



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

Eugenia H. Rho^{a,1} , Maggie Harrington^b, Yuyang Zhong^c , Reid Pryzant^d, Nicholas P. Camp^e, Dan Jurafsky^{d,e,1} , and Jennifer L. Eberhardt^{b,f,1}

Contributed by Jennifer Eberhardt; received September 29, 2022; accepted January 28, 2023; reviewed by Tracey Meares and Cristian Danescu-Niculescu-Mizil

Across the United States, police chiefs, city officials, and community leaders alike have highlighted the need to de-escalate police encounters with the public. This concern about escalation extends from encounters involving use of force to routine car stops, where Black drivers are disproportionately pulled over. Yet, despite the calls for action, we know little about the trajectory of police stops or how escalation unfolds. In study 1, we use methods from computational linguistics to analyze police body-worn camera footage from 577 stops of Black drivers. We find that stops with escalated outcomes (those ending in arrest, handcuffing, or a search) diverge from stops without these outcomes in their earliest moments—even in the first 45 words spoken by the officer. In stops that result in escalation, officers are more likely to issue commands as their opening words to the driver and less likely to tell drivers the reason why they are being stopped. In study 2, we expose Black males to audio clips of the same stops and find differences in how escalated stops are perceived: Participants report more negative emotion, appraise officers more negatively, worry about force being used, and predict worse outcomes after hearing only the officer's initial words in escalated versus non-escalated stops. Our findings show that car stops that end in escalated outcomes sometimes begin in an escalated fashion, with adverse effects for Black male drivers and, in turn, police–community relations.

policing | race | escalation | natural language processing (NLP) | body-worn cameras

The killing of George Floyd, after a Minneapolis police officer forcibly removed him from his car for using a counterfeit \$20 bill at a local store, led to the largest racial justice movement of the 21st century (1). Millions bore witness to the power of the camera to capture police–community interactions as they go awry (2). This filmed incident, perhaps more than any other, highlighted the need to reimagine public safety and renewed calls for police de-escalation training nationwide (3).

Yet, concerns about racial justice and police escalation can arise even when no force is used, and even in everyday, police–community encounters. The most common way members of the public come into contact with police is through car stops. Nearly 18.7 million drivers are stopped by police each year in the United States (4). Yet, not all groups share the same experience. In addition to being stopped at higher rates, Black drivers are more likely to be handcuffed, searched, and arrested than any other racial demographic (4–7). In fact, their stops are more likely to include these escalated outcomes even though they are much more likely than White drivers to be stopped for purely discretionary reasons (e.g., expired registration, broken license plate light) that pose little threat to public safety (8).

From the very beginning of the stop, then, Black drivers may experience fear and anxiety about what could unfold, regardless of the reason for the stop. *Will I be handcuffed? Will I be searched? Will I be arrested?* What cues in the interaction might Black drivers use to assess the probability of such outcomes? And to what extent might such outcomes increase Black drivers' worry about force? The current studies were designed to address these questions.

Much of what we know about Black Americans' attitudes toward policing involves their general views and experiences relative to other groups. For example, according to a 2021 Gallup Poll, only 27% of Black adults express “a great deal of confidence” in police, as compared to 56% of White adults (9). According to a recent Pew Survey, Black adults are also significantly more likely to report being subjected to unfair police stops because of their race (44%) than White adults (9%) (10).

Yet, we know little about how Black drivers experience police contact in the moment. During a police stop, to what are Black drivers attuned as the interaction begins—The officer's demeanor? Their own emotions? Their concerns about how the stop will end? Is there a linguistic signature to escalated stops—a manner in which officers speak to drivers that can telegraph the outcome of a stop well before that outcome occurs? How quickly does this signature emerge? And how quickly can drivers assess it?

Significance

Amid calls for police officers to de-escalate encounters with Black citizens, this work sheds light on when and how car stops escalate, as well as their psychological impact on Black men. Our analysis of police body-worn camera footage reveals that stops ultimately resulting in escalation differ in their conversational structure in the earliest moments of the encounter: in as little as the first 45 words the officer speaks. Listening to these escalated encounters evoked anxiety, suspicion, and worry about officer use of force for Black men, who are disproportionately subjected to escalated outcomes. The findings reported here not only inform approaches to de-escalation but also demonstrate the power and promise of systematic footage review more broadly. To improve police–community interactions, we could start by examining them.

Author contributions: E.H.R., R.P., D.J., and J.L.E. designed study 1 research; E.H.R. and R.P. performed research and analyzed the data for study 1; M.H., Y.Z., N.P.C., E.H.R., D.J., and J.L.E. designed study 2 research; M.H., Y.Z., and N.P.C. performed research and analyzed the data for study 2; E.H.R., M.H., N.P.C., R.P., D.J., and J.L.E. wrote the paper; and D.J. and J.L.E. served as PI on this project.

Reviewers: T.M., Yale University; and C.D.-N.-M., Cornell University.

The authors declare no competing interest.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: eugenia@vt.edu, jleberhardt@stanford.edu, or jurafsky@stanford.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2216162120/-/DCSupplemental>.

Published May 30, 2023.

Such questions have been nearly impossible to answer in the past. Now, with the widespread use of police body-worn cameras, we can begin to track the linguistic signature of escalated stops for the first time, as well as measure Black drivers' reactions to the language officers use in those stops.

The footage from these cameras allows us to directly observe each interaction as it progresses. We can play the recordings from these stops for others, effectively placing them in the seat of the driver—exposing them to what the driver heard. We can computationally analyze officers' language across hundreds (sometimes thousands) of cases to look for patterns. The camera not only allows us to lift the lid on what happens in horrific cases like Floyd but also on the mundane and ordinary, to understand the extent to which they too are fraught.

In the current work, we ask whether police stops involving a driver who is eventually searched, handcuffed, or arrested—what we call “escalated” interactions—are distinguished in the earliest moments of the encounter from non-escalated interactions (where none of these police actions occur). In study 1, we use natural language processing (NLP) models to determine how officers speak differently across escalated and non-escalated stops. In study 2, we examine how Black drivers react differently to these two classes of stops.

Data

We began with 588 vehicle stops of Black (192 F, 396 M) drivers conducted over the course of 1 mo in a medium-sized racially diverse US city. Nearly one out of six (15.3%) Black drivers stopped in this month were either arrested, searched, or handcuffed, whereas less than 1% of White drivers experienced one of these outcomes. Given that 97.8% of the drivers who experienced one of the three escalated outcomes were Black, we limit our sample of analysis to Black drivers only (*SI Appendix, Study 1 Description of Data and Sampling*).

Each car stop, which captures the interaction between the driver and the primary officer conducting the stop, was recorded on a police body-worn camera, professionally transcribed, and diarized. Of the 588 stops of Black drivers, we eliminated the only two stops that involved police use of force, as our work focuses on escalated stops that are relatively routine. In addition, we eliminated nine stops that included words or phrases directly connotative of the stop outcomes in the initial words spoken by the officer (e.g., “arrest”, “handcuff”, “I’m going to conduct a search”, etc., for details, see *Materials and Methods* and *SI Appendix, Study 1 Description of Data and Sampling*).

This final sampling criterion resulted in a total of 577 stops of Black drivers (191 F, 386 M) for our analyses (see *Table S2*, *SI* for

officer and driver demographics). Of the 577 stops, there were 76 searches, 56 handcuffings, and 24 arrests conducted across a total of 81 stops (10 F, 71 M) resulting in one of these three escalated outcomes. 81.6% (62 of 76) of the searches conducted on Black drivers did not result in the discovery of any firearms, drugs, or other incriminating evidence.

Study 1: Linguistic Signature of Officer Speech in Escalated Stops of Black Drivers

We first consider how officers communicate with Black drivers in the earliest moments of escalated and non-escalated stops by drawing on a linguistic model of one aspect of dialog structure: the *institutional dialog act* (11). Dialog acts are an extension of the traditional linguistic notion of speech acts and represent both the speaker’s intent (to ask a question, to issue a command, etc.) and the role the utterance plays in the surrounding dialog context, such as the expectations that it sets up for what comes next in the dialog. Institutional dialog acts are further enriched by the notion that in institutional settings, the acts that speakers choose are sensitive to their institutional roles (e.g., officer–driver, teacher–student, doctor–patient) and the tasks they are performing (12). In the context of car stops, police officers, for example, ask for documents, state the reason for the stop, or issue sanctions. How does the distribution of such dialog acts differ between escalated and non-escalated stops?

Where prior studies of police–citizen interactions have generally considered broad interactional variables like respect (13) or accommodation (14, 15), our approach captures whether, when, and in what sequence concrete institutional speech acts occur. This offers a more fine-grained and dynamic structural model of the interaction, tracking the temporal progression of vehicle stops through different acts. In so doing, we capture aspects of these interactions that are central to procedural justice (16–19): whether officers are transparent in providing the reason for a stop early in the interaction, or whether officers press drivers with accusatory questions or authoritative commands.

Specifically, we examine six institutional dialog acts developed by prior work (11), which analyzed body-worn camera transcripts of vehicle stops from a medium-sized police department in a racially diverse city. These dialog acts were developed based on a data-driven analysis of vehicle stop transcripts together with the procedural justice and conversational analysis literatures (14, 15, 20–23). Examples of each act from our data are shown in *Table 1*.

We use these six institutional dialog acts as a window into officer language in the earliest moments of car stop interactions: the first 45 words spoken by the officer to the driver during each stop. On average, these first 45 words constitute the initial 5% of the words

Table 1. Examples of institutional dialog acts from officer speech

	Institutional Dialog Act	Examples From Officer Speech
1	<i>Greeting</i> : greeting or identifying themselves as law enforcement	Hi, good morning. I'm officer [name] with the [city] police department.
2	<i>Reason</i> : explaining the offense for which driver is being stopped	The reason I stopped you is because your headlight is out in the front.
3	<i>Documentation</i> : requesting driver's documents, such as identification or insurance	Do you have your driver's license on you?
4	<i>Details</i> : asking about personal or demographic information (name, address, age, etc.)	What's your last name? Is this your current address? How old are you?
5	<i>Orders</i> : imperative statements commanding the driver	Keep your hands on the wheel; turn the car off.
6	<i>Legitimacy</i> : questioning the driver's legitimacy typically pertaining to their presence (why are you here), actions or behavior (what are you doing), or ownership of property (is this your car?)	What were you guys doing? Whose car is this?

spoken by the officer to the driver and comprise about the first 27 s of the stop (for details on selecting officer word threshold, see *Materials and Methods*).

Two research assistants (blind to condition) hand-annotated and counted each institutional dialog act in the first 45 words spoken by the officer to the driver for each of the 577 stops of Black drivers. Annotation agreement on all dialog acts [based on weighted squared kappa (κ)] ranged from moderate to strong agreement ($0.6 \leq \kappa \leq 0.90$) with the exception of Documentation and Details, which had lower consistency ($k = 0.533$, $k = 0.555$, respectively) (24). All discrepancies between the two annotators were manually verified and resolved by a third annotator. For more details on annotation agreement, see *Materials and Methods*.

To understand how officers adhere to institutional dialog norms when talking to Black drivers, we draw on prior work that shows how initial dialog strategies in conversations could be used to predict future toxicity and other problems in conversations online (25). We built a logistic regression model to test which of these six dialog acts were associated with escalated vs. non-escalated stops. Our model controlled for the driver's and officer's gender (M, F), the officer's race (Black, White, Other), and neighborhood crime statistics based on the stop location. Based on the model results in Table 2, in their first 45 words—in roughly their first 27 s of speaking—officers in escalated stops of Black drivers are significantly more likely to give orders (Odds Ratio = 1.29, $B = 0.26$, 95% CI = [0.08, 0.44], $P = 0.005$) and less likely to provide reasons for the stop (OR = 0.48, $B = -0.74$, CI = [-1.36, -0.11], $P = 0.02$) compared to officers in non-escalated stops. Results were similar when control variables were removed (*SI Appendix, Table S7*).

Alongside these hand-annotated analyses, we tested whether natural language processing (NLP) models could accurately distinguish escalated from non-escalated stops based only on the officer's initial words. We split our data into a training set and test set and built a series of models (e.g., logistic regression, random forests, support vector machines, neural BERT classifiers) to predict whether a stop was escalated vs. non-escalated based solely on the same 45 words (*SI Appendix, Predicting Escalated Outcomes with Humans and NLP Models*). Our best-performing neural classifier (DeBERTa V3) could predict whether an unseen stop had an escalated outcome with 70.83% accuracy (95% CI: 68.90, 72.60) on an equally balanced test set. These results provide further evidence of structural differences in the earliest moments of escalated interactions.

The fact that the initial words an officer speaks during a car stop can presage an escalated outcome suggests that escalation need not build over the length of the stop: stops that ended in escalation often began in escalation. Could officers' language in these encounters simply reflect their reaction to drivers' combative language and actions? Perhaps. However, this account is not supported by the data. A content coding of drivers' language in our thin slices of escalated encounters found no instances where the driver refused to comply with an officer's command or answer an officer's question (see *SI Appendix, Table S10* for details). Instead, the overwhelming majority of utterances are direct responses to officers: responses to questions, verbal acknowledgments of compliance, or explanations proffered for their behavior.

Nor do our results appear to be driven by officers' prestop decisions to search or arrest drivers with outstanding warrants or with probable cause for more serious offenses. In fact, 94% of the interactions in our dataset were identified as stops for traffic violations (e.g., having a broken taillight, rolling through a stop sign) rather than probable cause by the conducting officer, and the linguistic differences we observed between escalated and non-escalated stops persisted after controlling for the officer's stated legal justification for the stop (*SI Appendix, Table S9* for details). We note, however, that stops often blur the line between investigation and traffic enforcement. It is impossible to ascertain the totality of an officer's knowledge and motivation in conducting any particular stop, and an officer may use a traffic violation as a legal pretext to detain a driver they wish to investigate (26, 27). Thus, neither the reason listed by the officer in stop records nor what they communicate to the driver necessarily characterize the full motivation for the stop. While we cannot know the full reason for any given stop, study 1 suggests that officers' language early in the stop is indicative of how the stop may end, even in stops where there is no evidence to suggest probable cause.

From the perspective of those stopped by the police, and particularly of Black drivers, this ambiguity further heightens the tension of encounters with law enforcement. Providing the justification for a stop earlier rather than later in a stop comports with ideals of transparency and trustworthiness (21, 28, 29) and could bring the tension down. Yet, compared to non-escalated stops, we find that officers in escalated stops were 2.5 times more likely not to provide a reason for the stop (38% vs. 15%) and were nearly three times more likely to initiate the stop with an order (22% vs. 8%). Do these linguistic differences—which seem to undermine police

Table 2. Full logistic regression model (N = 577; Outcome: Escalated stops)

Independent Variables		B	Odds Ratio	95% CI		S.E.	P
Institutional Dialog Acts	Greeting	-0.27	0.76	-0.70	0.15	0.22	0.211
	Reason	-0.74	0.48	-1.36	-0.11	0.32	0.020 *
	Documentation	0.15	1.16	-0.18	0.46	0.16	0.357
	Details	0.06	1.06	-0.62	0.65	0.32	0.858
	Order	0.26	1.29	0.08	0.44	0.09	0.005 **
	Legitimacy	-0.13	0.88	-0.64	0.32	0.24	0.592
Control Variables	Driver's Gender: Male	1.54	4.66	0.84	2.33	0.38	0.000 ***
	Officer's Race: Other	0.02	1.02	-0.81	0.92	0.44	0.968
	Officer's Race: White	0.34	1.40	-0.47	1.23	0.43	0.430
	Officer's Gender: Male	0.53	1.69	-0.42	1.66	0.52	0.314
	Violent Crime Rate	0.00	1.00	-0.01	0.00	0.00	0.001 **
	Property Crime Rate	0.00	1.00	0.00	0.00	0.00	0.539
	Narcotics Crime Rate	0.01	1.01	0.00	0.01	0.00	0.000 ***
(Intercept)	-2.27	0.10	-3.95	-0.72	0.82	0.006 **	

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

legitimacy—impact how drivers perceive the encounter in its initial moments? In study 2, we examine how differences in the conversation structure between escalated and non-escalated stops translate to citizen perceptions of these disparate encounters.

Study 2: How Black Men Experience Escalated Stops

How do structural differences in officer language between escalated and non-escalated encounters of Black drivers impact community members' perceptions of these interactions? Do Black drivers discern these linguistic differences early in the stop to make predictions about how the stop will end? Further, do these linguistic differences elicit more negative emotions and more worry about the possibility of force being used? In our sample of interactions, there were nearly ten times as many Black men who encountered one of the three escalated outcomes as there were Black women; accordingly, we focus this examination on how Black men in the United States experience these interactions.

In study 2, a nationally representative sample of Black men ($N = 188$) listened to a subset of 100 audio clips from the same car stops analyzed in study 1, evenly divided between escalated and non-escalated stops in a block replication design. Each participant listened to audio clips of five escalated and five non-escalated car stops and was asked to mentally place themselves in the seat of the driver to imagine how they would feel, how they would perceive the officer, and what outcomes they would expect. As in study 1, each clip was limited to the first 45 words of an officer's speech. Driver speech was masked in order to isolate the effect of officer speech, and identifying information (e.g., officer name) was removed from the clips.

We tested four preregistered hypotheses. First, we predicted that Black men would experience more negative emotions after listening to thin slices of escalated stops. Second, they would rate the officers' demeanor more negatively in escalated stops. Third, they would be more confident that escalated stops in fact ended in a handcuffing, search, or arrest. Finally, they would be more worried about the potential use of force during escalated stops. Code, data, and preregistration for this study are available at <https://osf.io/cdy5w/>.

To compare perceptions of escalated stops to non-escalated stops, we submitted each of our dependent variables to a linear mixed-effects model with random intercepts for audio clips and participants. Given the hierarchical structure of our data (e.g., trials within participants), this model allows us to account for variation by clip and across participants (30). Missing data were imputed through multivariate imputation by chained equations (MICE) using the mice package in R (31).

Driver Emotions. After listening to each clip, we asked participants to rate how they would feel if they were the driver in the stop on a series of dimensions: fearful, anxious, confused, shaken, calm (reverse scored), helpless, alert, on guard, harassed, and safe (reverse scored; 1 = not at all, 5 = extremely). We averaged ratings across each item to create a composite index of negative emotions ($\alpha = 0.90$). As predicted, participants reported they would feel more negative emotions if they were the driver in an escalated stop than a non-escalated stop, $\beta = 0.27$ [0.11, 0.43], $t(93) = 3.28$, $P = 0.001$.

Officer Demeanor. In addition, we asked participants to rate to what extent they found the officer: aggressive, commanding, forceful, threatening, courteous (reverse scored), rigid, friendly (reverse scored), respectful (reverse scored), and condescending (1 = not at all, 5 = extremely). We again averaged ratings across each item to create a composite ($\alpha = 0.89$). As hypothesized, based on the first 45 words of the officer alone, participants rated officer demeanor more negatively in escalated vs. non-escalated stops, $\beta = 0.30$ [0.12, 0.49], $t(96) = 3.20$, $P = 0.002$.

Predicted Outcomes. We next asked participants whether they believed a handcuffing, search, or arrest would occur in each stop as a yes/no binary item and to rate their confidence in such an outcome occurring in that stop (1 = a little confident, 3 = very confident). We assigned numeric values to each binary item (1 = yes, -1 = no) and multiplied these by the confidence rating to create a six-point scale for each outcome ranging from "very confident [outcome] will not happen" to "very confident [outcome] will happen". As predicted, Black men were more confident that escalated stops would result in handcuffing ($\beta = 0.30$ [0.12, 0.47], $t(98) = 3.35$, $P = 0.001$), search ($\beta = 0.31$ [0.14, 0.48], $t(99) = 3.56$, $P = 0.001$), and arrest ($\beta = 0.24$ [0.07,

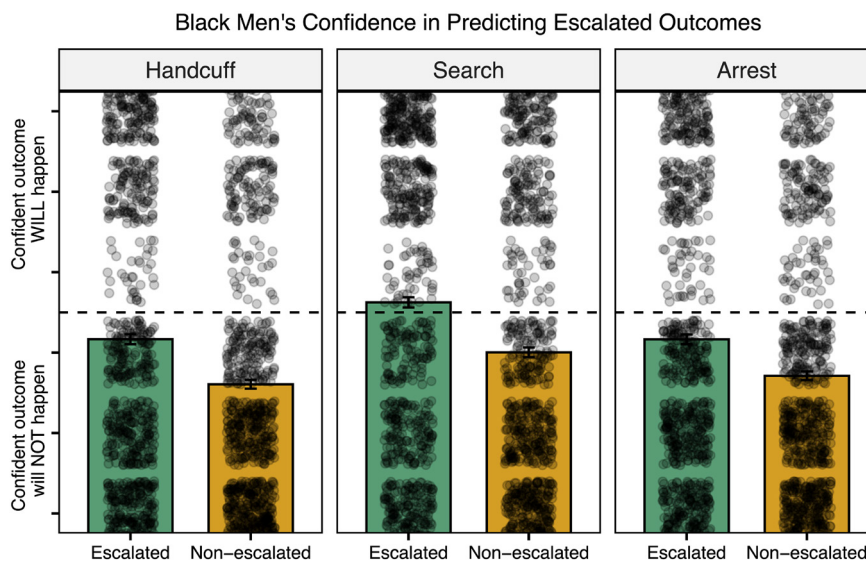


Fig. 1. Prediction outcome scores for handcuffing, search, and arrest by stop type (escalated vs. non-escalated). Prediction scores were transformed to a 1 to 6 scale representing confidence that a given outcome would occur in the stop, such that a score of 1 = "very confident [outcome] will not happen", and 6 = "very confident [outcome] will happen". Bars represent means, and dots represent individual data points across all stops ($N = 1,880$). Dotted line differentiates between yes/no outcome predictions; data points below the dotted line indicate responses of "no", those above the dotted line indicate "yes" response.

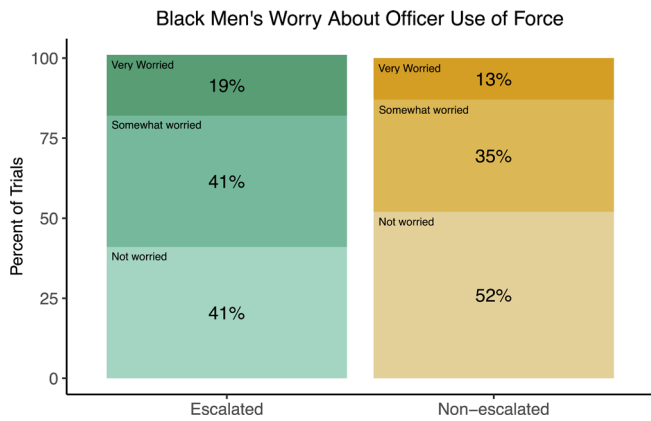


Fig. 2. Stacked bar plot of worry about use of force by stop type (escalated vs. non-escalated). Participants rated how worried they were about force being used in the stop on a 1 to 3 scale, 1 = Not at all worried, 3 = Very worried. Bar labels represent percentages across escalated vs. non-escalated trials (N = 1,880).

0.41], $t(107) = 2.80, P = 0.006$, relative to non-escalated stops (Fig. 1). Indeed, respondents were much more likely to state they were very confident (i.e., to provide a rating of 6 on the 6-point confidence scale) that an escalated interaction would result in a handcuffing ($X^2(1, N = 1,880) = 38.17, \phi = 0.14, P < 0.001$), search ($X^2(1, N = 1,880) = 28.03, \phi = 0.12, P < 0.001$), and arrest ($X^2(1, N = 1,880) = 26.52, \phi = 0.12, P < 0.001$) relative to non-escalated stops.

Worry about Use of Force. Although none of the stops presented in the study involved the use of force, we asked participants how worried they would be about each interaction ending with the officer using force or threatening to use force (1 = not at all worried, 3 = very worried). As predicted, participants felt more worried that the stop could involve force by the officer in escalated stops than in non-escalated stops, $\beta = 0.25 [0.11, 0.38], t(96) = 3.53, P = 0.001$ (Fig. 2). Further, exploratory multiple mediation analysis reveals that the relationship between the escalation status of the stop and worry about force is simultaneously mediated by participants' emotions ($\beta = 0.34, SE = 0.06, P < 0.001$), perceptions of officer demeanor ($\beta = 0.30, SE = 0.06, P = 0.001$), and the anticipated outcomes of the stop ($\beta = 0.14, SE = 0.04, P = 0.01$). For more details on the analysis, see *SI Appendix, Study 2 Exploratory Analyses*.

Role of Dialog Acts. Clearly, Black men are distinguishing between stops that escalate and those that do not. Yet, what information are participants using to inform these disparate perceptions? Are

they picking up on the same linguistic signature of escalation that we saw in study 1: the presence of orders and absence of reason? Exploratory analyses suggest they are (Table 3).

An analysis of variance (ANOVA) reveals a significant effect of reason and order dialog acts on negative emotion ratings ($F(3, 1874) = 61.67, P < 0.001, \eta^2 = 0.09$). A post hoc Tukey's HSD test reveals that average negative emotion ratings are significantly higher for stops that include orders and no reason ($M = 3.65, SD = 0.99$) compared to stops that include a reason and do not include orders ($M = 2.63, SD = 0.85, P < 0.001, 95\% CI = [0.80, 1.23]$). The same pattern exists between reason and order dialog acts and officer demeanor ratings ($F(3, 1876) = 86.07, P < 0.001, \eta^2 = 0.12$). Officer demeanor is perceived significantly more negatively in stops with orders and no reason ($M = 3.38, SD = 0.96$) relative to stops with reason and no order ($M = 2.31, SD = 0.81, P < 0.001, 95\% CI = [0.86, 1.27]$). In addition, we find a significant difference between dialog acts present in a stop and predictions of whether a stop ended in handcuffing ($X^2(3, N = 1,880) = 168.61, P < 0.001, \text{Cramer's } V = 0.30$), search ($X^2(3, N = 1,880) = 158.48, P < 0.001, \text{Cramer's } V = 0.29$), and arrest ($X^2(3, N = 1,880) = 121.61, P < 0.001, \text{Cramer's } V = 0.25$), such that stops that include orders and no reason are more likely to elicit predictions of escalated outcomes relative to stops with reason and no orders. Finally, reason and order dialog acts are associated with worry about the use of force ($X^2(6, N = 1,880) = 101.44, P < 0.001, \text{Cramer's } V = 0.16$), such that participants are more likely to be "very" or "somewhat" worried about force in stops where orders are present, but reason is absent.

General Discussion

What can the first moments of police stops tell us about the trajectory of those encounters for Black drivers? Here, we show that there is a linguistic signature to escalated car stops: officers are less likely to provide reasons and more likely to give orders in their first 45 words to the driver (which, on average, occur in the first 27 seconds of the stop). This linguistic signature could be discerned by trained coders, NLP models, and, most importantly, by Black citizens themselves. Not only do Black men experience more negative emotion and rate officer demeanor more negatively in escalated stops, they are also more likely to expect escalation and to worry that these interactions could lead to police threats of force or actual use of force. In fact, we show that Black men are specifically attuned to the officers' words: the presence or absence of orders and a reason for the stop, drive their perceptions of that stop.

Table 3. Participant perceptions by the presence/absence of reason and order dialog acts

Participant Perceptions	Reason absent/ Order present (N = 125 trials)	Reason present/ Order present (N = 241 trials)	Reason absent/ Order absent (N = 245 trials)	Reason present/ Order absent (N = 1,269 trials)
<i>Participant Emotion composite</i> (1[least negative] to 5 [most negative] scale)	3.65	3.12	2.81	2.63
<i>Officer Demeanor composite</i> (1[least negative] to 5 [most negative] scale)	3.38	2.94	2.40	2.31
Any escalated outcome prediction = yes	84.80%	62.24%	55.92%	36.57%
Handcuff prediction = yes	69.60%	44.40%	38.37%	21.59%
Search prediction = yes	81.60%	57.26%	52.24%	32.78%
Arrest prediction = yes	65.60%	41.49%	39.18%	23.64%
Worry about force = somewhat or very worried	80.80%	65.90%	56.30%	47.40%

Bold font indicates primary stop types of interest, while remaining stop types are included in regular font.

This is not to say that the outcome of a stop is preordained by an officer's choice of words. Indeed, while escalated and non-escalated encounters can be discerned above chance from the thin slices we employ, there is linguistic overlap between these two classes of encounters as well. The fuzzy boundary between "non-escalated" and "escalated" encounters may be one reason why police stops are so anxiety-inducing among citizens of color (32). For that matter, linguistically, stops that include handcuffing,

search, or arrest likely share much in common with encounters that escalate to police violence.

Millions of people are now familiar with how officers treated George Floyd after he was forcibly removed from his vehicle on May 25, 2020. Far fewer are familiar with what happened *before* he was removed—during the first moments of police contact. When considering Floyd's stop in terms of institutional dialog acts, we find that the first 45 words of the officer who made initial

No	Start	End	Speaker	Utterance	Speech Act
1	20:09:21	20:09:28	LANE	[OFFICER LANE walks across the street towards FLOYD's car].	
2	20:09:28	20:09:29	LANE	[OFFICER LANE taps the driver's seat window three times with the end of his flashlight].	
3	20:09:30	20:09:32	FLOYD	[FLOYD turns around immediately] <i>Oh, oh, oh!</i>	Exclamation
4	20:09:32	20:09:33	LANE	[OFFICER LANE raises the end of his flashlight towards FLOYD'S window as FLOYD is turning around]. <i>Let me see your hands.</i>	Order: hands
5	20:09:33	20:09:34	LANE	[OFFICER LANE taps FLOYD'S car window twice].	
6	20:09:34	20:09:35	FLOYD	[FLOYD opens the car door] <i>Hey, man. I'm sorry! I'm sorry -</i>	Apology
7	20:09:36	20:09:37	LANE	[overlapping] <i>Stay in the car, let me see your other hand.</i>	Order: stay in car, hands
8	20:09:37	20:09:38	FLOYD	<i>I'm sorry, I'm sorry! I'm sorry! I'm sorry -</i>	Apology
9	20:09:38	20:09:39	LANE	[overlapping] <i>Let me see your other hand!</i>	Order: hands
10	20:09:39	20:09:40	FLOYD	[overlapping] <i>Please, please, Mr. Officer.</i>	Plea
11	20:09:40	20:09:41	LANE	<i>Both hands!</i>	Order: hands
12	20:09:41	20:09:41	FLOYD	[overlapping] <i>I didn't do nothing.</i>	Proclaim Innocence
13	20:09:41	20:09:42	LANE	[OFFICER LANE points a gun at FLOYD'S face and uses his left hand to open the car door wider]. <i>Put your fucking hands up right now!</i>	Order: hands
14	20:09:42	20:09:43	FLOYD	[FLOYD with hands raised] <i>Please! Please! I do nothing! I do nothing!</i>	Plea; Proclaim Innocence
15	20:09:44	20:09:45	LANE	<i>Let me see 'em all your hands.</i>	Order: hands
16	20:09:45	20:09:46	HILL	<i>Let him see your other hand.</i>	
17	20:09:46	20:09:47	FLOYD	[FLOYD turns to the passenger's side while keeping his hands up] <i>All right. What I do though? What'd we do though?</i>	Request for Reason
18	20:09:47	20:09:48	FLOYD	[FLOYD turns to face Officer Lane while keeping his hands up] <i>What we do Mr. Officer? I didn't do nothing.</i>	Request for Reason; Proclaim Innocence
19	20:09:48	20:09:49	LANE	[OFFICER LANE points to the steering wheel with his left hand] <i>Put your hand up there.</i>	Order: hands
20	20:09:49	20:09:50	FLOYD	<i>God man -</i>	Plea
21	20:09:50	20:09:51	LANE	[overlapping] <i>Put your fucking hand up there!</i> [OFFICER LANE pushes FLOYD'S shoulder with his left hand while still pointing his gun at FLOYD'S face with his right hand]	Order: hands
22	20:09:51	20:09:52	FLOYD	<i>I got shot before. I got shot.</i>	Explanation
23	20:09:52	20:09:55	LANE	[overlapping] <i>Jesus Christ, keep your fucking hands on the wheel!</i>	Order: hands
24	20:09:55	20:09:55	FLOYD	<i>I got shot.</i>	Explanation

Fig. 3. Timestamped transcript of police body-camera footage capturing the first 27 s (no. 2 to 24) of the interaction between the primary officer (Thomas Lane) and George Floyd. The officer's first 45 words occur within the first 23 s of interaction (no. 2 to 21). Body camera footage from the primary officer's camera was transcribed by the authors (33) and verified by a professional transcription agency. Language is denoted in italics while additional context is provided in closed brackets.

contact with Floyd occurred in the first 23 s of the interaction (Fig. 3, no. 2 to 21); in the first 27 s of the interaction, the officer spoke a total of 57 words to Floyd (Fig. 3, no. 2 to 24).

In those 27 s, there are a total of nine officer speech turns, the entirety of which consist only of physical orders. These commands are responses to Floyd's speech turns that span a much wider range of speech acts. For example, in the same 27 s, Floyd communicates with the officer through a total of 11 turns, which contain two apologies ("Hey, man. I'm sorry! I'm sorry -"), two requests for reason for the stop ("What we do Mr. Officer?"), three proclaiming his innocence ("I do nothing! I do nothing!"), two explanations signaling fear ("I got shot before. I got shot."), one exclamation ("Oh, oh, oh!"), and three pleas ("Please, please, Mr. Officer."). As shown in the transcript excerpt, all of Floyd's apologies, requests, proclamations, explanations, exclamations, and pleas are consistently responded to with a singular response: an order.

There are many studies of George Floyd's last encounter with police, including discourse analyses of news reports (34, 35), official press conferences (36), and responses by academic organizations (37, 38). There are even studies of how Black and White parents talked to their children about race before and after George Floyd's death (39). Yet, what remains surprisingly understudied is the language of the police officers and Floyd himself during the initial moments of their encounter. What can these early moments tell us about how use-of-force cases may transpire?

The current work reveals how officers' language can act as a signal of the trajectory of a stop that can be discerned by Black drivers: *I could be handcuffed. I could be searched. I could be arrested on the side of the road.* Although only 1 to 2% of police contacts per year involve the use of force (40), our data suggest the percentage of contacts where a driver is worried about force being used is much higher. It is not only the presence of force but also the potential for force that drives negative police encounters. Given well-known racial disparities in policing, Black male drivers may be more likely to be concerned about that possibility than any other group. An officer's speech, then, can set into motion negative perceptions and emotions of the driver—sparkling a dynamic that erodes trust and undermines the relationship between police and those they are meant to serve.

The case of George Floyd is just one example of what footage can tell us about police–community interactions more broadly. Body-worn cameras document police encounters as they unfold, beat by beat, word by word. In contrast to administrative records, which describe stops as a single event in time from the perspective of the officer, camera footage lets us chart the trajectory of police encounters, from their first moments to their final outcomes. In tandem, it lets us observe the impact of these encounters on community members, giving a better understanding of both policing and being policed. Yet, despite holding valuable and dynamic insight into police–community relations, footage analysis remains a largely underutilized resource. Police departments cite increasing community trust as a primary reason for adopting body-worn cameras (41). However, the vast majority of BWC footage goes unwatched.

Systematically analyzing footage can not only allow us to diagnose the health of police–community relations but could also assist us in improving those relations. Institutional dialog acts such as those analyzed here are a vehicle by which individual officers enact institutional power. Officers are trained on appropriate dialog acts in car stops—for example, to give a reason for the stop and to ask for documents. However, they are also given broad discretion to communicate as they see fit; indeed, we can detect significant differences in language across escalated and non-escalated stops. These differences impact Black drivers' emotional experience, impression of the officer, and predictions about

the stop's ultimate outcomes. What if systematic, rigorous footage review—even of routine encounters—became an institutionalized process within police departments as well as outside of them? The development of such a review process across a wide variety of encounters—from car stops to calls for service—could lead to changes to agency practices that promote the communication of trust and transparency (21), rather than invoking fear and anxiety. While language is just one element in the matrix of forces that contribute to police–community relations, our findings highlight its importance in addressing the current crisis of police legitimacy.

Materials and Methods

Study 1:

Choice of 45 word threshold for officer speech. We chose a low threshold of 45 words (approximately 27 s of speech) to keep only the officer's initial speech turns and simultaneously to minimize the occurrences of words directly associated with escalated outcomes, such as "arrest," "search," or "handcuff" (the longer the initial segment of officer speech, the more likely one of these words would occur). Because both study 1 and study 2 investigate the association between the officer's initial speech and the acts constitutive of escalation (arrests, searches, or handcuffing), we eliminated nine stops in which the initial speech contained one of these words. The choice of 45 words kept the number of stops to be eliminated at less than 10 stops. We found that longer thresholds unacceptably increased the number of stops that would have to be deleted from our dataset (*SI Appendix, Fig. S1*).

Inclusion Criteria for Institutional Dialog Acts. Of the 11 institutional dialog acts identified in prior literature (14), we focus on six: Greeting, Reason, Documentation, Details, Order, and Legitimacy as shown in the main paper. We excluded Sanction, Offering Help, Positive Closing, History, and Search Inquiry dialog acts (*SI Appendix, Table S3* for description of each act) for several reasons. First, we wanted to avoid acts that are confounded by being direct predictors of our output variables. We therefore excluded History and Search Inquiry dialog acts from our main analysis because they naturally contain words directly implying the escalated outcomes. We also removed dialog acts that did not occur with sufficient frequency in the officers' initial words. No stops included Positive Closing or Offering Help in the first 45 words, and only two stops included Sanction acts during our annotation process. Hence, these three dialog acts were also omitted from our analysis.

Annotation agreement. Two annotators rated 45-word segments of officer speech from a sample of 577 vehicle stops for the count of each of the 11 institutional dialog acts. We then measured inter-annotator consistency using weighted kappa for each dialog act (*SI Appendix, Table S4*). Of the six dialog acts we focus on in our analyses, four of them (Greeting, Reason, Order, Legitimacy) had moderate to strong agreement ($0.6 \leq \kappa \leq 0.90$), while Documentation and Details had low consistency (24).

Study 2:

Participants. We recruited 200 Black male US citizens to participate in the study via an online recruitment platform. Our sample was nationally representative in terms of age, region, education, and political ideology (*SI Appendix, Table S13* for sample demographics). All research was approved by the University of Michigan Institutional Review Board, informed consent was obtained from all participants before their participation, and participants were compensated for their time.

Consistent with our preregistration, we excluded five participants who failed an attention check (in which they were asked to leave a block of items blank), five participants who did not identify as male, and two participants who did not identify as Black. Thus, our final sample consisted of 188 Black male US citizens.

We note that, in our nationally representative sample of Black male US citizens, 41% have previously been handcuffed, searched, and arrested; 65% know people who have been handcuffed, searched, and arrested; and 26.1% have previously had force used on them by an officer (*SI Appendix, Table S13*). Thus, not only do Black men generally experience escalated outcomes at higher rates than other segments of the US population, Black male participants in study 2 do so as well.

Study Procedure. Participants were told that they would be listening to anonymized audio clips of real car stops and were asked to place themselves in the position of the driver as they listened to each recording. Participants then responded to 10 audio clips: five from escalated stops and five from non-escalated stops, presented in random order. Stimuli were allocated across participants by, first, randomly sorting the 100 clips in our stimulus set into batches of ten stimuli and then randomly assigning participants one of the 10 batches. For each clip, participants indicated a) the extent to which they would feel a number of emotions if they were the driver in the stop (e.g., fearful, anxious; 1 = not at all, 5 = extremely); b) the demeanor of the officer (e.g., commanding, respectful; 1 = not at all, 5 = extremely); c) whether they expected the stop to end with a search, handcuffing, and/or arrest (yes/no), along with their confidence in their predictions (1 = a little confident, 3 = very confident); and d) how worried they would be about the officer in the clip using force or threatening to use force (1 = not worried at all, 3 = very worried).

Data, Materials, and Software Availability. Anonymized data [study 1: codes for the logistic regressions (R script) and NLP models (python script), as well as the data (count of dialog acts in each 45-word segment of officer speech per stop) as a CSV file; study 2: codes for the statistical models (R script), anonymized survey

response data, and participant demographic data (CSV files)] have been deposited in Github (study 1) (42) and OSF (study 2) (43).

ACKNOWLEDGMENTS. This research was supported by grants from the John D. and Catherine T. MacArthur Foundation (G-1512-150464 and G-1805-153038). We thank the city and the police department whose provision of officer body-worn camera footage enabled this research. The funding, collaborations, and research referenced in this publication were administered and supported by Stanford SPARQ, a center that builds research-driven partnerships with industry leaders and changemakers to combat bias, reduce disparities, and drive culture change. We thank the researchers at Stanford SPARQ for their feedback, Kristyn Jones, Amrita Maitreyi, Lee Hickman, and Kaitlene Poblete for their annotations, and Xiaohan Ding and Chenguang Zhu for their support and input on the language analyses.

Author affiliations: ^aDepartment of Computer Science, Virginia Tech, Blacksburg, VA 24061; ^bDepartment of Psychology, Stanford University, Stanford, CA 94305; ^cDepartment of Organizational Studies, University of Michigan, Ann Arbor, MI 48109; ^dDepartment of Computer Science, Stanford University, Stanford, CA 94305; ^eDepartment of Linguistics, Stanford University, Stanford, CA 94305; and ^fGraduate School of Business, Stanford University, Stanford, CA 94305

- E. Shuman, S. Hasan-Asli, M. van Zomeren, T. Saguy, E. Halperin, Protest movements involving limited violence can sometimes be effective: Evidence from the 2020 BlackLivesMatter protests. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2118990119 (2022).
- J. Stern, They used smartphone cameras to record police brutality—and change history. *Wall Street Journal*, 13 June 2020. <https://www.wsj.com/articles/they-used-smartphone-cameras-to-record-police-brutality-and-change-history-11592020827>. Accessed 25 July 2022.
- D. Sharp, US police departments clamoring for de-escalation training. Associated Press, 24 September 2021. <https://apnews.com/article/death-of-george-floyd-police-new-york-new-york-city-maine-a85463f8955fc4be53e7b66287c1a87>. Accessed 25 July 2022.
- E. Harrell, E. Davis, "Contacts between police and the public, 2018 - statistical tables." Bureau of Justice Statistics Report, NCJ, 255730. (2020) Accessed 25 July 2022.
- National Institute of Justice, "Racial profiling and traffic stops" (National Institute of Justice, 2013).
- I. Ayres, J. Borowsky "Study of racially disparate outcomes in the los angeles police department | office of justice programs" (Office of Justice Programs, 2008) (July 25, 2022).
- W. R. Smith et al., The North Carolina highway traffic study: Final report. *National Institute of Justice*, **407**, (2004).
- F. R. Baumgartner, D. A. Epp, K. Shoub, *Suspect Citizens: What 20 Million Traffic Stops Tell Us About Policing and Race* (Cambridge University Press, 2018).
- J. Jones, In U.S., Black Confidence in Police Recovers From 2020 Low. Gallup, 14 July 2021. <https://news.gallup.com/poll/352304/black-confidence-police-recovers-2020-low.aspx>. Accessed 19 July 2022.
- D. Desilver, M. Lipka, D. Fahmy, "Race and policing in America: 10 things we know about race and policing in the U.S." (2020) (June 5, 2021).
- V. Prabhakaran et al., Detecting institutional dialog acts in police traffic stops. *TAFL* **6**, 467-481 (2018).
- B. J. Grosz, *The Representation and Use of Focus in Dialogue Understanding* (University of California, Berkeley, 1977).
- R. Voigt et al., Language from police body camera footage shows racial disparities in officer respect. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6521-6526 (2017).
- H. Giles et al., "Accommodation and Institutional Talk: Communicative dimensions of police-civilian interactions" in *Language, Discourse and Social Psychology, Palgrave Advances in Linguistics*, W. Ann, M. W. Bernadette, G. Cindy, Eds. (Palgrave Macmillan UK, 2007), pp. 131-159.
- H. Giles et al., "Communication accommodation: Law enforcement and the public" in *Applied research in interpersonal communication: Family communication, health communication and communicating across social boundaries*, R.M. Dailey, B.A. Le Poire, Eds. (Peter Lang, Vol. 5, 2006, pp. 241-269).
- T. R. Tyler, P. A. Goff, R. J. MacCoun, The impact of psychological science on policing in the United States: Procedural justice, legitimacy, and effective law enforcement. *Psychol. Sci. Public Interest* **16**, 75-109 (2015).
- T. R. Tyler, Procedural justice, legitimacy, and the effective rule of law. *Crime Justice* **30**, 283-357 (2003).
- T. Meares, Policing and procedural justice: Shaping citizens' identities to increase democratic participation. *Nw. U. L. Rev.* **111**, 1525 (2016).
- J. Sunshine, T. R. Tyler, The role of procedural justice and legitimacy in shaping public support for policing. *Law Soc. Rev.* **37**, 513-548 (2003).
- T. R. Tyler, What is procedural justice - criteria used by citizens to assess the fairness of legal procedures. *Law Soc. Rev.* **22**, 103 (1988).
- M. Quattlebaum, T. L. Meares, T. Tyler, Principles of Procedurally Just Policing (2018). <https://doi.org/10.2139/ssrn.3179519> (September 20, 2022).
- J. Heritage, "Conversation analysis and institutional talk" in *Handbook of Language and Social Interaction* (Psychology Press, Vol. **103**, 2005, p. 47).
- Office of Community Oriented Policing Services, *The President's Task Force on 21st Century Policing Implementation Guide: Moving from Recommendations to Action* (Office of Community Oriented Policing Services, Washington, DC, 2015).
- M. L. McHugh, Interrater reliability: The kappa statistic. *Biochem. Med.* **22**, 276-282 (2012).
- J. Zhang, "Conversations gone Awry: Detecting early signs of conversational failure" in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, 2018), Vol. **1**, Long Papers, pp. 1350-1361.
- C. R. Epp, S. Maynard-Moody, D. P. Haider-Markel, *Pulled Over: How Police Stops Define Race and Citizenship* (University of Chicago Press, 2014).
- Whren v. United States, 517 U. S. 806, 1996. Accessed 29 August 2022.
- Office of Community Oriented Policing Services, Final report of the president's task force on 21st century policing (Office of Community Oriented Policing Services, Washington, DC, 2015).
- T. R. Tyler, Y. Huo, *Trust in the law: Encouraging public cooperation with the police and courts* (Russell Sage Foundation, 2002).
- A. Galecki, T. Burzykowski, "Linear mixed-effects model" in *Linear Mixed-Effects Models Using R: A Step-by-Step Approach*, A. Galecki, T. Burzykowski, Eds. (Springer Texts in Statistics, Springer, 2013), pp. 245-273.
- S. van Buuren, K. Groothuis-Oudshoorn, Mice: Multivariate imputation by chained equations in R. *J. Stat. Software* **45**, 1-67 (2011).
- J. T. Pickett, A. Graham, F. T. Cullen, The American racial divide in fear of the police. *Criminology* **60**, 291-320 (2022).
- 10 Tampa Bay, RAW: George Floyd Minneapolis police body camera footage. *YouTube* (2020) (25 July 2022).
- C. Lin, Who is to Blame?: A discourse analytic comparison of the news reports on the killing of George Floyd in Mainland China and Taiwan. *Chinese Language and Discourse*, **13**, 197-225 (2021). <https://doi.org/10.1075/cld.20030.lin>.
- R. N. Canevez, M. Karabelnik, J. S. Winter, Police brutality and racial justice narratives through multi-narrative framing: Reporting and commenting on the george floyd murder on youtube. *J. Mass Commun. Quart.* **99**, 696-717 (2022). [10.1177/1076990221108722](https://doi.org/10.1177/1076990221108722).
- A. Oursler, A. DalCortivo, Grieving for George Floyd. *Nation* **312**, 46-46 (2021).
- M. Stack, The murder of George Floyd and the mediatization of solidarity by top-ranked universities. *Lang. Intercult. Commun.* **21**, 749-764 (2021).
- A. Brown, E. Auguste, F. Omobhude, N. Bakana, J. Sukhera, Symbolic solidarity or virtue signaling? A critical discourse analysis of the public statements released by academic medical organizations in the wake of the killing of George Floyd. *Acad. Med.* **97**, 867-875 (2022).
- J. N. Sullivan, J. L. Eberhardt, S. O. Roberts, Conversations about race in Black and White US families: Before and after George Floyd's death. *PNAS Proc. Natl. Acad. Sci. U.S.A.* **118**, e2106366118 (2021).
- M. R. Smith et al., *A multi-method evaluation of police use of force outcomes: Final report to the national institute of justice* (U.S. Department of Justice, Washington, DC, 2010).
- Police Executive Research Forum, "Cost and benefits of body-worn camera deployments" (Police Executive Research Forum, 2018) (August 30, 2022).
- E. H. Rho, R. Pryzant, Data from "Escalated police stops of Black men are linguistically and psychologically distinct in their earliest moments." GitHub. Available at <https://github.com/eugeniarho/police-escalation-project>. Deposited 3 October 2022.
- E. H. Rho et al., Data from "Escalated police stops of Black men are linguistically and psychologically distinct in their earliest moments." Open Science Framework. Available at <https://osf.io/cdy5w/>. Deposited 29 August 2022.