# 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 speak with consistently less respect toward black versus white community members, even after controlling for the race of the officer, the severity of the infraction, the location of the stop, and the outcome of the stop. Such disparities in common, everyday interactions between police and the communities they serve have important implications for procedural justice and the building of police–community trust.**

racial disparities | natural language processing | procedural justice | traffic stops | policing

Over the last several years, our nation has been rocked by an onslaught of incidents captured on video involving police officers' use of force with black suspects. The images from these cases are disturbing, both exposing and igniting police–community conflict all over the country: in New York, Missouri, Ohio, South Carolina, Maryland, Illinois, Wisconsin, Louisiana, Oklahoma, and North Carolina. These images have renewed conversations about modern-day race relations and have led many to question how far we have come (1). In an effort to increase accountability and transparency, law enforcement agencies are adopting body-worn cameras at an extremely rapid pace (2, 3).

Despite the rapid proliferation of body-worn cameras, no law enforcement agency has systematically analyzed the massive amounts of footage these cameras produce. Instead, the public and agencies alike tend to focus on the fraction of videos involving high-profile incidents, using footage as evidence of innocence or guilt in individual encounters.

Left unexamined are the common, everyday interactions between the police and the communities they serve. By best estimates, more than one quarter of the public (ages 16 y and over) comes into contact with the police during the course of a year, most frequently as the result of a police-initiated traffic stop (4, 5). Here, we examine body-worn camera footage of routine traffic stops in the large, racially diverse city of Oakland, CA.

Routine traffic stops are not only common, they are consequential, each an opportunity to build or erode public trust in the police. Being treated with respect builds trust in the fairness of an officer's behavior, whereas rude or disrespectful treatment can erode trust (6, 7). Moreover, a person's experiences of respect or disrespect in personal interactions with police officers play a central role in their judgments of how procedurally fair the police are as an institution, as well as their willingness to support or cooperate with the police (8, 9).

Blacks report more negative experiences in their interactions with the police than other groups (10). Across numerous studies, for example, blacks report being treated less fairly and respectfully in their contacts with the police than whites (6, 11). Indeed, 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?

We address this question by analyzing officers' language during vehicle stops of white and black community members. Although many factors may shape these interactions, an officer's words are undoubtedly critical: Through them, the officer can communicate respect and understanding of a citizen's perspective, or contempt and disregard for their voice. Furthermore, the language of those in positions of institutional power (police officers, judges, work superiors) has greater influence over the course of the interaction than the language used by those with less power (12–16). Measuring officer language thus provides a quantitative lens on one key aspect of the quality or tone of police–community interactions, and offers new opportunities for advancing police training.

Previous research on police–community interactions has relied on citizens' recollection of past interactions (10) or researcher observation of officer behavior (17–20) to assess procedural fairness. Although these methods are invaluable, they offer an indirect view of officer behavior and are limited to a small number of interactions. Furthermore, the very presence of researchers may influence the police behavior those researchers seek to measure (21).

## Significance

**Police officers speak significantly less respectfully to black than to white community members in everyday traffic stops, even after controlling for officer race, infraction severity, stop location, and stop outcome. This paper presents a systematic analysis of officer body-worn camera footage, using computational linguistic techniques to automatically measure the respect level that officers display to community members. This work demonstrates that body camera footage can be used as a rich source of data rather than merely archival evidence, and paves the way for developing powerful language-based tools for studying and potentially improving police–community relations.**

Author contributions: R.V., N.P.C., D. Jurafsky, and J.L.E. designed research; R.V. and N.P.C. performed research; V.P., W.L.H., R.C.H., C.M.G., and D. Jurgens contributed new reagents/analytic tools; R.V. and N.P.C. analyzed data; R.V., N.P.C., D. Jurafsky, and J.L.E. wrote the paper; and D. Jurafsky and J.L.E. served as PI on this project.

Reviewers: J.P., University of Texas at Austin; and T.T., Yale Law School.

Conflict of interest statement: J.L.E. was invited by a federal judge and monitor to serve as a Subject Matter Expert to assist with the Oakland Police Department's reform efforts. The assignment began prior to the studies reported here.



[1]To whom correspondence may be addressed. Email: robvoigt@stanford.edu or jleberhardt@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1702413114/-/DCSupplemental.

In study 1, human participants rated officer utterances on several overlapping dimensions of respect. With a high degree of agreement, participants inferred these dimensions from officer language. Even though they were not told the race of the stopped driver, participants judged officer language directed toward black motorists to be less respectful than language directed toward whites. In study 2, we build statistical models capable of predicting aspects of respect based on linguistic features derived from theories of politeness, power, and social distance. We discuss the linguistic features that contribute to each model, finding that particular forms of politeness are implicated in perceptions of respect. In study 3, we apply these models to all vehicle stop interactions between officers of the Oakland Police Department and black/white community members during the month of April 2014. We find strong evidence that utterances spoken to white community members are consistently more respectful, even after controlling for contextual factors such as the severity of the offense or the outcome of the stop.

## Data

Our dataset consists of transcribed body camera footage from vehicle stops of white and black community members conducted by the Oakland Police Department during the month of April 2014. We examined 981 stops of black ($N = 682$) and white ($N = 299$) drivers from this period, 68.1% of the 1,440 stops of white and black drivers in this period. These 981 stops were conducted by 245 different officers (see *SI Appendix, Data Sampling Process* for inclusion criteria). Per Oakland Police Department policy, officers turn on their cameras before making contact with the driver and record for the duration of the stop. From the 183 h of footage in these interactions, we obtain 36,738 usable officer utterances for our analysis.

**Study 1: Perceptions of Officer Treatment from Language.** We first test whether human raters can reliably judge respect from officers' language, and whether these judgments reveal differences in officer respect toward black versus white community members.

Respect is a complex and gradient perception, incorporating elements of a number of correlated constructs like friendliness and formality. Therefore, in this study, we ask participants to rate transcribed utterances spoken by officers along five conceptually overlapping folk notions related to respect and officer treatment. We randomly sampled 414 unique officer utterances (1.1% of all usable utterances in the dataset) directed toward black ($N = 312$) or white ($N = 102$) community members. On each trial, participants viewed the text of an officer utterance, along with the driver's utterance that immediately preceded it. All proper names and places were anonymized, and participants were not told the race or gender of the driver. Participants indicated on four-point Likert scales how respectful, polite, friendly, formal, and impartial the officer was in each exchange. Each utterance was rated by at least 10 participants.

Could participants reliably glean these qualities from such brief exchanges? Previous work has demonstrated that different perceivers can arrive at similar judgments from "thin slices" of behavior (22). In a similar vein, participants showed consistency in their perceptions of officer language, with reliability for each item ranging from moderate (Cronbach's $\alpha = 0.73$) to high ($\alpha = 0.91$) agreement (see *SI Appendix, Annotator Agreement*). These results demonstrate that transcribed language provides a sufficient and consensual signal of officer communication, enough to gain a picture of the dynamics of an interaction at a given point in time.

To test whether participant ratings uncovered racial group differences, we averaged scores across raters to calculate a single rating on each dimension for each utterance, then built a linear mixed-effects regression model to estimate the fixed effect of community member race across interactions, controlling for variance of a random effect at the interaction level. Officer utterances directed toward black drivers were perceived as less respectful [$b = -0.23$, 95% confidence interval ($-0.34$, $-0.11$)], polite [$b = -0.23$ ($-0.35$, $-0.12$)], friendly [$b = -0.24$ ($-0.36$, $-0.12$)], formal [$b = -0.16$ ($-0.30$, $-0.03$)], and impartial [$b = -0.26$ ($-0.39$, $-0.12$)] than language directed toward white drivers (Fig. 1). These differences persisted even when controlling for the age and sex of the driver (see *SI Appendix, Model Outputs for Each Rated Dimension*).

Given the expected conceptual overlap in the five perceptual categories we presented to the participants, we used principal component analysis to decompose the ratings into their underlying components. Two principal components explained 93.2% of the variance in the data (see *SI Appendix, Principal Component Analysis (PCA) Loadings* for loadings). The first component, explaining 71.3% of the variance and composed of positive loadings on the impartial, respectful, friendly, and polite dimensions with some loading on the formal dimension, we characterize as Respect, broadly construed. The second, explaining 21.9% of the variance and composed primarily of a very high positive loading on the formal dimension and a weak negative loading on the friendly dimension, we characterize as Formality. This component captures formality as distinct from respect more generally, and is likely related to social distance.

Standardizing these factor scores as outcome variables in mixed-effects models, we find that officers were equal in Formality with white and black drivers [$\beta = -0.01$ ($-0.19$, $0.16$)], but higher in Respect with white drivers [$\beta = 0.17$ ($0.00$, $0.33$)] (Fig. 1).

Study 1 demonstrates that key features of police treatment can be reliably gleaned from officer speech. Participant ratings from thin slices of police–community interactions reveal racial disparities in how respectful, impartial, polite, friendly, and formal officers' language to community members was perceived. Such differences were driven by differences in the Respect officers communicated toward drivers rather than the Formality with which officers addressed them.

**Study 2: Linguistic Correlates of Respect.** The methods of study 1 (human coding of 414 individual utterances), although effective at discovering racial disparities in officer respect toward community members in our dataset, cannot offer a general solution to the analysis of body camera data. One problem is scale: Each year, on the order of 26 million vehicle stops are made (5). Furthermore, using only a small sample of individual utterances makes it impossible to study how police treatment varies over officers, or how the interaction progresses across time in each stop.

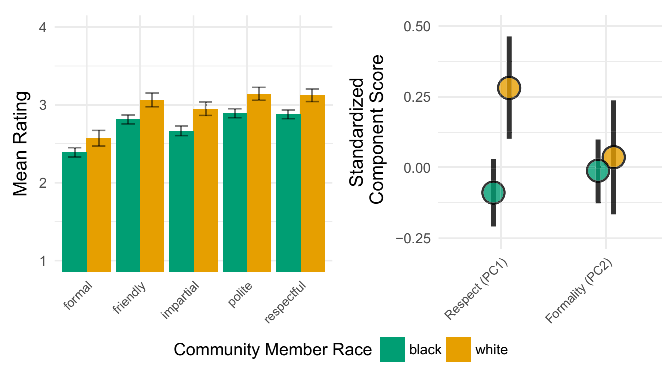


**Fig. 1.** (*Left*) Differences in raw participant ratings between interactions with black and white community members. (*Right*) When collapsed to two uncorrelated components, Respect and Formality, we find a significant difference for Respect but none for Formality. Error bars represent 95% confidence intervals. PC, principal component.

In this study, we therefore develop computational linguistic models of respect and formality and tune them on the 414 individual utterances; in study 3, we apply these models to our full dataset of 36,738 utterances. Our method is based on linguistic theories of respect that model how speakers use respectful language (apologizing, giving agency, softening of commands, etc.) to mitigate "face-threatening acts." We use computational linguistic methods (e.g., refs. 23–26) to extract features of the language of each officer utterance. The log-transformed counts of these features are then used as independent variables in two linear regression models predicting the perceptual ratings of Respect and Formality from study 1.

Our model-assigned ratings agree with the average human from study 1 about as well as humans agree with each other. Our model for Respect obtains an adjusted $R^2$ of 0.258 on the perceptual ratings obtained in study 1, and a root-mean-square error (RMSE) of 0.840, compared with an RMSE of 0.842 for the average rater relative to other raters. Our model for Formality obtains an adjusted $R^2$ of 0.190, and an RMSE of 0.882 compared with 0.764 for the average rater (see *SI Appendix, Model Comparison to Annotators* for more details on how these values were calculated). These results indicate that, despite the sophisticated social and psychological cues participants are likely drawing upon in rating officers' utterances, a constrained set of objectively measurable linguistic features can explain a meaningful portion of the variance in these ratings.

Fig. 2 lists the linguistic features that received significant weights in our model of Respect (arranged by their model coefficients). For example, apologizing, gratitude, and expressions of concern for citizen safety are all associated with respect. The bars on the right show the log-odds of the relative proportion of interactions in our dataset taken up by each feature, where negative numbers mean that a feature comprised a larger proportion of officers' speech in interactions with black community members and positive numbers mean the same for interactions

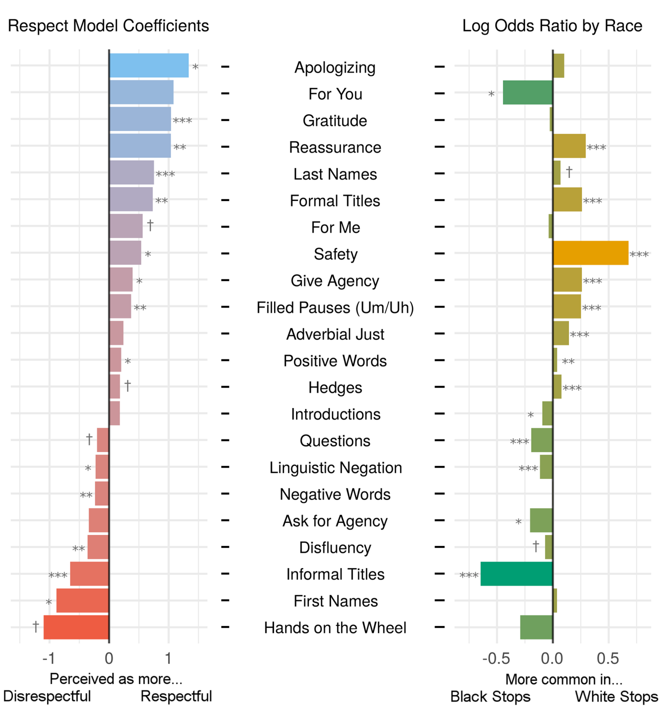


**Fig. 2.** (*Left*) Respect weights assigned by final model to linguistic features and (*Right*) the corresponding log-odds of those features occurring in officer speech directed toward black versus white community members, calculated using Fisher's exact test. $^{\dagger}P < 0.1$; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$.

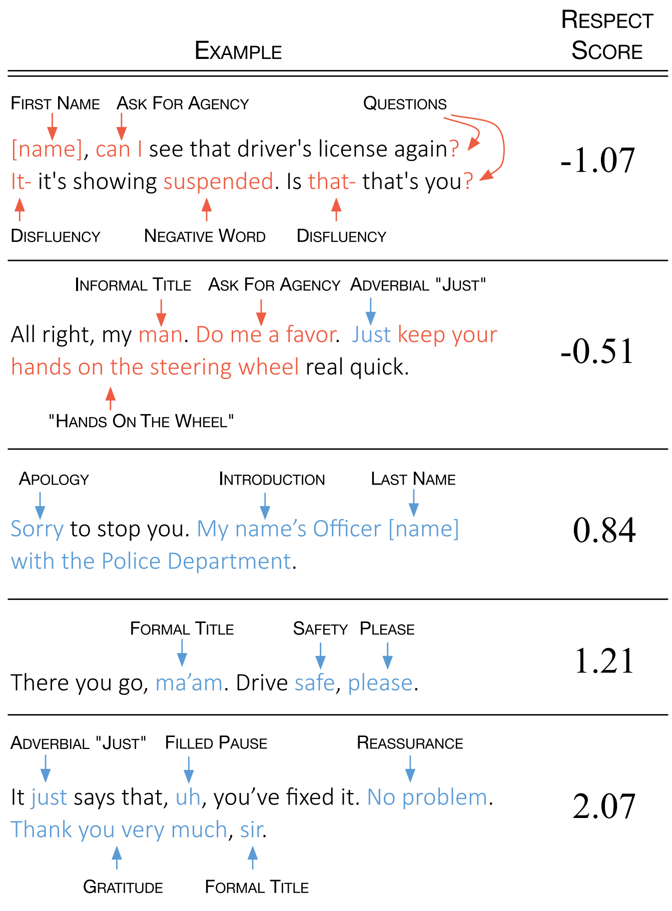


**Fig. 3.** Sample sentences with automatically generated Respect scores. Features in blue have positive coefficients in the model and connote respect, such as offering reassurance ("no problem") or mentioning community member well-being ("drive safe"). Features in red have negative coefficients in the model and connote disrespect, like informal titles ("my man"), or disfluencies ("that- that's").

with white community members. Example utterances containing instances of the highest-weighted features for the Respect model are shown in Fig. 3. See *SI Appendix, Study 2* for full regression outputs and more detailed discussion of particular linguistic findings.

**Study 3: Racial Disparities in Respect.** Having demonstrated that people can reliably infer features of procedural justice from officer speech (study 1), and that these ratings can be reliably predicted from statistical models of linguistic features (study 2), we are now able to address our central question: Controlling for contextual factors of the interaction, is officers' language more respectful when speaking to white as opposed to black community members?

We apply our models from study 2 to the entire corpus of transcribed interactions to generate predicted scores for Respect and Formality for each of the 36,738 utterances in our dataset. We then build linear mixed-effects models for Respect and Formality over these utterances. We include, as covariates in our primary model, community member race, age, and gender; officer race; whether a search was conducted; and the result of the stop (warning, citation, or arrest). We include random intercepts for interactions nested within officers.

Controlling for these contextual factors, utterances spoken by officers to white community members score higher in Respect [$\beta = 0.05$ (0.03, 0.08)]. Officer utterances were also higher in

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Respect when spoken to older [$\beta = 0.07$ (0.05, 0.09)] community members and when a citation was issued [$\beta = 0.04$ (0.02, 0.06)]; Respect was lower in stops where a search was conducted [$\beta = -0.08$ ($-0.11$, $-0.05$)]. Officer race did not contribute a significant effect. Furthermore, in an additional model on 965 stops for which geographic information was available, neither the crime rate nor density of businesses in the area of the stop were significant, although a higher crime rate was indicative of increased Formality [$\beta = 0.03$ (0.01, 0.05)].

One might consider the hypothesis that officers were less respectful when pulling over community members for more severe offenses. We tested this by running another model on a subset of 869 interactions for which we obtained ratings of offense severity on a four-point Likert scale from Oakland Police Department officers, including these ratings as a covariate in addition to those mentioned above. We found that the offense severity was not predictive of officer respect levels, and did not substantially change the results described above.

To consider whether this disparity persists in the most "everyday" interactions, we also reran our analyses on the subset of interactions that did not involve arrests or searches ($N = 781$), and found the results from our earlier models were fundamentally unchanged. Full regression tables for all models described above are given in *SI Appendix*, *Study 3*.

Another hypothesis is that the racial disparities might have been caused by officers being more formal to white community members, and more informal or colloquial to black community members. However, we found that race was not associated with the formality of officers' utterances. Instead, utterances were higher in Formality in interactions with older [$\beta = 0.05$ (0.03, 0.07)] and female [$\beta = 0.02$ (0.00, 0.04)] community members.

Are the racial disparities in the respectfulness of officer speech we observe driven by a small number of officers? We calculated the officer-level difference between white and black stops for every officer ($N = 90$) in the dataset who had interactions with both blacks and whites (Fig. 4). We find a roughly normal distribution of these deltas for officers of all races. This contrasts with the case of stop-and-frisk, where individual outlier officers account for a substantial proportion of racial disparities (27); the disparities we observe here cannot be explained by a small number of extreme officers.

Because our model is able to generate scores across all utterances in our dataset, we can also consider aspects of the trajectory of interactions beyond the mean level of respect (Fig. 5). Growth-curve analyses revealed that officers spoke with greater Respect [$b = 0.35$ (0.29, 0.40)] and reduced Formality [$b = -0.57$ ($-0.62$, $-0.53$)] as interactions progressed. However, these trajectories varied by community member race: Although stops of white and black drivers converged in the Formality expressed during the interaction [$b = -0.09$ ($-0.13$, $-0.05$)], the gap in Respect increased over time [$b = 0.10$ (0.05, 0.15)]. That is, officer Respect increased more quickly in interactions with white drivers [$b = 0.45$ (0.38, 0.54)] than in interactions with black drivers [$b = 0.24$ (0.19, 0.29)].

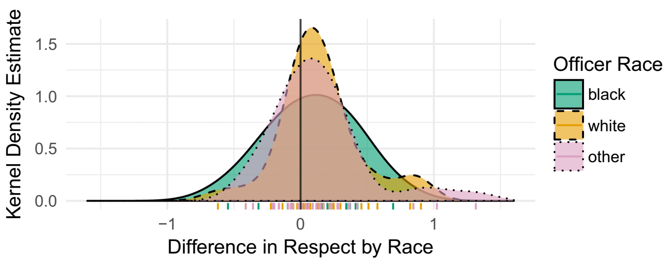


**Fig. 4.** Kernel density estimate of individual officer-level differences in Respect when talking to white as opposed to black community members, for the 90 officers in our dataset who have interactions with both blacks and whites. More positive numbers on the *x* axis represent a greater positive shift in Respect toward white community members.

**Discussion.** Despite the formative role officer respect plays in establishing or eroding police legitimacy (7), it has been impossible to measure how police officers communicate with the public, let alone gauge racial disparities in officer respect. However, body-worn cameras capture such interactions every day. Computational linguistic techniques let us examine police–community contacts in a manner powerful enough to scale to any number of interactions, but sensitive enough to capture the interpersonal qualities that matter to the police and public alike.

In doing so, we first showed that people make consistent judgments about such interactions from officers' language, and we identified two underlying, uncorrelated constructs perceived by participants: Respect and Formality. We then built computational linguistic models of these constructs, identifying crucial positive and negative politeness strategies in the police–community interactional context. Applying these models to an entire month of vehicle stops, we showed strong evidence for racial disparities in Respect, but not in Formality: Officers' language is less respectful when speaking to black community members.

Indeed, we find that white community members are 57% more likely to hear an officer say one of the most respectful utterances in our dataset, whereas black community members are 61% more likely to hear an officer say one of the least respectful utterances in our dataset. (Here we define the top 10% of utterances to be most respectful and the bottom 10% to be least respectful.)

This work demonstrates the power of body camera footage as an important source of data, not just as evidence, addressing limitations with methodologies that rely on citizens' recollection of past interactions (10) or direct researcher observation of police behavior (17–20). However, studying body camera footage presents numerous hurdles, including privacy concerns and the raw scale of the data. The computational linguistic models presented here offer a path toward addressing both these concerns, allowing for the analysis of transcribed datasets of any size, and generating reliable ratings of respect automatically. These models have the potential to allow for useful information about an interaction to be extracted while maintaining officer and community member privacy.

The racial disparities in officer respect are clear and consistent, yet the causes of these disparities are less clear. It is certainly possible that some of these disparities are prompted by the language and behavior of the community members themselves, particularly as historical tensions in Oakland and preexisting beliefs about the legitimacy of the police may induce fear, anger, or stereotype threat. However, community member speech cannot be the sole cause of these disparities. Study 1 found racial disparities in police language even when annotators judged that language in the context of the community member's utterances. We observe racial disparities in officer respect even in police utterances from the initial 5% of an interaction, suggesting that officers speak differently to community members of different races even before the driver has had the opportunity to say much at all.

Regardless of cause, we have found that police officers' interactions with blacks tend to be more fraught, not only in terms of disproportionate outcomes (as previous work has shown) but also interpersonally, even when no arrest is made and no use of force occurs. These disparities could have adverse downstream effects, as experiences of respect or disrespect in personal interactions with police officers play a central role in community members' judgments of how procedurally fair the police are as an institution, as well as the community's willingness to support or cooperate with the police (8, 9).

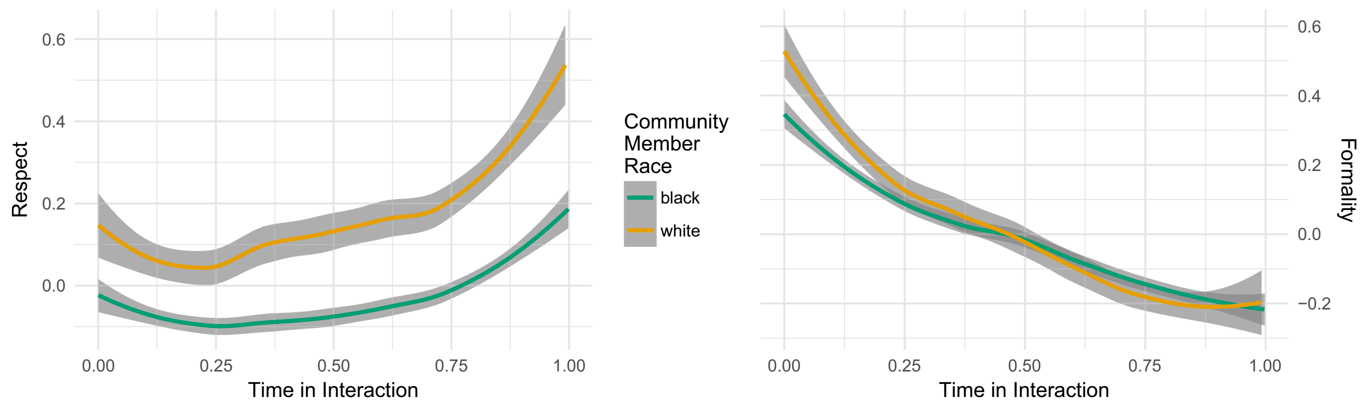


**Fig. 5.** Loess-smoothed estimates of the (*Left*) Respect and (*Right*) Formality of officers' utterances relative to the point in an interaction at which they occur. Respect tends to start low and increase over an interaction, whereas the opposite is true for Formality. The race discrepancy in Respect is consistent throughout the interactions in our dataset.

We now have a method for quantifying these troubled interactions. Although the circumstances of any particular stop can vary dramatically, our approach allows us to measure aggregate department-level trends, revealing disparities across hundreds of interactions. These disparities are part of a constellation of differences in officer language spoken toward black versus white community members; a simple classifier trained on only the words used by officers is able to correctly predict the race of the community member in over two thirds of the interactions (see *SI Appendix, Linguistic Classification Accuracy of Race*).

Future research could expand body camera analysis beyond text to include information from the audio such as speech intonation and emotional prosody, and video, such as the citizen's facial expressions and body movement, offering even more insight into how interactions progress and can sometimes go awry. In addition, footage analysis could help us better understand what linguistic acts lead interactions to go well, which can inform police training and quantify its impacts over time.

The studies presented here open a path toward these future opportunities and represent an important area of research for the study of policing: Computational, large-scale analyses of language give us a way to examine and improve police–community interaction that we have never had before.

## Materials and Methods

**Data and Processing.** The video for each traffic stop was transcribed into text by professional transcribers, who transcribed while listening to audio and watching the video. Extensive measures were taken to preserve privacy; data were kept on a central server, and transcribers (as well as all researchers) underwent background checks with the Oakland Police Department. Transcribers also "diarized" the text (labeling who was speaking at each time point). We used the diarization to automatically remove all officer speech to the dispatcher or to other officers, leaving only speech from the officer directed toward the community member. After transcription, transcripts were manually cleaned up, heuristically fixing transcriber diarization errors, and correcting typographical errors involving utterance timing so that all transcripts were automatically readable. Every utterance in the dataset was processed with Stanford CoreNLP 3.4.1 (28) to generate sentence and word segmentation, part-of-speech tags, and dependency parses used for feature extraction and analysis.

The raw video footage associated with this paper was available for our research purposes with the cooperation of the Oakland Police Department, and naturally cannot be publicly distributed. However, we make available deidentified data frames for each study described here, so that other researchers can replicate our results. We also release all of the code for the computational linguistic models, as well as pretrained models that can be run on arbitrary text.

**Human Annotation of Utterances.** A subset of 420 exchanges, consisting of one officer utterance (defined as a "turn" of one or more sentences by transcribers) and, if applicable, the immediately preceding community member utterance were sampled from the corpus for annotation. Utterances were sampled with the constraint that at least 15 words were spoken between the two speakers, and that at least five words were spoken by the officer. These utterances were grouped into seven "batches" of 60 utterances apiece. Due to a data error, six duplicate utterances were annotated, but were excluded from subsequent analyses, resulting in 414 unique utterances toward black ($N$ = 312) and white ($N$ = 102) community members.

Each of 70 participants (39 female, $M_{age}$ = 25.3) rated a batch of 60 of these utterances, such that each utterance was rated by at least 10 participants. On each trial, participants viewed the text of an exchange between a police officer and a community member: the text of the officer utterance, as well as the text of the community member utterance that immediately preceded it, if there was one. They then indicated, on four-point bipolar Likert scales, how respectful, polite, friendly, formal, and impartial the officer was in each exchange. Participants were allowed to indicate that they could not rate an utterance on a particular dimension, but were encouraged to nonetheless indicate their best guess. Participants had no other information about the interaction besides the officer's utterance and the immediately preceding community member utterance.

All research was approved by the Stanford University Institutional Review Board, and written informed consent was obtained from all raters before their participation.

**Computational Annotation of Utterances.** Our model draws on linguistic theories of politeness; the technical term "politeness" refers to how concepts like respect, formality, and social distance take shape in language. These theories suggest that speakers use polite or respectful language to mitigate face-threatening acts (29–31).

Negative politeness is used to mitigate direct commands or other impositions that limit the freedom of action of the listener, for example, by minimizing the imposition or emphasizing the agency of the interlocutor. Such strategies are central to police–community interactions because of the inherently coercive nature of a traffic stop. For instance, the use of the word "please" can soften requests and provide a sense of agency or choice; apologizing ("sorry," "excuse me") can admit regret on the part of the officer that some request is necessary; the use of hedges ("may," "kinda," "probably") may reduce the perception of imposition.

Positive politeness is used to show that the speaker values the interlocutor and their interests, or to minimize the impact of actions that could damage such a perception. Positive politeness strategies are also crucial for police–community interactions, where the inherently unequal social roles at play may necessitate a particular sensitivity to the community member's positive face. For instance, greetings and introductions can establish a friendly context at the beginning of an interaction and convey openness. Expressions of reassurance ("no big deal," "don't worry") seek to assuage the community member's potential concerns in tense circumstances, and expressions of gratitude ("thank you") serve to reduce the perceived power differential by deferring to the actions of the community member. Mentions of safety ("Drive safely now") explicitly acknowledge concern for the community member's personal well-being. Referring expressions are another important component of positive politeness;

PSYCHOLOGICAL AND COGNITIVE SCIENCES

formal titles ("sir," "ma'am," "Mr.," "Ms.") and surnames may convey a contrast with informal titles ("dude," "bro," "bud") and first names (31–33).

We also include features we expect to capture officer anxiety, such as speech disfluencies ("w- well") and commands to keep "hands on the wheel," which may contribute to a community member's perception of disrespect. These are of a different character than the politeness strategies discussed above, but we found that all analyses presented here hold true even if these features are not included.

We use standard techniques to automatically extract features from the text of each utterance (23–26). These features include lexicons (lists of words). For example, to detect informal titles, we used an augmented version of a word list from ref. 34. We also used regular expressions, such as for detecting tag questions ("do that for me, will you?"), and syntactic parse features, such as a feature that detects when "just" is used in constructions as an adverbial modifier.

Features were modeled as log-transformed counts in each utterance, and were used as independent variables in two linear regression models predicting the human perceptual ratings of respect and formality obtained in study 1. They were introduced into the regression using stepwise forward selection by $R^2$ to remove features that don't substantially contribute to the model's accuracy.

**ACKNOWLEDGMENTS.** This research was supported by the John D. and Catherine T. MacArthur Foundation, with additional support from the Stanford Institute for Research in the Social Sciences, the Stanford School of Humanities and Sciences, and the Stanford Data Science Initiative. We also thank the City of Oakland and the Oakland Police Department for their support and cooperation.

1. President's Task Force on 21st Century Policing (2015) *Final Report of the President's Task Force on 21st Century Policing* (Off Commun Oriented Policing Serv, Washington, DC).
2. The White House (December 1, 2014) Fact sheet: Strengthening community policing. Press release (Washington, DC). Available at https://obamawhitehouse.archives.gov/the-press-office/2014/12/01/fact-sheet-strengthening-community-policing. Accessed February 1, 2017.
3. Reaves B (2015) *Local Police Departments, 2013: Personnel, Policies, and Practices* (US Dep Justice, Washington, DC), NCJ 248677.
4. Eith C, Durose M (2011) *Contacts Between Police and the Public, 2008* (Bur Justice Stat, Washington, DC).
5. Langton L, Durose M (2013) *Special Report: Police Behavior During Traffic and Street Stops, 2011* (Bur Justice Stat, Washington, DC).
6. Tyler TR, Huo Y (2002) *Trust in the Law: Encouraging Public Cooperation with the Police and Courts* (Russell Sage Found, New York).
7. Tyler TR, Blader SL (2003) The group engagement model: Procedural justice, social identity, and cooperative behavior. *Pers Soc Psychol Rev* 7:349–361.
8. Tyler TR, Bies RJ (1990) Beyond formal procedures: The interpersonal context of procedural justice. *Applied Social Psychology and Organizational Settings*, ed Carroll JS (Lawrence Erlbaum, Hillsdale, NJ), pp 77–98.
9. Mazerolle L, Antrobus E, Bennett S, Tyler TR (2013) Shaping citizen perceptions of police legitimacy: A randomized field trial of procedural justice. *Criminology* 51:33–63.
10. Epp CR, Maynard-Moody S, Haider-Markel DP (2014) *Pulled Over: How Police Stops Define Race and Citizenship* (Univ Chicago Press, Chicago).
11. Peffley M, Hurwitz J (2010) *Justice in America: The Separate Realities of Blacks and Whites* (Cambridge Univ Press, New York).
12. Giles H, Coupland J, Coupland N (1991) Accommodation theory: Communication, context and consequences. *Contexts of Accommodation: Developments in Applied Sociolinguistics*, eds Giles H, Coupland J, Coupland N (Cambridge Univ Press, New York), pp 1–68.
13. Gnisci A (2005) Sequential strategies of accommodation: A new method in courtroom. *Br J Soc Psychol* 44:621–643.
14. Ng SH, Bell D, Brooke M (1993) Gaining turns and achieving high influence ranking in small conversational groups. *Br J Soc Psychol* 32:265–275.
15. Nguyen VA, et al. (2014) Modeling topic control to detect influence in conversations using nonparametric topic models. *Mach Learn* 95:381–421.
16. Prabhakaran V, Rambow O (2014) Predicting power relations between participants in written dialog from a single thread. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Assoc Comput Linguist, Stroudsburg, PA), pp 339–344.
17. Mastrofski SD, Parks RB, McCluskey JD (2010) Systematic social observation in criminology. *Handbook of Quantitative Criminology*, eds Piquero AR, Weisburd D (Springer, New York), pp 225–247.
18. Dai M, Frank J, Sun I (2011) Procedural justice during police-citizen encounters: The effects of process-based policing on citizen compliance and demeanor. *J Crim Justice* 39:159–168.
19. Jonathan-Zamir T, Mastrofski SD, Moyal S (2015) Measuring procedural justice in police-citizen encounters. *Justice Q* 32:845–871.
20. Mastrofski SD, Jonathan-Zamir T, Moyal S, Willis JJ (2016) Predicting procedural justice in police–citizen encounters. *Crim Justice Behav* 43:119–139.
21. Mastrofski S, Parks RB (1990) Improving observational studies of police. *Criminology* 28:475–496.
22. Ambady N, Bernieri FJ, Richeson JA (2000) Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Adv Exp Soc Psychol* 32:201–271.
23. Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 29:24–54.
24. Prabhakaran V, Rambow O, Diab M (2012) Predicting overt display of power in written dialogs. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics* (Assoc Comput Linguist, Stroudsburg, PA), pp 518–522.
25. Danescu-Niculescu-Mizil C, Lee L, Pang B, Kleinberg J (2012) Echoes of power: Language effects and power differences in social interaction. *Proceedings of the 21st International Conference on World Wide Web* (Assoc Comput Mach, New York) , pp 699–708.
26. Danescu-Niculescu-Mizil C, Sudhof M, Jurafsky D, Leskovec J, Potts C (2013) A computational approach to politeness with application to social factors. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (Assoc Comput Linguist, Stroudsburg, PA), pp 250–259.
27. Goel S, Rao JM, Shroff R (2016) Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy. *Ann Appl Stat* 10:365–394.
28. Manning CD, et al. (2014) The Stanford CoreNLP natural language processing toolkit. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Assoc Comput Linguist, Stroudsburg, PA), pp 55–60.
29. Goffman E (1967) On face-work. *Interaction Ritual: Essays on Face-to-Face Behavior* (Anchor, Garden City, NY), pp 5–45.
30. Lakoff RT (1973) The logic of politeness: Minding your p's and q's. *Papers from the 9th Regional Meeting of the Chicago Linguistic Society*, eds Corum C, Smith-Stark T, Weiser A (Chicago Linguist Soc, Chicago), pp 292–305.
31. Brown P, Levinson SC (1987) *Politeness: Some Universals in Language Usage* (Cambridge Univ Press, Cambridge, UK).
32. Wood LA, Kroger RO (1991) Politeness and forms of address. *J Lang Soc Psychol* 10:145–168.
33. Boxer D (1993) Social distance and speech behavior: The case of indirect complaints. *J Pragmat* 19:103–125.
34. Krishnan V, Eisenstein J (2015) "You're Mr. Lebowski, I'm the Dude": Inducing address term formality in signed social networks. *Proceedings of the North American Chapter of the Association for Computational Linguistics*, eds Elangovan V, Eisenstein J (Assoc Comput Linguist, Stroudsburg, PA), pp 1616–1626.

# Supplementary Material for "Language from police body camera footage shows racial disparities in officer respect"

## Contents

# 1 Data

## 1.1 Video Matching and Transcription

To observe how police officers communicated with community members, we sought body-worn camera footage from traffic stops conducted by the Oakland Police Department in a one-month period (April, 2014). Per OPD policy, officers are required to activate their camera for each stop they make, prior to making contact with the driver. Additionally, officers complete an electronic stop data form for each stop they make. This form links together information about the officer and community member demographics, along with information about the context (e.g. the date, time, and reason for the encounter) and outcomes of the stop (whether the officer issued a citation or searched the driver, for example).

We used these stop data forms to identify vehicle stops of black and white community members. Researchers manually matched videos to stops using the officer ID for the officer who wrote the stop report and the timestamp of the stop as identifiers. Videos were matched to 85.6% of stops.

Once videos were matched, a preliminary check was made to determine eligibility for transcription. To be transcribed, the officer recording the stop had to be the primary interlocutor for the community member, and the entirety of the stop had to be captured one recording to ensure adequate timestamping. We excluded stops where the recording we matched was from an officer who stood by the passenger door of the driver's car, for example, as well as stops where the officer activated their camera in the middle of the interaction. These exclusions resulted in 701 black and 308 white stops for transcription.

Transcribers underwent background checks and fingerprinting by the OPD in order to participate. They watched the videos and listened to the audio while transcribing via a secure streaming network connection. Each utterance was transcribed with an annotation for the speaker (OFFICER or MALE/FEMALE community member), the start time and end time of each utterance to the nearest second, and an indication of the audience of the officer's utterance (only for utterances not directed to the community member, that is, directed to the dispatch or another officer). Because, in a few cases, transcribers neglected to mark the audience of utterances to dispatch or other officers, we filtered officer utterances to only those that were marked as directed to a community member and also occurred immediately before or after a community member utterance.

Transcribers were instructed to mark disfluencies like word fragments ("th- th- that's all folks"), repetitions ("he he he said so"), filled pauses ("uh," "um"), and backchannels ("oh," "uh-huh"). Interruptions and word overlap were also marked, although were not analyzed in this study.

The transcribed interactions included 19 cases where there were fewer than 3 of officer utterances, which were excluded from analysis. The final dataset thus contained transcriptions of 68.3% of black stops and 70.9% of white stops.

## 1.2 Data Sampling Process

| | | |
|---|---|---|
| Total Vehicle Stops in April 2014 | 2159 | |
| | Black | White |
| Race of Community Member | 998 | 422 |
| **Unsuccessful Matches** | | |
| Officer Body-Worn Camera Not Activated | 1 | 1 |
| Video File Could Not be Opened | 3 | 2 |
| No Body-Worn Camera Issued to Officer | 48 | 35 |
| Could Not Locate File | 63 | 32 |
| Stops Matched | 883 | 352 |
| Proportion of Total Stops Matched | 0.884 | 0.834 |
| **Stops Marked Ineligible for Transcription** | | |
| Single Video Does Not Capture Entire Duration of Stop | 22 | 3 |
| Recording Officer Not Primary Interlocutor | 160 | 41 |
| Stops Transcribed | 701 | 308 |
| Proportion of Total Stops Transcribed | 0.702 | 0.729 |
| **Transcribed Stops Excluded From Analysis** | | |
| Fewer than 3 Turns | 19 | 9 |
| Stops in Dataset | 682 | 299 |
| Proportion of Total Stops in Dataset | 0.683 | 0.709 |

Table 1: Accounting of all vehicle stops conducted by the Oakland Police Department in April 2014, and the sampling process by which they were included in the final dataset. We attempted to obtain as clean and complete a full sample of all vehicle stops of black and white community members conducted as possible.

### 1.3 Description of the Data

Tables 2 and 3 give a brief overview of the data at the stop and officer level, respectively. Relative to white drivers, a greater proportion of black drivers were male, ($\chi^2$(2, N=981)= 6.9, p $<$ 0.001). On average, stops of black drivers lasted longer than stops of white drivers ($t$=8.56, df=978.94, p $<$ 0.001) , and black drivers were more likely to be searched than whites ($\chi^2$(2, N=981)= 50.8, p $<$ 0.001). Stop outcomes also differed by race, such that stops of black drivers were more likely to end in arrest than stops of white drivers, ($\chi^2$(2, N=981)= 17.7, p $<$ 0.001).

As illustrated in Table 2, the 245 officers captured in our data were largely male and a plurality were white. Officers varied greatly in the number of stops they conducted in our dataset, from a single stop (the modal value) to 40 stops. As a result, stop counts for individual officers were highly dispersed (M=4.0, $\sigma^2$=22.6). While multilevel models are robust to estimating marginal effects across groups of unequal size [1], the low number of datapoints per officer in our data limits our power to detect within-officer effects.

| Community Member Race | | Black | White |
|---|---|---|---|
| Total | | 682 | 299 |
| Gender | M | 463 | 177 |
| | F | 219 | 122 |
| Mean Age | | 35.5<br>SD=13.6 | 38.4<br>SD=13.4 |
| Stop Result | Arrest | 40 | 1 |
| | Citation | 369 | 185 |
| | Warning | 273 | 113 |
| Search Conducted | Yes | 113 | 2 |
| | No | 569 | 297 |
| Mean Stop Duration (Minutes) | | 12.6<br>SD=11.5 | 8.0<br>SD=5.1 |

Table 2: Characteristics of the data (Community Members)

| Total Officers | | 245 |
|---|---|---|
| Race | White | 102 |
| | Black | 39 |
| | Asian | 36 |
| | Hispanic | 57 |
| | Other | 11 |
| Gender | M | 224 |
| | F | 21 |
| Mean Age | | 35.5<br>SD=8.2 |
| Mean Years of Experience | | 7.1<br>SD=6.8 |
| Mean Number of Stops in Dataset | | 4<br>SD=4.8 |

Table 3: Characteristics of the data (Officers)

### 1.4 Severity Annotations

One factor we control for in our analysis of vehicle stops is the justification for the stop — i.e., the type of violation and its severity. This is necessary since it is possible that an officer's language may differ depending on the severity of the violation. We might expect officers to use less respectful language for, say, a driver who ran a stop sign versus a motorist driving with a broken taillight.

The Stop Data Form contains a "narrative" field where officers provide a written account of the interaction and its surrounding circumstances, and we obtained measurements for the severity of each stop by analyzing data from this narrative field. 13 senior OPD staff rated the severity of the violation associated with each stop an 1 (*Very Severe*) to 4 (*Very Minor*) scale.

For our analysis, we used narrative ratings from a set of 1,010 stop data forms completed in April 2014 of white and black community members. Narratives from ten stops were used for training purposes. Subsequently, each officer analyzed the narratives of 100 stops, first applying the coding scheme for severity over the course of one week. Out of the 1,010 stop data reports, at least two officers coded 300 of these stops in common so that we could establish inter-rater agreement. We obtained fair to moderate agreement (Kappa = 0.57-0.71) for the severity coding. We coded these values as a numerical variable in our regression models that ranged from 0 (*very minor*) to 3 (*very severe*).

# 2 Study 1

## 2.1 Annotator Agreement

For annotations, utterances were grouped together in “batches", such that ten annotators rated the same set of utterances. Annotators rated each utterance using a four-point biploar scale (*Very* or *Somewhat Impolite/Polite, Disrespectful/Respectful, Judgmental/Impartial*, and *Informal/Formal* ). We measured inter-annotator consistency (Cronbach’s $\alpha$) along each dimension of each batch (Table 4). Agreement varied depending on the batch and rating dimension, ranging from moderate ($\alpha$=.73) to high ($\alpha$=.91).

| 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 |

Table 4: Annotator consistency (Cronbach’s $\alpha$) across batches and dimension for the utterance-level thin-slice judgments in Study 1.

Participant ratings were averaged for each utterance to get a single rating for each utterance along these dimensions.

## 2.2 Model Outputs for Each Rated Dimension

Table 5 shows the results of linear mixed-effects models predicting score on each dimension as a function of the driver’s race, sex, and age (standardized), with random intercepts for each stop.

| | *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** | | | | | | | | | | | | | | | |
| $\sigma^2$ | | 0.17 | | | 0.19 | | | 0.21 | | | 0.22 | | | 0.25 | |
| $\tau_{00,\text{Stop}}$ | | 0.05 | | | 0.04 | | | 0.07 | | | 0.05 | | | 0.06 | |
| $N_{\text{Stop}}$ | | 251 | | | 251 | | | 251 | | | 251 | | | 251 | |
| $ICC_{\text{Stop}}$ | | 0.22 | | | 0.19 | | | 0.24 | | | 0.17 | | | 0.18 | |
| Observations | | 414 | | | 414 | | | 414 | | | 414 | | | 414 | |
| $R^2$ / $\Omega_0^2$ | | .52 / .39 | | | .48 / .35 | | | .56 / .42 | | | .47 / .33 | | | .47 / .34 | |

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

## 2.3 Principal Component Analysis (PCA) Loadings

To decompose participants' ratings into underlying components, we conducted a principal components analysis of their responses.

| | 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% |

Table 6: Loadings for the first two principal components (referred to throughout the paper as *Respect* and *Formality*) of the annotated ratings from Study 1.

## 2.4 Full Regression Model Output

| | *Respect* | | | *Formality* | | |
|---|---|---|---|---|---|---|
| | $\beta$ | CI | $p$ | $\beta$ | 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** | | | | | | |
| $\sigma^2$ | | 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 | |
| $N_{\text{Officer}}$ | | 118 | | | 118 | |
| $ICC_{\text{Stop:Officer}}$ | | 0.011 | | | 0.000 | |
| $ICC_{\text{Officer}}$ | | 0.132 | | | 0.110 | |
| Observations | | 414 | | | 414 | |
| $R^2$ / $\Omega_0^2$ | | .358 / .335 | | | .255 / .213 | |

Table 7: Mixed-effects regression outputs on observed ratings from participants in Study 1 for models with *Respect* and *Formality* (PC1 and PC2) as dependent variables; fixed effects for the community member's race, age, and gender; and random effects at the officer and interaction level. Reference levels are black male community members, a white officer, and a warning issued with no citation, arrest, or search. Standardized coefficients are reported. P-values computed via the Wald-statistics approximation with the sjPlot R Package [2].

# 3 Study 2

## 3.1 Linguistic Feature Engineering

In Table 3.1 in this section we provide a listing of every feature used in our models in Study 2, its implementation, and its source or justification. All features were implemented using the Python programming language (Python Software Foundation, https://www.python.org/) with word segmentation and syntactic annotations generated by version 3.4.1 of the publicly available Stanford CoreNLP toolkit [3].

<table>
<tr><th>Feature Name</th><th>Implementation</th><th>Source</th></tr>
<tr><td>Adverbial "Just"</td><td>"Just" occurs in a dependency arc as the head of an `advmod` relation</td><td></td></tr>
<tr><td>Apologizing</td><td>Lexicon: `"sorry", "oops", "woops", "excuse me", "forgive me", "apologies", "apologize", "my bad", "my fault"`</td><td>[4]</td></tr>
<tr><td>Ask for Agency</td><td>Lexicon: `"do me a favor", "let me", "allow me", "can i", "should i", "may i", "might i", "could i"`</td><td>[4]</td></tr>
<tr><td>Bald Command</td><td>The first word in a sentence is a bare verb with part-of-speech tag VB (`"look", "give", "wait"` etc.) but is not one of `"be", "do", "have", "thank", "please", "hang"`.</td><td></td></tr>
<tr><td>Colloquialism</td><td>Regular expression capturing `"y'all", "ain't"` and words ending in `"in'"` such as `"walkin'", "talkin'"`, etc., as marked by transcribers</td><td></td></tr>
<tr><td>Conditional</td><td>Lexicon: `"if"`</td><td></td></tr>
<tr><td>Disfluency</td><td>Word fragment ("Well I thi-") as indicated by transcribers</td><td>[5, 6]</td></tr>
<tr><td>Filled Pauses</td><td>Lexicon: `"um", "uh"`</td><td>[7, 8]</td></tr>
<tr><td>First Names</td><td>Top 1000 most common first names from the 1990 US Census, where first letter is capitalized in transcript</td><td>[9, 10][1]</td></tr>
<tr><td>Formal Titles</td><td>Lexicon: `"sir", "ma'am", "maam", "mister", "mr*", "ms*", "madam", "miss", "gentleman", "lady"`</td><td>[9, 10]</td></tr>
<tr><td>For Me</td><td>Lexicon: `"for me"`</td><td></td></tr>
<tr><td>For You</td><td>Lexicon: `"for you"`</td><td></td></tr>
<tr><td>Give Agency</td><td>Lexicon: `"let you", "allow you", "you can", "you may", "you could"`</td><td>[4]</td></tr>
<tr><td>Gratitude</td><td>Lexicon: `"thank", "thanks", "appreciate"`</td><td>[4]</td></tr>
<tr><td>Goodbye</td><td>Lexicon: `"goodbye", "bye", "see you later"`</td><td></td></tr>
<tr><td>Hands on the Wheel</td><td>Regular expression capturing cases like "keep your hands on the wheel" and "leave your hands where I can see them": `"hands?  ([.,?!:;]+ )?(wheel|see)"`</td><td></td></tr>
<tr><td>Hedges</td><td>All words in the "Tentat" LIWC lexicon</td><td>[11]</td></tr>
<tr><td>Impersonal Pronoun</td><td>All words in the "Imppron" LIWC lexicon</td><td>[4, 11]</td></tr>
<tr><td>Informal Titles</td><td>Lexicon: `"dude*", "bro*", "boss", "bud", "buddy", "champ", "man", "guy*", "guy", "brotha", "sista", "son", "sonny", "chief"`</td><td>[9, 10, 12]</td></tr>
<tr><td>Introductions</td><td>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))"`</td><td>[4]</td></tr>
<tr><td>Last Names</td><td>Top 5000 most common last names from the 1990 US Census, where first letter is capitalized in transcript</td><td>[9, 10][2]</td></tr>
<tr><td>Linguistic Negation</td><td>All words in the "Negate" LIWC lexicon</td><td>[11]</td></tr>
<tr><td>Negative Words</td><td>All words in the "Negativ" category in the Harvard General Inquierer, matching on word lemmas</td><td>[4, 13]</td></tr>
<tr><td>Positive Words</td><td>All words in the "Positiv" category in the Harvard General Inquierer, matching on word lemmas</td><td>[4, 13]</td></tr>
</table>

<table>
<tr><td>Please</td><td>Lexicon: <code>"please"</code></td><td>[4]</td></tr>
<tr><td>Questions</td><td>Occurrence of a question mark</td><td></td></tr>
<tr><td>Reassurance</td><td>Lexicon: <code>"'s okay", "n't worry", "no big deal", "no problem", "no worries", "'s fine", "you 're good", "is fine", "is okay"</code></td><td></td></tr>
<tr><td>Safety</td><td>Regular expression for all words beginning with the prefix "safe", such as <code>"safe", "safety", "safely"</code></td><td></td></tr>
<tr><td>Swear Words</td><td>All words in the "Swear" LIWC lexicon</td><td>[11]</td></tr>
<tr><td>Tag Question</td><td>Regular expression capturing cases like "..., right?" and "..., don't you?": <code>", (((all right|right|okay|yeah|please|you know)( sir| ma'am| miss| son)?)|((are|is|do|can|have|will|won't) (n't )?(i|me|she|us|we|you|he|they|them))) [?]"</code></td><td>[14, 15]</td></tr>
<tr><td>The Reason for the Stop</td><td>Lexicon: <code>"reason", "stop* you", "pull* you", "why i", "why we", "explain", "so you understand"</code></td><td></td></tr>
<tr><td>Time Minimizing</td><td>Regular expression capturing cases like "in a minute" and "let's get this done quick": <code>"(a|one|a few) (minute|min|second|sec|moment)s?|this[␣,?!]+quick|right back"</code></td><td></td></tr>
</table>

[2] https://www2.census.gov/topics/genealogy/1990surnames/

## 3.2 Full Regression Model Output

| | *Respect* | | | *Formality* | | |
|---|---|---|---|---|---|---|
| | $\beta$ | CI | p | $\beta$ | CI | p |
| **Fixed Parts** | | | | | | |
| (Intercept) | -0.18 | -0.36 – 0.00 | .052 | 0.26 | 0.07 – 0.45 | **.008** |
| Adverbial "Just" | 0.24 | -0.07 – 0.53 | .118 | | | |
| Apologizing | 1.34 | 0.15 – 2.52 | **.027** | -1.56 | -2.80 – -0.32 | **.014** |
| Ask for Agency | -0.34 | -0.90 – 0.22 | .230 | 0.37 | -0.23 – 0.96 | .225 |
| Bald Commands | | | | -0.25 | -0.68 – 0.18 | .255 |
| Colloquialism | | | | -1.10 | -1.97 – -0.23 | **.013** |
| Conditional | | | | -0.27 | -0.74 – 0.21 | .271 |
| Disfluency | -0.36 | -0.63 – -0.09 | .009 | | | |
| Filled Pauses (Um/Uh) | 0.37 | 0.14 – 0.60 | **.002** | -0.40 | -0.64 – -0.16 | **.001** |
| First Names | -0.88 | -1.66– -0.11 | **.026** | | | |
| Formal Titles | 0.73 | 0.20 – 1.26 | **.007** | 0.96 | 0.43 – 1.49 | $<$**.001** |
| For Me | 0.56 | -0.08 – 1.21 | .086 | | | |
| For You | 1.08 | -0.70 – 2.87 | .234 | -1.26 | -3.10 – 0.58 | .178 |
| Give Agency | 0.39 | 0.01 – 0.78 | **.047** | 0.40 | -0.02 – 0.82 | .063 |
| Gratitude | 1.04 | 0.44 – 1.64 | $<$**.001** | | | |
| Hands on the Wheel | -1.09 | -2.27 – 0.07 | .065 | 1.33 | 0.10 – 2.55 | **.034** |
| Hedges | 0.18 | 0.00 – 0.37 | .053 | | | |
| Impersonal Pronouns | | | | -0.10 | -0.27 – 0.07 | .269 |
| Informal Titles | -0.65 | -1.03 – -0.28 | $<$**.001** | -1.06 | -1.45 – -0.68 | $<$**.001** |
| Introductions | 0.18 | -0.12 – 0.48 | .235 | | | |
| Last Names | 0.75 | 0.39 – 1.12 | $<$**.001** | 0.26 | -0.10 – 0.62 | .156 |
| Linguistic Negation | -0.22 | -0.43 – -0.03 | **.027** | 0.22 | 0.01 – 0.43 | **.045** |
| Negative Words | -0.24 | -0.40 – -0.07 | **.005** | -0.17 | -0.34 – 0.01 | .056 |
| Positive Words | 0.20 | 0.03 – 0.37 | **.020** | -0.16 | -0.32 – 0.00 | .056 |
| Questions | -0.20 | -0.43 – 0.02 | .075 | 0.26 | 0.02 – 0.49 | **.031** |
| Reassurance | 1.04 | 0.34 – 1.74 | **.004** | -0.73 | -1.46 – 0.00 | **..049** |
| Safety | 0.54 | 0.06 – 1.02 | **.027** | | | |
| The Reason for the Stop | | | | 0.41 | 0.08 – 0.75 | **.015** |
| Time Minimizing | | | | -0.66 | -1.31 – 0.00 | **.049** |
| Observations | | 414 | | | 414 | |
| $R^2$ / $\Omega_0^2$ | | .298 / .258 | | | .229 / .190 | |

Table 9: Linear regression outputs, with stepwise feature selection by $R^2$, for all annotated utterances with *Respect* and *Formality* (PC1 and PC2) as dependent variables and utterance-level log counts of linguistic features as independent variables. The swear words, please, goodbye, and tag question features were selected out in both models.

### 3.3 Other Model Possibilities

Though we report results using simple linear regression models, we note that we also tried using this same set of features with more complex machine learning algorithms including lasso regression, support vector regression, and random forest regression, none of which exceeded the performance of the models reported here.

### 3.4 Model Comparison to Annotators

The method for the RMSE comparison between our model and the human annotators in Study 2 in the paper is as follows. Each of our 70 annotators was part of a batch of annotators who annotated the same set of around 60 utterances. We converted each annotator's set of 5 ratings for a given utterance to the two PCA dimensions - Respect and Formality. For a given annotator, for every utterance they annotated we calculated the average rating on each dimension for all the other annotators in their batch. We could then treat the average rating from the other annotators as a "gold" label, and calculate each annotator's error with respect to all the others.

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

Table 10: Human RMSE scores for *Respect* and *Formality* across annotators relative to other annotators.

These numbers establish a comparative context in which to understand the RMSE scores of our *Respect* (0.840) and *Formality* (0.882) models, which are calculated across the entire dataset with reference to the average annotator ratings across all 414 utterances.

### 3.5 Linguistic Implications

The main body of the paper addresses the primary goal of this work, namely, the question of whether racial disparities in respect can be observed in officer language. The linguistic features used and computational models developed are in some sense secondary tools in the service of this goal. We found these models to be of a sufficient predictive accuracy overall, but we caution the reader against accepting the results for any individual feature as definitive since our training set for these models is relatively small (414 utterances). However, the linguistic findings of this study are still interesting on their own for several reasons.

Existing work on politeness in linguistics has tended to focus on the character of face-threatening acts in general [16] and the case of requests in particular [4, 17, 18], as well as cross-cultural differences that emerge in politeness strategies [19–21]. In contrast, the police-community interactions captured on body camera footage we work with in this study constitute data from a unique domain that is heretofore unstudied linguistically.

Many of the features selected in the models patterned as we might expect given the predictions of the linguistic politeness literature. We find positive politeness strategies - aimed at showing the speaker values the hearer and their self-image - are generally perceived as respectful in this context, including introductions, reassurance, gratitude, safety, and referential politeness like formal titles and last names; the inverse features of informal titles and first names are in turn perceived as disrespectful. Negative politeness strategies - aimed at mitigating the magnitude of the imposition - are perceived as respectful, including apologizing (which is our top weighted feature) and numerous forms of softening (hedges, framing with "for you" and "for me", and adverbial "just"). In line with [4], we find that in general positive words and greetings contribute to respect while negative words and questions contribute to disrespect.

This study demonstrates the distinction between politeness in the traditional theoretical linguistic sense and perceived respect in this particular context. We propose this distinction may have much to do with the set of expectations which accompany any involuntary police stop. For instance, linguistic theories of politeness might predict that giving agency ("let you", "you can") and asking for agency ("can I", "may I") would both be perceived as respectful since on the surface they empower the hearer. However, we found that asking for agency is selected with a negative weight in our model, associated with disrespect. It may be that since in the context of a police-community interaction, all parties know that the officer is the one with the hierarchical position of power, requests for agency are perceived as disingenuous; everyone knows the officer has agency available and does not need to ask for it.

Finally, we include features potentially having to do with officers' comfort or anxiety, including speech disfluencies ("th- that"), filled pauses ("um" and "uh"), and direct commands to "keep your hands on the wheel". We find "hands on the wheel" and speech disfluencies are perceived as disrespectful, while filled pauses are respectful. This result is in line with existing research such as [5] who found the production of all manner of disfluencies increased with anxiety in talk with the exception of filled pauses. Filled pauses, on the other hand, have been argued to be conventional words that are planned by speakers with particular interactional functions [8]. These results suggest that unlike traditional theories of politeness, in this context perceived respect is not only about the choice of linguistic strategies but also contextual and emotive factors like an officer's anxiety level.

### 3.6 Respect vs. Formality

Our findings also highlight a contrast between officers' respect and formality. While the referential features (titles and the use of names) largely pattern in the same direction in both models, several politeness strategies (apologizing, reassurance, softening with "for you") are perceived as respectful, but informal. At the same time, certain other linguistic features which might traditionally be considered to be associated with respect (softening with tag questions, colloquialisms, minimizing the time imposition such as "real quick") were perceived as informal but not relevant for respect.

Furthermore, giving the reason for the stop is not only important for procedural justice considerations, but is also required by department policy and is in fact more common with white community members (log odds ratio of 0.349, $p < 0.001$); however, it is only selected in our model for formality, but not for respect.

### 3.7 Linguistic Classification Accuracy of Race

We mention in the discussion of the main paper that the results we uncover are only one part of a number of diverse linguistic differences between officer language in talking to white versus black community members. Specifically, we show that a simple classifier trained only on officer language is able to predict the race of the community member to whom an utterance was directed at much higher than chance performance. In this section we briefly describe that model.

We first take a random balanced subsample of our dataset to allow for equal comparison; our final dataset contains 13,910 utterances directed towards white community members, so we sample 13,910 utterances towards black community members for a balanced dataset of 27,820 utterances. For each of these, we extract all the linguistic features described in the paper, as well as n-grams up to length three. That is, every window of words up to length three that occurs in any utterance is extracted as a feature. We then select the 5,000 most informative features using the chi squared criterion, and train logistic regression classifiers to predict race based on the features in an utterance.

We perform a hyperparameter search for this model with training set comprising 80% of the data and a development set containing an additional 10% of the data on each fold in a 10-fold cross validation scheme. We find a regularization strength of 1 and an $l2$ regularization penalty to provide the strongest performance on this development set. We test this model in a 10-fold cross validation scheme on the previously held-out 10% of each fold, training on the training set of that fold.

With this model we find a mean performance on the test sets in our 10-fold cross validation of 67.6%, compared to a most-common-class baseline of 50%. This result again confirms the finding that there are significant, observable differences in officer speech based on the race of the community member.

# 4 Study 3

## 4.1 Full Regression Model Output

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

| | *Respect* | | | *Formality* | | |
|---|---|---|---|---|---|---|
| | $\beta$ | CI | p | $\beta$ | 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** | | | | | | |
| $\sigma^2$ | | 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 | |
| $N_{\text{Officer}}$ | | 245 | | | 245 | |
| $ICC_{\text{Stop:Officer}}$ | | 0.045 | | | 0.029 | |
| $ICC_{\text{Officer}}$ | | 0.029 | | | 0.015 | |
| Observations | | 36738 | | | 36738 | |
| $R^2$ / $\Omega_0^2$ | | .100 / .097 | | | .064 / .059 | |

Table 11: Linear mixed-effects model outputs on computationally-generated ratings on all utterances in the dataset for models with *Respect* and *Formality* (PC1 and PC2) as dependent variables; fixed effects for the community member's race, age, and gender, as well as whether a search was conducted, whether a citation was issued, whether an arrest occurred, the race of the officer (Black, White, or Other), and an interaction effect between community member race and officer race; and random effects at the officer and interaction level. P-values computed via the Wald-statistics approximation with the sjPlot R Package [2]. Reference levels are black male community members, white officers, and no arrest, citation, or search.

## 4.2 *Respect* and *Formality* Over Time

To analyze how *Respect* and *Formality* varied over time, we regressed each score separately in a linear mixed-effects model, with driver race, utterance position in the interaction (scaled from 0 to 1), and the interaction of these terms as fixed effects. The trajectory was allowed to vary across interactions by including a random slope of utterance position within each stop. A comparison of this model with a random intercept-only model revealed that the trajectory of *Respect* over time varied significantly, $\chi^2(2) = 127.08$, p <.001. However, a similar random slope model predicting *Formality* failed to converge; as a result, we fit a random intercept-only model for this outcome. The results of these analyses are given below.[3]

| | *Respect* | | | *Formality* | | |
|---|---|---|---|---|---|---|
| | *b* | CI | p | *b* | CI | p |
| **Fixed Parts** | | | | | | |
| Intercept | 0.05 | 0.01 – 0.08 | **<.001** | 0.00 | -0.02 – 0.02 | .72 |
| Race (W) | 0.20 | 0.15 – 0.25 | **<.001** | 0.00 | -0.04 – 0.04 | .88 |
| Utterance Position (mean-centered) | 0.24 | 0.19 – 0.29 | **<.001** | -0.48 | -0.52 – -0.45 | **<.001** |
| Utterance Position: Race (W) | 0.20 | 0.10 – 0.31 | **<.001** | -0.18 | -0.27 – -0.10 | **<.001** |
| **Random Parts** | | | | | | |
| $\sigma^2$ | | 0.90 | | | 0.93 | |
| $\tau_{00,\text{Stop}}$ | | 0.09 | | | 0.05 | |
| $\tau_{11,\text{Utterance Position}}$ | | 0.23 | | | | |
| $cor_{\tau_{00},\tau_{11}}$ | | -0.24 | | | | |
| $N_{\text{Stop}}$ | | 981 | | | 981 | |
| $ICC_{\text{Stop}}$ | | 0.09 | | | 0.05 | |
| Observations | | 36,738 | | | 36,738 | |
| $R^2$ / $\Omega_0^2$ | | .13 / .12 | | | .09 / .08 | |

[3] While estimates of lower-order effects of race and utterance position are estimated using effects coding (black= -1, white= 1) in the body of the paper, we dummy code race here (black= 0, white= 1) for consistency with other models reported in this supplement.

### 4.3 Raw Respect Means

For reference, in this section we provide figures depicting the raw estimated respect level from our computational annotations across the cells of Table 2 in Section 1.3, representing different community member attributes and stop outcomes.

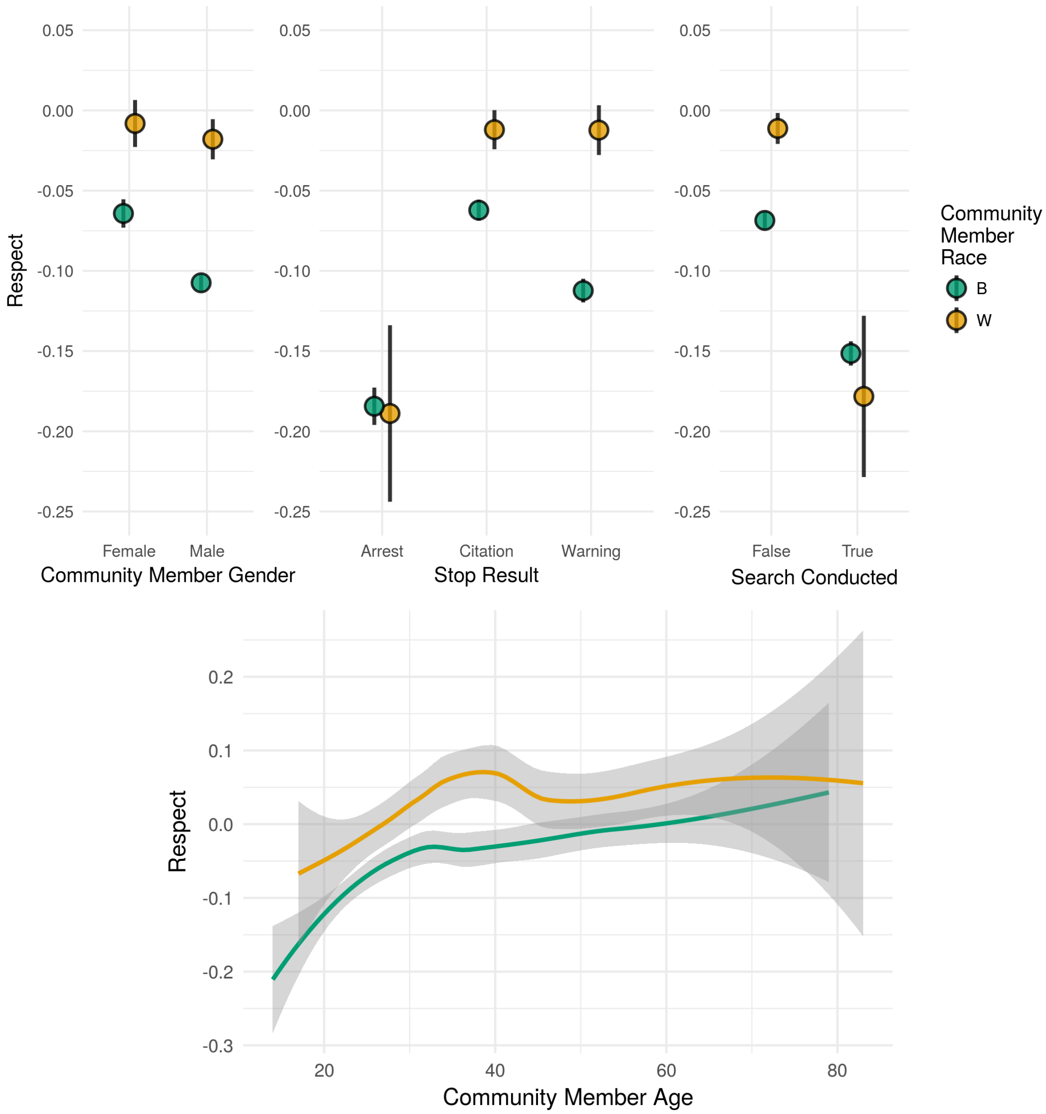


Figure 1: Raw mean estimated respect levels across different community member attributes.

## 4.4 Alternative Models

In addition to the model presented above in Section 4.1, we ran several additional models to include other possible variables that could confound our results. We find that none of these alter the significant effect of community member race on respect.

### 4.4.1 Additional Covariates

| | *Respect* | | | *Formality* | | |
|---|---|---|---|---|---|---|
| | $\beta$ | CI | p | $\beta$ | CI | p |
| **Fixed Parts** | | | | | | |
| Arrest Occurred | 0.00 | -0.03 – 0.04 | .938 | 0.01 | -0.02 – 0.04 | .531 |
| Citation Issued | 0.04 | 0.01 – 0.06 | **.002** | 0.01 | -0.01 – 0.03 | .256 |
| Search Conducted | -0.08 | -0.11 – -0.05 | <**.001** | 0.00 | -0.03 – 0.02 | .933 |
| 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 | **.032** |
| Race (W) | 0.05 | 0.03 – 0.08 | <**.001** | -0.01 | -0.04 – 0.01 | .281 |
| Crime Rate in Census Tract | 0.01 | -0.01 – 0.04 | .278 | 0.03 | 0.01 – 0.05 | **.014** |
| Businesses per Square Mile | 0.00 | -0.02 – 0.03 | .702 | -0.01 | -0.03 – 0.01 | .222 |
| Race Known Before Stop | -0.01 | -0.04 – 0.01 | .255 | 0.01 | -0.01 – 0.02 | .612 |
| Officer Years of Experience | 0.00 | -0.03 – 0.03 | .831 | 0.00 | -0.03 – 0.02 | .754 |
| Officer Race (B) | 0.00 | -0.03 – 0.04 | .795 | 0.00 | -0.03 – 0.03 | .939 |
| Officer Race (O) | -0.01 | -0.04 – 0.03 | .741 | 0.00 | -0.03 – 0.02 | .761 |
| Officer Race (B) : Race (W) | -0.01 | -0.03 – 0.01 | .471 | 0.01 | -0.01 – 0.03 | .225 |
| Officer Race (O) : Race (W) | -0.01 | -0.03 – 0.02 | .470 | 0.00 | -0.02 – 0.02 | .882 |
| **Random Parts** | | | | | | |
| $\sigma^2$ | | 0.919 | | | 0.954 | |
| $\tau_{00,\text{Stop:Officer}}$ | | 0.046 | | | 0.030 | |
| $\tau_{00,\text{Officer}}$ | | 0.029 | | | 0.015 | |
| $N_{\text{Stop:Officer}}$ | | 965 | | | 965 | |
| $N_{\text{Officer}}$ | | 241 | | | 241 | |
| $ICC_{\text{Stop:Officer}}$ | | 0.046 | | | 0.030 | |
| $ICC_{\text{Officer}}$ | | 0.029 | | | 0.015 | |
| Observations | | 36137 | | | 36137 | |
| $R^2$ / $\Omega_0^2$ | | .101 / .098 | | | .065 / .060 | |

Table 12: Model from Section 4.1 with additional control variables: officer years of experience, crime rate in the census tract where the stop took place, businesses per mile in the census tract, and whether the officer marked that they were aware of the community member's race before stopping them.

### 4.4.2 Accounting for Infraction Severity

To ensure that differences in Respect were not due to differences in the severity of the traffic offense, we ran a model including severity as a covariate on the subset of 869 stops annotated by officers as “Equipment" or “Moving Violation" stops.

| | *Respect* | | | *Formality* | | |
|---|---|---|---|---|---|---|
| | $\beta$ | CI | p | $\beta$ | CI | p |
| **Fixed Parts** | | | | | | |
| Arrest Occurred | -0.01 | -0.04 – 0.03 | .708 | -0.04 | -0.06 – -0.01 | **.007** |
| Citation Issued | 0.03 | 0.00 – 0.05 | **.020** | 0.01 | -0.01 – 0.03 | .419 |
| Search Conducted | -0.06 | -0.08 – -0.03 | **<.001** | 0.00 | -0.02 – 0.02 | .971 |
| Age | 0.07 | 0.05 – 0.09 | **<.001** | 0.04 | 0.03 – 0.06 | **<.001** |
| Gender (F) | 0.02 | 0.00 – 0.04 | .125 | 0.01 | 0.00 – 0.03 | .139 |
| Race (W) | 0.05 | 0.02 – 0.08 | **<.001** | -0.02 | -0.04 – 0.01 | .195 |
| Severity | 0.00 | -0.02 – 0.02 | .945 | 0.00 | -0.02 – 0.02 | .893 |
| Officer Race (B) | 0.01 | -0.02 – 0.05 | .467 | 0.00 | -0.03 – 0.03 | .847 |
| Officer Race (O) | 0.00 | -0.04 – 0.04 | .905 | -0.01 | -0.04 – 0.02 | .669 |
| Officer Race (B) : Race (W) | -0.01 | -0.03 – 0.02 | .508 | 0.01 | -0.01 – 0.04 | .219 |
| Officer Race (O) : Race (W) | -0.01 | -0.03 – 0.02 | .615 | 0.00 | -0.02 – 0.03 | .788 |
| **Random Parts** | | | | | | |
| $\sigma^2$ | | 0.964 | | | 0.977 | |
| $\tau_{00,\text{Stop:Officer}}$ | | 0.044 | | | 0.026 | |
| $\tau_{00,\text{Officer}}$ | | 0.032 | | | 0.017 | |
| $N_{\text{Stop:Officer}}$ | | 869 | | | 869 | |
| $N_{\text{Officer}}$ | | 220 | | | 220 | |
| $ICC_{\text{Stop:Officer}}$ | | 0.043 | | | 0.026 | |
| $ICC_{\text{Officer}}$ | | 0.031 | | | 0.017 | |
| Observations | | 28786 | | | 28786 | |
| $R^2$ / $\Omega_0^2$ | | .095 / .090 | | | .065 / .060 | |

Table 13: Model from Section 4.1 including a variable for severity of the infraction on the subset of the dataset for which we have these ratings annotated by OPD officers. Severity ratings were recoded on an increasing scale from 0 (very minor infraction) to 3 (very severe infraction).

### 4.4.3 "Everyday" Stops

As seen in Section 1.3, stops with black drivers are much more likely than those with white drivers to involve an arrest or search; therefore, to confirm that the effect we find is not only a side effect of these more charged circumstances, we run our model on a subset of the data including only "everyday" traffic stops in which no arrest or search occurred.

| | *Respect* | | | *Formality* | | |
|---|---|---|---|---|---|---|
| | $\beta$ | CI | p | $\beta$ | CI | p |
| **Fixed Parts** | | | | | | |
| Citation Issued | 0.04 | 0.01 – 0.06 | **<.002** | 0.00 | -0.02 – 0.02 | .922 |
| Age | 0.07 | 0.05 – 0.09 | **<.001** | 0.05 | 0.03 – 0.07 | **<.001** |
| Gender (F) | 0.02 | 0.00 – 0.04 | .109 | 0.02 | 0.00 – 0.03 | .103 |
| Race (W) | 0.06 | 0.03 – 0.09 | **<.001** | -0.02 | -0.04 – 0.01 | .199 |
| Officer Race (B) | 0.01 | -0.03 – 0.05 | .554 | 0.00 | -0.03 – 0.04 | .750 |
| Officer Race (O) | 0.00 | -0.04 – 0.04 | .890 | -0.01 | -0.04 – 0.02 | .609 |
| Officer Race (B) : Race (W) | -0.01 | -0.04 – 0.02 | .459 | 0.01 | -0.01 – 0.04 | .222 |
| Officer Race (O) : Race (W) | -0.01 | -0.04 – 0.02 | .423 | 0.00 | -0.03 – 0.03 | .963 |
| **Random Parts** | | | | | | |
| $\sigma^2$ | | 0.946 | | | 0.936 | |
| $\tau_{00,\text{Stop:Officer}}$ | | 0.047 | | | 0.027 | |
| $\tau_{00,\text{Officer}}$ | | 0.032 | | | 0.015 | |
| $N_{\text{Stop:Officer}}$ | | 864 | | | 864 | |
| $N_{\text{Officer}}$ | | 221 | | | 221 | |
| $ICC_{\text{Stop:Officer}}$ | | 0.046 | | | 0.027 | |
| $ICC_{\text{Officer}}$ | | 0.031 | | | 0.016 | |
| Observations | | 26270 | | | 26270 | |
| $R^2$ / $\Omega_0^2$ | | .099 / .093 | | | .064 / .056 | |

Table 14: Model from Section 4.1 on only stops in which no arrest was made and no search occurred.

#### 4.4.4 Accounting for Racial Homophily

Our main model in Section 4.1 includes variables for community and officer race as well as the interaction between these. However, this model may not capture potential effects of racial homophily. Might officers communicate more respect towards community members of their own race (white officers with white community members, for example, or black officers with black community members)?

| | *Respect* | | | *Formality* | | |
|---|---|---|---|---|---|---|
| | $\beta$ | CI | p | $\beta$ | CI | p |
| **Fixed Parts** | | | | | | |
| Arrest Occurred | 0.00 | -0.03 – 0.03 | .925 | 0.01 | -0.02 – 0.04 | .533 |
| Citation Issued | 0.04 | 0.02 – 0.06 | <**.001** | 0.01 | -0.01 – 0.03 | .186 |
| Search Conducted | -0.08 | -0.11 – -0.05 | <**.001** | 0.00 | -0.03 – 0.02 | .862 |
| Age | 0.07 | 0.05 – 0.09 | <**.001** | 0.05 | 0.03 – 0.07 | <**.001** |
| Gender (F) | 0.02 | 0.00 – 0.04 | .059 | 0.02 | 0.00 – 0.04 | **.021** |
| Race (W) | 0.04 | 0.01 – 0.07 | **.003** | -0.01 | -0.03 – 0.02 | .590 |
| Race Homophily (T) | 0.00 | -0.03 – 0.04 | .842 | 0.00 | -0.03 – 0.02 | .806 |
| Race Homophily (T) : Race (W) | 0.01 | -0.03 – 0.05 | .677 | 0.00 | -0.03 – 0.03 | .867 |
| **Random Parts** | | | | | | |
| $\sigma^2$ | | 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 | |
| $N_{\text{Officer}}$ | | 245 | | | 245 | |
| $ICC_{\text{Stop:Officer}}$ | | 0.045 | | | 0.029 | |
| $ICC_{\text{Officer}}$ | | 0.029 | | | 0.015 | |
| Observations | | 36738 | | | 36738 | |
| $R^2$ / $\Omega_0^2$ | | .100 / .097 | | | .064 / .059 | |

Table 15: Model from Section 4.1 substituting the variable for race of the officer with a variable for race homophily (community member and officer race are the same).

# References


[1] Gelman A, Hill J (2006) *Data analysis using regression and multilevel/hierarchical models.* (Cambridge university press).

[2] Lüdecke D (2015) sjPlot: Data visualization for statistics in social science—R package version 1.8. 1.

[3] Manning CD et al. (2014) The Stanford CoreNLP natural language processing toolkit. in *ACL (System Demonstrations).* pp. 55–60.

[4] Danescu-Niculescu-Mizil C, Sudhof M, Jurafsky D, Leskovec J, Potts C (2013) A computational approach to politeness with application to social factors in *Proceedings of ACL.*

[5] Kasl SV, Mahl GF (1965) Relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of personality and social psychology* 1(5):425.

[6] Mahl GF (1987) Everyday disturbances of speech in *Language in psychotherapy.* (Springer), pp. 213–269.

[7] Christenfeld N (1994) Options and ums. *Journal of Language and Social Psychology* 13(2):192–199.

[8] Clark HH, Fox Tree JE (2002) Using uh and um in spontaneous speaking. *Cognition* 84(1):73–111.

[9] Brown R, Ford M (1961) Address in American English. *The Journal of Abnormal and Social Psychology* 62(2):375.

[10] Krishnan V, Eisenstein J (2015) "You're Mr. Lebowski, I'm the Dude": Inducing address term formality in signed social networks in *Proceedings of NAACL.*

[11] Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology* 29(1):24–54.

[12] Kiesling SF (2004) Dude. *American speech* 79(3):281–305.

[13] Stone PJ, Bales RF, Namenwirth JZ, Ogilvie DM (1962) The General Inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science* 7(4):484–498.

[14] Holmes J (1984) Modifying illocutionary force. *Journal of pragmatics* 8(3):345–365.

[15] Cameron D, McAlinden F, O'Leary K (1988) Lakoff in context: The social and linguistic functions of tag questions. *Women in their speech communities* pp. 74–93.

[16] Brown P, Levinson SC (1978) Universals in language usage: Politeness phenomena in *Questions and politeness: Strategies in social interaction.* (Cambridge University Press), pp. 56–311.

[17] Clark HH, Schunk DH (1980) Polite responses to polite requests. *Cognition* 8(2):111–143.

[18] Blum-Kulka S (1987) Indirectness and politeness in requests: Same or different? *Journal of pragmatics* 11(2):131–146.

[19] Holtgraves T, Joong-nam Y (1990) Politeness as universal: Cross-cultural perceptions of request strategies and inferences based on their use. *Journal of personality and social psychology* 59(4):719.

[20] Reiter RM (2000) *Linguistic politeness in Britain and Uruguay: A contrastive study of requests and apologies.* (John Benjamins Publishing) Vol. 83.

[21] Lakoff RT, Ide S (2005) *Broadening the horizon of linguistic politeness.* (John Benjamins Publishing) Vol. 139.