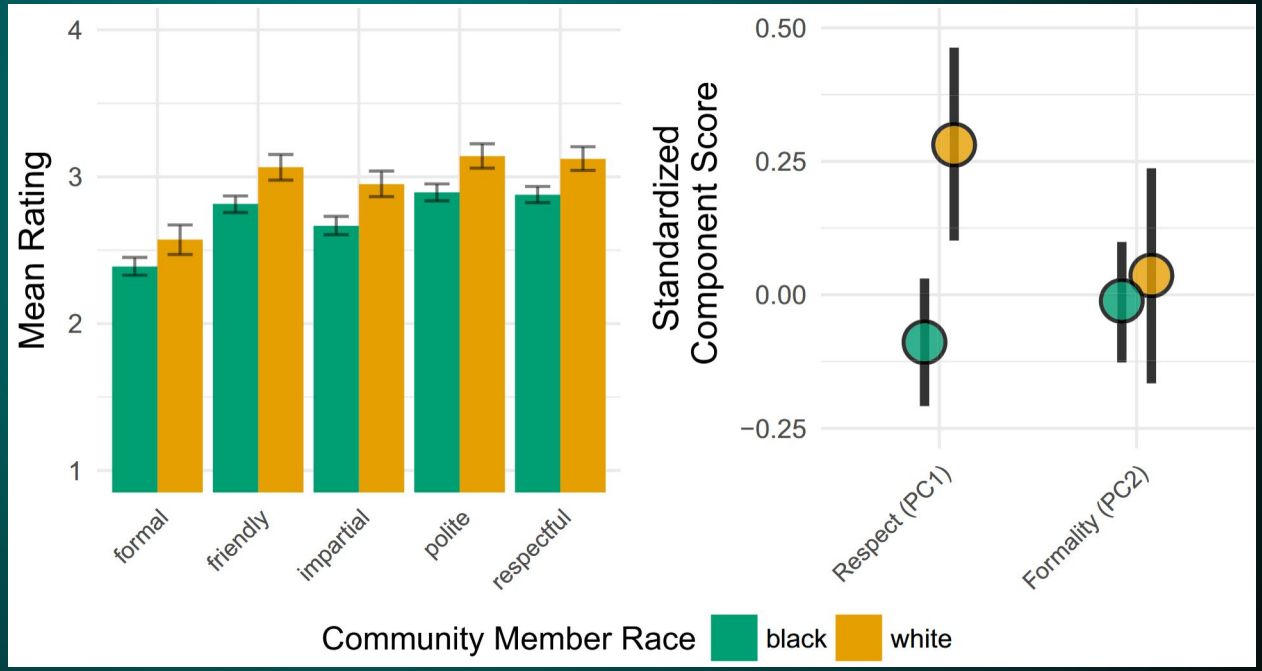
Full Regression Model Output

Beta is change in dependent variable, CI is confidence interval (range of values expected where 95% is expected to lie), P is the probability of observing data (small values means unlikely to be due to chance

	Respect			*Formality*		
	β	CI	p	β	CI	p
Fixed Parts						
Arrest Occurred	-0.08	-0.20 – 0.04	.210	0.04	-0.09 – 0.17	.532
Citation Issued	0.05	-0.06 – 0.16	.387	0.13	0.02 – 0.25	**.023**
Search Conducted	-0.23	-0.34 – -0.11	**<.001**	0.04	-0.08 – 0.17	.470
Age	0.05	-0.05 – 0.15	.321	0.11	0.01 – 0.21	**.036**
Gender (F)	-0.03	-0.12 – 0.07	.608	0.09	-0.01 – 0.19	.089
Race (W)	0.17	0.00 – 0.33	**.046**	-0.01	-0.19 – 0.16	.873
Officer Race (B)	-0.03	-0.18 – 0.11	.646	0.04	-0.11 – 0.20	.565
Officer Race (O)	0.00	-0.15 – 0.14	.966	-0.08	-0.23 – 0.07	.291
Officer Race (B) : Race (W)	0.02	-0.12 – 0.16	.799	-0.03	-0.18 – 0.11	.658
Officer Race (O) : Race (W)	-0.07	-0.22 – 0.09	.405	0.01	-0.15 – 0.18	.869
Random Parts						
σ^2		0.751			0.870	
$\tau_{00,\text{Stop:Officer}}$		0.010			0.000	
$\tau_{00,\text{Officer}}$		0.115			0.107	
$N_{\text{Stop:Officer}}$		254			254	
N_{Officer}		118			118	
$ICC_{\text{Stop:Officer}}$		0.011			0.000	
ICC_{Officer}		0.132			0.110	
Observations		414			414	
R^2 / Ω_0^2		.358 / .335			.255 / .213	

Differences in raw participant ratings

Where the two components, respect and formality, were derived from.



Issues with study 1

- Small sample size, doesn’t represent data as a whole
- Impossible to see how interactions progress, since only the single officer utterance with the preceding driver’s utterance was shown to participants of the study

Study 2

Based on linguistic features

Tuned on 414 unique officer utterances (1.1% of total)

Based on linguistic theories of respect, which is a form of honorifics

Model-assigned ratings agreed with the average human from study 1

RMSE of Respect from study 1: 0.840

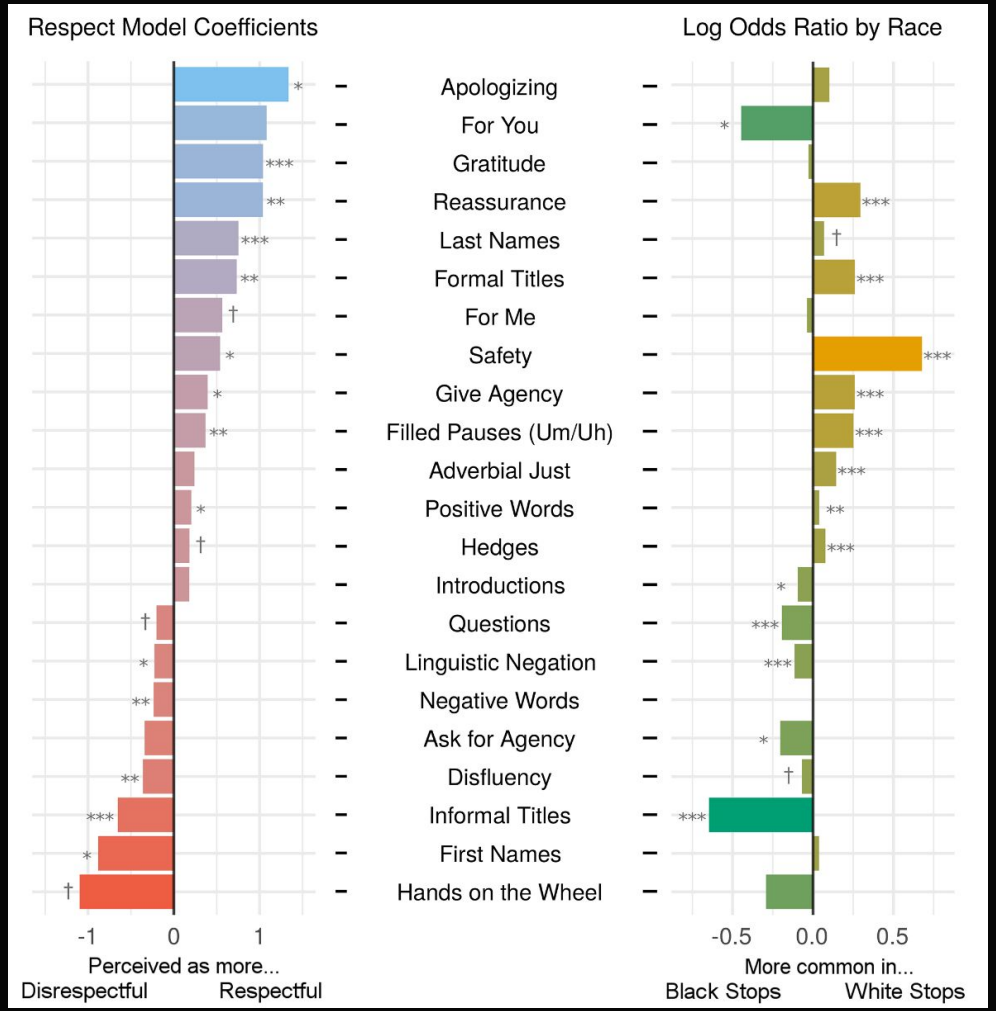
RMSE of Formality from study 1: 0.882

RMSE (root mean square error)

	Mean	Median	Max	Min
Respect	0.842	0.826	1.677	0.497
Formality	0.764	0.718	1.703	0.518

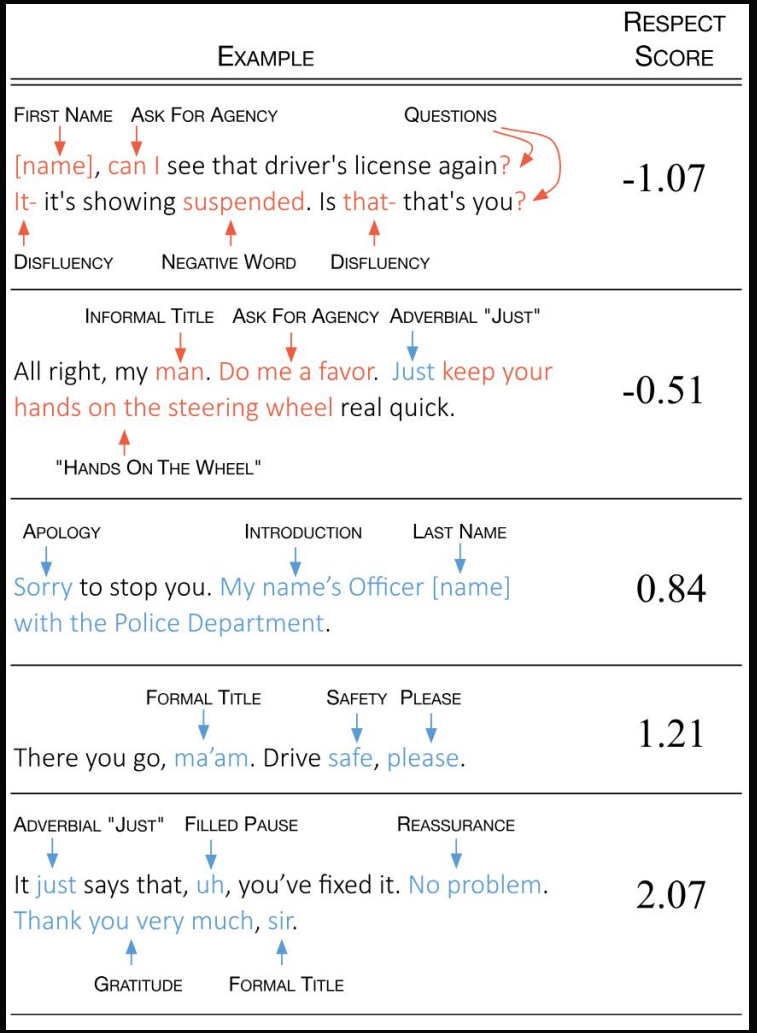
Respect weights by final model

Natural logarithmic scale of features for right graph, positive numbers represent more white community members, negative represents more black community members



Examples utterances of highest-weighted features

blue is positive connotations, red is negative



Study 3

Study 2 model applied to full corpus, in order to generate predicted scored of respect and formality for each of the 26,738 utterances

Built a linear mixed-effects model, and included a number of covariates in the primary model

Covariates include: community member race, age, gender, officer race, whether the search was conducted, and the result of being stopped (warning, citation, or arrest).

Utterances by officers to white people were higher in respect

Utterances spoken to older people were higher in respect

Respect as lower in stops where a search ended up being conducted.

Officer race did not contribute any significant effect

Race was not indicative of formality

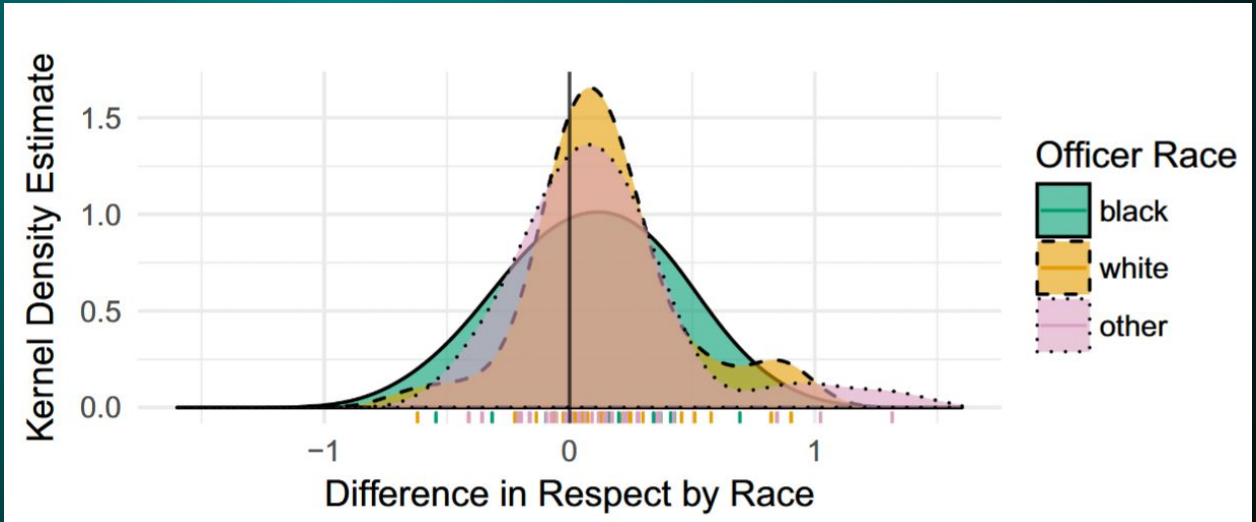
Formality was higher for women and older people

The main model presented and discussed in the paper is given below.

	Respect			*Formality*		
	β	CI	p	β	CI	p
Fixed Parts						
Arrest Occurred	0.00	-0.03 – 0.03	.933	0.01	-0.02 – 0.04	.528
Citation Issued	0.04	0.02 – 0.06	**<.001**	0.01	-0.01 – 0.03	.209
Search Conducted	-0.08	-0.11 – -0.05	**<.001**	0.00	-0.03 – 0.02	.848
Age	0.07	0.05 – 0.09	**<.001**	0.05	0.03 – 0.07	**<.001**
Gender (F)	0.02	0.00 – 0.04	.062	0.02	0.00 – 0.04	**.025**
Race (W)	0.05	0.03 – 0.08	**<.001**	-0.01	-0.04 – 0.01	.236
Officer Race (B)	0.00	-0.03 – 0.04	.884	0.00	-0.03 – 0.03	.987
Officer Race (O)	0.00	-0.04 – 0.03	.809	0.00	-0.03 – 0.02	.783
Officer Race (B) : Race (W)	-0.01	-0.03 – 0.02	.583	0.01	-0.01 – 0.03	.188
Officer Race (O) : Race (W)	-0.01	-0.03 – 0.02	.486	0.00	-0.02 – 0.02	.928
Random Parts						
σ^2		0.918			0.954	
$\tau_{00,\text{Stop:Officer}}$		0.045			0.029	
$\tau_{00,\text{Officer}}$		0.029			0.015	
$N_{\text{Stop:Officer}}$		981			981	
N_{Officer}		245			245	
$ICC_{\text{Stop:Officer}}$		0.045			0.029	
ICC_{Officer}		0.029			0.015	
Observations		36738			36738	
R^2 / Ω_0^2		.100 / .097			.064 / .059	

Density estimate of individual officer-level differences in Respect

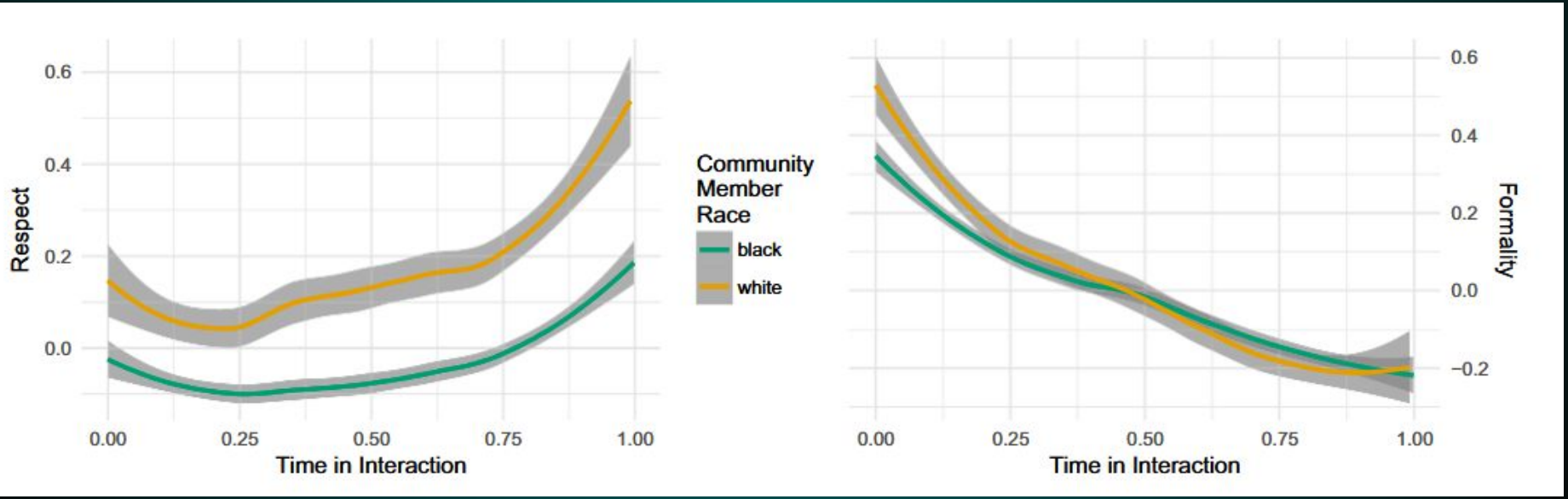
Normal distribution used to show that the racial disparity seen was not by a few extreme outlier police officers, but rather a consistent trend among most cops



Respect and Formality Over Time

Officers spoke with increased respect and decreased formality over the course of an interaction.

Officers became more respectful to white drivers quicker than black drivers.



Findings

- People made consistent judgements from language
- Respect and formality were the most notable factors
- Strong evidence in racial disparities for respect, but not formality
- White community members 57% more likely to hear one of the most respectful utterances
- Black community members 61% more likely to hear one of the least respectful utterances
- Computational linguistic model able to overcome hurdle of privacy concerns and scale of data
- Classifier trained on words that officers used can accurately predict race of community member 2/3rds of the time

Commentary

I found this research paper to be interesting and the methods they used seem to be thorough and successful.

I thought they repeated themselves a bit too much at times, reiterating their findings throughout the entirety of the paper.

I found it interesting they used the term “blacks”, since it’s seen as more disrespectful than “black people”.

Quiz Time:

What of these was not an issue the previous research papers had on this topic?

A. Limited number of interactions

B. Indirect view of the officers behavior

C. Presence of researches might influence police behavior

D. Impossible to see how interaction progresses over time

What were the findings of this study?

A. Officers were more respectful and formal to black people than white people.

B. Officers were equally formal to both white and black people, but were less respectful to black people.

C. Racial disparity only occured between white cops and black citizens

D. They were inconclusive.

If officers became more respectful as interactions went on but showed respect to white drivers more quickly than to Black drivers, how might this difference in timing and rate of respect explain why Black community members were 61% more likely to hear one of the least respectful remarks?

A. The slower increase in respect toward Black drivers meant they were more exposed to disrespectful comments early in interactions.

B. Black drivers were more likely to engage in longer conversations, leading to more opportunities for disrespect.

C. Officers reserved respectful language only for the end of conversations with all drivers.

D. White drivers used more polite language, prompting greater respect from officers overall.

Because earlier studies were limited by small samples and biased recollections, how did using computational linguistics to analyze 36,738 officer utterances from body camera footage help researchers confirm the pattern of racial disparities in respect?

A. It allowed them to test whether differences in respect appeared across a large number of officers, showing the pattern wasn't just due to a few individuals.

B. It helped them replace statistical methods with personal observations.

C. It enabled researchers to collect only the most respectful utterances for analysis.

D. It reduced the need to compare racial differences altogether.

QUESTIONS?