INNOVATIONS IN SOCIAL PSYCHOLOGY

The Thin Blue Waveform: Racial Disparities in Officer Prosody Undermine Institutional Trust in the Police

Nicholas P. Camp[1], Rob Voigt[2], Dan Jurafsky[3], and Jennifer L. Eberhardt[4]
[1] Department of Organizational Studies, University of Michigan
[2] Department of Linguistics, Northwestern University
[3] Department of Linguistics and Department of Computer Science, Stanford University
[4] Department of Psychology, Stanford University

How do routine police encounters build or undermine community trust, and how might they contribute to racial gaps in citizen perceptions of the police? Procedural justice theory posits that officers' interpersonal communication toward the public plays a formative role, but experimental tests of this hypothesis have been constrained by the difficulties of measuring and manipulating this dimension of officer-citizen interactions. Officer-worn body camera recordings provide a novel means to overcome both of these challenges. Across five studies with laboratory and community samples, we use footage from traffic stops to examine how officers communicate to drivers and whether racial disparities in officers' communication erode institutional trust in the police. Specifically, we consider the cumulative effects of one subtle interpersonal cue: an officer's tone of voice. In Studies 1A, 1B, and 1C, participants rated thin slices of officer speech. Participants were blind to the content of the officer's words and the race of their interlocutor, yet they evaluated officers' tone toward White (vs. Black) men more positively. By manipulating participants' exposure to repeated interactions, we demonstrate that even these paraverbal aspects of police interactions shape how citizens construe the police generally (Study 2), and that racial disparities in prosodic cues undermine trust in institutions such as police departments (Study 3). Participants' trust in the police, and personal experiences of fairness, in turn, correlated with their perceptions of officer prosody across studies. Taken together, these data illustrate a cycle through which interpersonal aspects of police encounters erode institutional trust across race.

Keywords: policing, race, prosody, procedural justice, social cognition

Supplemental materials: https://doi.org/10.1037/pspa0000270.supp

Racial gaps in police-community trust are pervasive and persistent in the United States: 42% of Whites but only 14% of African Americans report having a great deal of confidence in their local police department (Pew Research Center, 2016). These differences in institutional trust mirror disparities in Black and White Americans' reported experiences with the police. African Americans are more likely than Whites to report having been subjected to intrusive questioning (Epp et al., 2014) and disrespectful treatment (Tyler, 2005; Tyler & Huo, 2002) in their interactions with law enforcement.

These routine encounters are consequential because they are at once interpersonal and institutional interactions. As representatives of the state, police officers literally give voice to the law. In encounters with identical legal outcomes, an officer can communicate deference and understanding toward a citizen, or address them with condescension and indifference. At the same time, interpersonal aspects of these exchanges have institutional consequences. Citizens' experiences of respectful treatment at the hands of the police affect their perceptions of legal legitimacy, support for law enforcement, and even whether they themselves follow the law (Sunshine & Tyler, 2003b; Tyler, 1990; Tyler & Lind, 1992).

The current work sheds light on a cycle through which such institutional interactions erode police-community trust across race. First, officers communicate different levels of respect, warmth, and ease toward Black and White citizens. Second, these interpersonal cues accumulate across interactions to shape citizens' perceptions of and trust in law enforcement. Third, citizens interpret officer communication in light of this trust in their subsequent encounters.

Nicholas P. Camp https://orcid.org/0000-0001-9704-6133
Rob Voigt https://orcid.org/0000-0003-3743-0322
Dan Jurafsky https://orcid.org/0000-0002-6459-7745
Jennifer L. Eberhardt https://orcid.org/0000-0003-1687-8128

Correspondence concerning this article should be addressed to Nicholas P. Camp, Department of Organizational Studies, University of Michigan, 500 Church Street, Ann Arbor, MI 48109–1042, United States. Email: npcamp@umich.edu

Such an account is consistent with theories of procedural justice, which stress the relational nature of authority (Tyler & Blader, 2003). It comports with well-documented racial disparities in who the police stop (Gelman et al., 2007; Pierson et al., 2020) and search (Goel et al., 2016; Knowles et al., 2001). However, we lack the key source of data needed to test these premises: how police officers actually communicate with the public. The interpersonal dimension of police encounters is all but invisible in administrative records. Stop data reports can reveal racial disparities in officers' decisions to search or sanction citizens, but they cannot reveal whether officers address community members with respect or contempt. Interactions that are indistinguishable in administrative data may unfold quite differently in the experiences of community members, and have divergent consequences for their trust in law enforcement.

Community surveys can capture citizens' subjective experiences of their own police encounters, and relate these perceptions to their trust in law enforcement, but carry their own limitations. For one, it is impossible to dissociate citizens' perceptions of interpersonal treatment from officers' demeanor in those encounters. Variation in citizen self-reports could reflect differences in how police officers communicate with citizens, differences in how community members interpret similar conversations with police officers, or both. Further, since these data are correlational in nature, we are limited in the causal inferences we can make between officers' interpersonal treatment and citizens' perceptions of law enforcement.

We address these shortcomings with a novel source of data on police-citizen interactions: officer body-worn camera (BWC) footage. BWCs capture a rapidly increasing number of encounters; by some estimates, about half of all U.S. law enforcement agencies (Hyland, 2018), and 97% of major city police departments (Major Cities Chiefs Major County Sheriffs, 2015) have adopted or plan to adopt these devices. The bulk of scientific research concerns either the ambiguities of assigning blame from BWC video recordings (Jones et al., 2017, 2019; Turner et al., 2019), or whether the presence of BWCs themselves affects officer behavior (Ariel et al., 2015; cf. Yokum et al., 2019). Little research has considered the routine interactions that compose the bulk of BWC footage.

Body-worn cameras grant access to the interpersonal dimensions of these encounters for the first time. By capturing conversations between officer and citizen, they can reveal how these exchanges differ across race. For example, Voigt and colleagues (2017) harnessed BWC footage to examine the respectfulness of officers' language usage in routine traffic stops. Their analysis of transcribed footage revealed that police officers used more respectful language in stops of White versus Black drivers.

This footage can inform our understanding of the consequences of police interactions as well as their content, however. In contrast to studies that use a small number of lab-created stimuli (e.g., Maguire et al., 2017), a large number of BWC recordings can be sampled from distributions of actual police encounters, maintaining a high level of ecological validity while capturing the heterogeneity in these interactions. As a result, we can decouple the variability among police encounters from perceiver-level differences in how these stimuli are interpreted. Furthermore, by selectively presenting individuals with different recordings, we can estimate their causal influence on perceivers' institutional trust in the police.

In short, body cameras make the relational aspects of policing visible. This lets us test mechanisms through which police interactions translate to institutional mistrust or trust, alongside the racial dynamics of such encounters. Here, we consider one subtle but socially important channel of communication that can only be accessed from BWC recordings: prosody, or the acoustic features of one's voice.

Prosody and Social Hierarchy

Qualities such as the pitch, rhythm, and intonation of voice convey a wealth of information about the speaker and their relationship to their interlocutor (Belin et al., 2011). Prosody is essential to parse the social meaning of verbal communication, such as the difference between a polite utterance and a sarcastic one (Culpeper, 2011; DePaulo & Friedman, 1998), social evaluations like friendliness or authoritativeness (Jeong, 2016; Liscombe et al., 2003), or the presentation of particular social personas (Jeong, 2017; Podesva, 2011). From even brief exposure to vocal tone, observers infer a speaker's personality (McAleer et al., 2014), their race, gender, and physical size (Belin et al., 2011) and their emotional state (Belin et al., 2008; Scherer et al., 2003).

Prosody is a particularly strong indicator of hierarchical relationships between speakers (Cheng et al., 2016; Ko et al., 2015; Kraus et al., 2017), more so than other nonverbal cues like posture or eye contact (Hall et al., 2005). Research from Ko and colleagues (2015) illustrates how prosody both reflects and reifies social hierarchies. Participants randomly assigned to a dominant negotiating position spontaneously adopted a lower-pitched, more monotone vocal tone relative to those placed in a weaker position; a subsequent sample could accurately gauge the original participant's negotiating position from these prosodic cues alone.

How might such dynamics operate in interactions between police and policed? The state grants officers immense power to detain, sanction, and even deploy force against citizens, but police authority is grounded in citizens' perceptions of their legitimacy (Tyler, 1990). Citizens form these impressions, in large part, from interpersonal aspects of their interactions with the police: the extent to which they perceive officers to be benevolent, respectful, and neutral actors (Lind & Tyler, 1988; Tyler, 1990). An officer's manner of interpersonal communication can contribute to this image, by expressing respect and friendliness, or undermine it with condescension and coldness. In turn, these perceptions guide citizens' assessments in subsequent police encounters (Tyler & Huo, 2002), although it remains to be seen whether officer prosody is one vehicle through which citizens' perceptions are formed.

Evidence from other institutional settings suggests that police officers may adopt a more respectful, inclusive tone toward White citizens. Doctors (R. L. Johnson et al., 2004) and teachers (Simpson & Erickson, 1983) use a more positive tone with White interlocutors, and subtle racial biases in paraverbal and nonverbal behavior operate across a wide range of interactions (Dovidio et al., 1997, 2002; Hebl & Dovidio, 2005). Given the relational basis of police legitimacy, racial disparities in prosodic cues could undermine police-community trust, yet these aspects of officer communication have gone heretofore unexamined.

The Present Research

Using BWC recordings, we test for racial disparities in the content of officers' communication, and experimentally expose citizens to different sets of police interactions to understand their effects. We examine the role of prosody in the most common

police-citizen interaction in the United States: the traffic stop (Davis et al., 2018; Eith & Durose, 2011). Since disparities in police treatment (Baumgartner et al., 2018; Pierson et al., 2020) and citizen trust (Krogstad, 2014, Pew Research Center, 2016) are most pronounced between Black and White citizens in the United States, we focus our analysis on these two groups.

In Studies 1A, 1B, and 1C we compare the extent to which an officer's tone communicates respect, comfort, and friendliness toward Black and White men in thin slices of officer speech. By filtering out higher frequencies in the audio and removing drivers' speech, we were able to mask the content of officers' utterances and the race of their interlocutor in testing for racial disparities. We then test the consequences of this interpersonal cue for citizens' perceptions of police institutions: their mental representations of officers (Study 2) and their trust in police departments (Study 3).

To complete this cycle, alongside differences among officer-citizen interactions, we measured variation in perceivers' perspectives: their own general trust in the police and their experiences with law enforcement. Past research has found that citizens' identification with the police, itself a consequence of procedural justice (Sunshine & Tyler, 2003a), guides the interpretation of recorded police-citizen interactions (Braga et al., 2014; Granot et al., 2014). Other studies find that witnessing just or unjust interactions can, in turn, influence citizens' global assessments of the police (Maguire et al., 2017). In light of these findings, a secondary goal of this research was to consider whether perceivers who held more trust in law enforcement or reported fairer treatment would perceive the same officer prosody more favorably than those less trusting and, conversely, whether the tenor of police interactions might influence participants' general trust in the police.

Sampling Thin Slices of Officer Prosody

We sampled officer prosody from body camera footage of traffic stops conducted over the course of a month in a medium-sized U.S. city. Our corpus of footage had been previously matched to stop data, checked to ensure recordings captured the speech of the primary officer conducting the stop, and professionally transcribed.[1] To constrain the variability among interactions, we limited our analyses to routine traffic stops (i.e., those that did not result in the arrest of the driver) and stops of Black and White men, who comprised the majority of stopped drivers in the corpus. These constraints provided a pool of 433 recordings of Black male drivers and 180 recordings of White male drivers from which to sample our stimuli.

From this pool of stops, we generated thin-slice clips of officer prosody directed toward Black and White drivers. Previous studies of doctors' and teachers' tone of voice were conducted using raw audio of these interactions. However, one limitation of this approach is that listeners could discern information about the speaker's interlocutor, including their race, alongside the speaker's communication. A second limitation is that a speaker's words can interfere with the interpretation of their tone of voice, a potential confound given differences in the words officers use with Black and White drivers during traffic stops (Voigt et al., 2017).

To address limitations in prior research, here we filtered out the frequencies required to discern the content of speakers' utterances, a standard approach to isolating a speaker's prosody from their language (Rogers et al., 1971; Weisbuch & Ambady, 2008). We

further removed the drivers' speech from the audio channel, blinding the listener to the race of the officer's interlocutor. Stimuli for the present research were generated in three steps: first, by sampling interactions from the month of stops; second, by sampling a random window of officer speech within each interaction; lastly, by removing driver speech and filtering out the content of officers' utterances.

Sampling Interactions

We generated two sets of stimuli for the current research. To create our primary stimulus set, we sampled 200 stops from the month of stop data. Half of these stops were of Black drivers, and half were of White drivers. The stops in the sample were conducted by 116 officers (11 female, 105 male; 47.4% White, 19.0% Latinx, 17.2% Asian, 13.8% Black, 2.6% Other), with an average of 1.72 stops per officer ($SD = 1.59$).

Due to a coding error, six clips from stops of female drivers were sampled in this set, and driver race was unavailable for one additional clip. We created a second stimulus set of 250 clips (125 stops of Black men, and 125 stops of White men) for a replication study, Study 1C. These stops were conducted by 132 officers (15 female, 117 male; 43.9% White, 18.1% Latinx, 16.6% Asian, 18.2% Black, 3.2% Other) with an average of 1.88 stops per officer ($SD = 1.79$).

Sampling Officer Speech

In a typical traffic stop, a police officer greets the driver, gives a brief explanation for the stop, and requests their license and registration before returning to their patrol car. We sought to capture a short clip of the officer's contact with the driver in this initial phase of each interaction. For each turn of officer speech occurring in this period, we created a 10-s window, extended to the end of the nearest utterance. The clip was saved as a potential stimulus in the study if the officer spoke for at least 4 s in that window. In pilot testing stimuli, this cutoff ensured the widest possible sampling frame while capturing sufficient officer speech to render a judgment.

For each stop, we determined all possible clips meeting this criterion, then sampled one clip at random from this set. On average, clips were 11.9 s in length ($SD = 4.18$ s), with 7.46 s of officer speech ($SD = 5.81$ s). Stimuli for the replication set were sampled in the same manner, with the additional constraint that the clips could not have been sampled for the original stimulus set. Clips in the second (replication) set averaged 11.7 s in length ($SD = 3.67$ s), with 8.27 s of officer speech ($SD = 6.13$ s). Black and White-directed clips did not significantly differ in their duration or in the amount of officer speech in either stimulus set.

Isolating Officer Prosody

We subsequently edited stimulus clips to only contain officers' prosody. To blind participants to the identity of the driver, we replaced the driver's speech in each clip with brown noise,

---

[1] The corpus covers approximately 70% of stops in this period, as some stops were captured by a secondary officer's BWC (and contained no speech) and others could not be manually matched to stop metadata. Coverage does not differ by driver race ($z = -.94$, $p = .35$).

retaining only the officer's speech. On average, approximately 2.71 s of driver speech were removed ($SD = 2.34$ s), the amount of driver speech edited out did not differ across driver race. We then applied a low-pass filter to remove frequencies above 500 MHz (600 MHz for female officers). In effect, this process masks the content of the speaker's utterances, while preserving prosodic features such as the tone, rhythm, and quality of their voice (Rogers et al., 1971).

To confirm that we successfully obscured the meaning of officers' utterances, a professional transcriptionist recorded any words they could discern from a random subset of 100 content-filtered clips in our stimulus set. We compared the content-filtered transcriptions against transcriptions produced from the unfiltered BWC video to compute the Word Information Loss (Morris, 2002; Morris et al., 2004). This metric, originally from the automatic speech recognition literature, estimates the quantity of word-level information lost between the true source (in this case, transcription of the original audio) and the new estimate (transcription of the filtered audio). The mean WIL for our clips was 93.8% ($SD = 7.7$%), indicating that the vast majority of words were rendered unintelligible. WIL did not significantly differ by driver race. Of the words recognized by the transcriptionist, almost two-thirds (65%) were function words: common words that are not essential to parse the meaning of an utterance (e.g., "you," "is," "there"; Grosjean & Gee, 1987). Thus, we could be confident that our procedure masked the meaning of officers' speech, and indeed rendered the vast majority of words unintelligible.

Studies 1A–1C: Racial Disparities in Officer Prosody

In Studies 1A, 1B, and 1C, participants rated the extent to which an officer's tone sounded friendly, at ease, and respectful toward the driver, or, conversely, whether the officer talked down to the driver and in a cold or tense manner. Study 1A was conducted in a laboratory on a college campus. In Study 1B, we expanded our sample to more closely approximate the population of motorists: patrons at a Department of Motor Vehicles office located in the same city from which our interactions were sampled. Study 1C was a preregistered replication of Study 1A with the second, larger sample of content-filtered clips. All studies were conducted with Institutional Review Board (IRB) approval.

Study 1A

Based on a pilot study, we estimated that a sample of 200 clips and 80 raters would be sufficient to detect a small racial disparity ($\beta = .2$) with at least 80% power in a mixed-effects analysis (Judd et al., 2017). While we set a recruitment goal of 80 participants, we allowed all participants who had signed up for the study in advance to complete the protocol. A final sample size of 84 participants (57 female, M_{age} 23.44, $SD = 9.88$) participated in Study 1A for payment or course credit. The racial composition of the sample was 39% White, 12% Latinx, 14% Black, 26% Asian, and 8% multiracial/some other group.

Participants from a university population listened to clips in separate rooms in a 30-minute study session. To balance the demands of stimuli sampling and participant fatigue, we adopted a replication design (Judd et al., 2017; Westfall et al., 2014): participants were blocked into four "replications" of 50 stimuli each (half from

stops of Black drivers, half from stops of White drivers), such that each replication had approximately 20 participants.

Study 1B

Although the lab setting in Study 1A provided a controlled environment for judging our stimuli clips, university participants on a residential campus may have qualitatively different interactions with law enforcement than the general public. We sought to replicate Study 1A in a more representative sample of motorists. Thus, participants in Study 1B were recruited from a Department of Motor Vehicles (DMV) office located in the same city from which the stimuli were sampled.

DMV participants completed the study on tablets while they waited for service. To meet our desired power for the study, we set a recruitment goal of 200 participants for the quarter in which data was collected at the field site, allowing data collection to continue until the end of the day the recruitment goal was reached. Ultimately, 209 participants (98 female, $M_{age} = 34.65$, $SD = 12.80$) were recruited, but 180 participants completed all trials of the clip rating task. A post hoc sensitivity analysis found that this reduced sample provided 80% power to detect an effect size of at least .17. Participants who rated at least one clip ($N = 200$) are included in the pooled analysis, but constraining our sample to participants who rated all clips did not affect our outcomes of interest. The composition of the sample was 38.3% White, 11.7% Latinx, 23.3% Black, 11.7% Asian, and 15% multiracial/some other group.

Participants completed the same task as in Study 1A, with two main changes. First, each participant listened to a smaller number of clips ($N = 16$) on a tablet (running OpenSesame software; Mathôt et al., 2012). Second, where Study 1A adopted a replication design, in Study 1B, clips were selected at random from the pool of 200 clips. This design change was made to make the study easier to implement in the field, since up to eight participants could complete the study simultaneously. As a result, the number of raters per clip ranged from 7 to 23 raters.

Study 1C

Study 1C was a preregistered replication of Study 1A with a second, independently sampled set of content-filtered stimuli (preregistration materials are available at https://aspredicted.org/ty5ky .pdf). Given the effect sizes we observed in Studies 1A and 1B, we increased the sample of clips from 200 to 250 to have 80% power to detect a smaller effect ($\beta = .15$). While we specified a recruitment goal of 100 participants in our preregistration document, we decided before data analysis to let participants enroll in the study through the end of the academic quarter in which it was conducted.

Ultimately 121 university participants (84 female, $M_{age} = 21.12$, $SD_{age} = 7.33$) completed Study 1C for course credit or payment. The composition of the sample was as follows: 46.2% White, 6.6% Latinx, 7.4% Black, 22.3% Asian, and 17.4% multiracial/some other group. Stimulus set aside, the design of Study 1C was identical to Study 1A, with an additional replication of 50 clips to accommodate the larger stimulus set. Participants in Study 1C completed the experimental protocol in the same lab setting and with the same equipment as Study 1A participants.

Procedure

All participants were recruited for a study on interpersonal interactions. Upon giving consent to participate, participants were informed that the purpose of the study was to see how much information people could glean from the tone of another person's voice, and that they would be listening to interactions between police officers and drivers that had been edited so they could only hear the police officer's tone of voice. Participants were instructed to provide a gut impression of the officer's tone in each clip. They were then provided stereo headphones and proceeded through the study at a self-guided pace.

On each trial, participants first heard a content-filtered clip, then rated it on three dimensions: tension (whether the officer sounded tense or at ease), friendliness (whether the officer sounded cold or friendly toward the driver), and respectfulness (whether the officer sounded like they respected the driver or was talking down to him). For each of these dimensions, participants responded on a six-point bipolar scale (e.g., 1 = *very cold*, 6 = *very warm*), as well as a binary categorical judgment (e.g., *The officer was cold/warm toward this driver*). Lastly, participants indicated if they were unable to rate the clip along the dimension via checkbox. Participants could click a button to replay a clip as many times as they saw fit.

Upon completing the clip-rating task, participants provided demographic information. In addition, participants responded to five items assessing their general trust in the police (Tyler & Huo, 2002). Specifically, participants indicated their agreement on a 7-point scale (1 = *strongly disagree* to 7 = *strongly agree*) whether the police were generally honest, whether they usually treat people with respect, whether they respect the participant's basic rights, whether they usually listen to people's views before making a decision, and whether the participant generally trusted the police (α_{1A} = .88, α_{1B} = .81, α_{1C} = .86).

Participants in Study 1B and 1C completed three additional items pertaining to their personal experiences with the police: whether the participant had been stopped before, and, if so, how fairly they thought the officer treated them in their most recent encounter, and how positive the outcome was for them personally (5-point Likert scale ranging from 1 = *not at all* to 5 = *a great deal*).

Results

Our primary question of interest was whether officers' tone communicated more positive interpersonal treatment—less tense, more friendly, and more respectful—toward White men than Black men. Participants indicated that they were able to rate officer speech on the provided dimensions in 96.5% of trials. Participants' scale ratings were combined by averaging scores on these dimensions into a single index of perceived officer treatment (α = .84). Here, we present a pooled analysis of officer prosody from these samples (N = 412 participants), with additional analyses for each study detailed in the online supplemental materials. Data and scripts for analyses can be found at https://osf.io/89wmk/?view_only=7b2cc545e7354f83852c74fc929a5a48.

Analysis Strategy

In all analyses, we applied generalized linear mixed-effects models with cross-specified random effects for participants and stimuli (Baayen et al., 2008; Judd et al., 2017). Such an approach lets us account for sampling variability in both clips and participants in estimating the fixed effect of driver race. Since our data were pooled across three separate samples of participants, we controlled for study with a series of effects-coded contrasts in the analyses below.

We further accounted for the fact that some traffic stops were sampled in both of our stimulus sets by nesting stimulus clip within a higher-order stop term. Including this term significantly improved model fit, $\chi^2(1)$ = 13.86, p < .001, confirming that there was significant correlation among clips sampled from the same stop. However, a greater proportion of the variance in participants' responses was explained by differences among stimuli ICC_{Clip} = .13 (intraclass correlation coefficient), 95% confidence interval (CI) [.10, .17], relative to differences among individual participants $ICC_{Subject}$ = .09 [.08, .11] or at the stop level ICC_{Stop} = .08 [.04, .12].

Racial Disparities in Prosody

This model revealed that officers' interpersonal treatment differed when addressing Black and White drivers, β_{Pooled} = .19 [.10, .29], $t(326.18)$ = 3.98, p < .001. Participants perceived officers' prosody toward Whites as more positive (M = 3.72 [3.64, 3.81]) than their tone toward Black drivers, which was viewed as neither positive nor negative (M = 3.50 [3.41, 3.58]). Figure 1 plots this aggregate fixed-effect contrast, along with separate estimates for Study1A (β = .13 [.00, .26], $t(194.93)$ = 2.01, p = .05), Study 1B (β = .12 [−.02, .22], $t(195.50)$ = 1.70, p = .09), and Study 1C (β = .26 [.13, .40], $t(246.82)$ = 382, p < .001).

This difference persisted even controlling for other features of the stop (driver age, driver sex, whether a citation was issued, whether a search was conducted), officer (officer race and gender), and participant rater (age, political orientation, gender, and race), β_{Pooled} = .21 [.11, .31], $t(314.84)$ = 4.10, p < .001.[2] None of these additional covariates reached statistical significance. A comparison of our random-intercept model with a model accommodating random slopes for the race effect within each participant revealed that this effect did not differ significantly among participants $\chi^2(2)$ = 2.55, p = .28.

Participants' categorical judgments followed a similar pattern of results in mixed-effects binomial regression model (with the same random effects structure). Participants were more likely to categorize officer speech toward Black drivers clips as *talking down* (odds ratio [OR] = 1.43 [1.18, 1.73], z = 3.72, p < .001) and *tense* (OR = 1.51 [1.23e, 1.85], z = 3.97, p < .001), and less likely to categorize them as *friendly* (OR = .67 [.56, .81], z = −4.25, p < .001).

---

[2] Estimates for the fixed effect of driver race were equivalent whether the female-directed clips mistakenly sampled were included, excluded, or controlled for with a dummy variable.

Figure 1
Responses for Individual Scale Items in Study 1, by Study and Driver Race
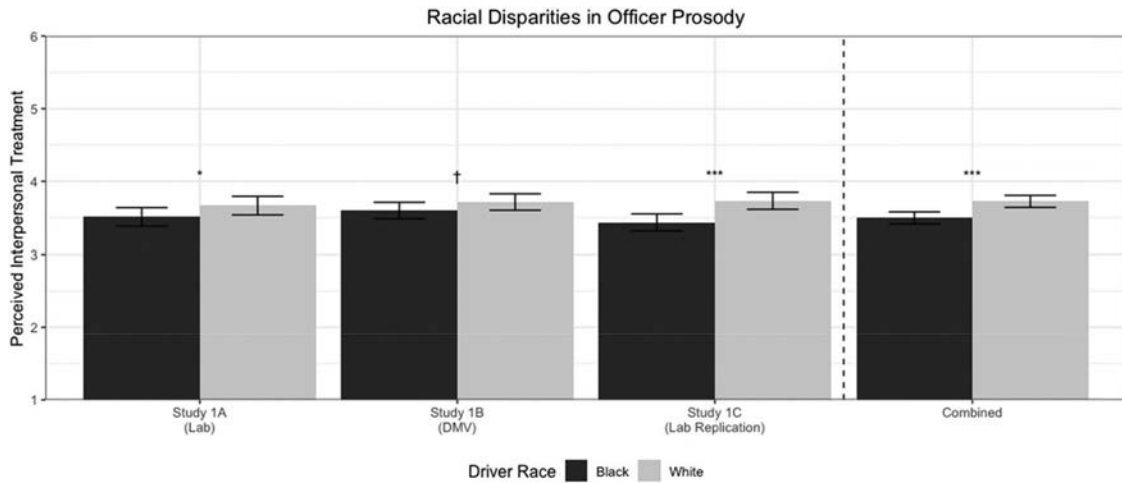
Note. Error bars represent 95% bootstrapped confidence intervals for the fixed effect of driver race; asterisks denote significant differences.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Perceiver-Level Variation in the Perception of Officer Prosody

Alongside these stimulus-level disparities in officer prosody, we examined the relationship between participant-level general trust in the police and their perceptions of these interpersonal cues. Since we exposed participants to the same stimulus set, we could measure the extent to which participants' global trust and past experiences of fairness influenced their impressions of identical stimulus clips.

Participant Global Trust

We computed a linear mixed-effects model predicting participants' interpersonal treatment ratings, with the same random effects structure as our primary model (i.e., random intercepts for participants, and intercepts for clips nested within stops). Participants with greater general trust in the police perceived officer language in the same clips more positively than participants who had less trust in the police ($\beta = .11$ [.07, .16], $t(332.8) = 5.44$, $p < .001$).

Participant Experiences. In these studies, 81.0% of the DMV sample (Study 1B, $N = 168$) and 41.3% of the replication sample (Study 1C, $N = 50$) indicated that they had been stopped by a police officer at least once in their lifetime. Among this subset of participants, we were further able to gauge the extent to which their experiences corresponded with their perceptions of officer prosody. We ran a mixed-effects model with three fixed effect terms: the clip driver race, how fairly the participant felt they had been treated in their last police encounter, and how favorable the outcome of that encounter was for the participant.

Participants who reported more fair treatment in their most recent police encounter perceived more positive interpersonal treatment in officers' prosody ($\beta = .08$ [.02, .15], $t(200.2) = 2.58$, $p = .01$), in contrast, these judgments were unrelated to the favorability of the outcome of these interactions ($\beta = .02$ [−.05, .08],

$t(195.61) = .48$, $p = .63$). Thus, participants' trust in the police—both individual attitudes and personal experiences of fairness—were associated with the interpersonal treatment they felt officers' tone conveyed.

Interim Discussion

In three studies, participants detected disparities in police officers' tone toward Black and White men. These differences were robust across stimuli and participant samples, and impressive considering that participants were blind to the content of officers' words and the race of their interlocutor. A secondary finding of Study 1 is that citizens resolve these ambiguities, in part, based on their preexisting trust in the police. That is, participants who generally trusted the police heard more warmth, respect, and ease in the same encounters than those who were more wary. These perceptions tracked participants' experiences of fairness in their most recent police encounter, and not whether the outcome was personally beneficial to them.

Since we replaced interlocutors' speech with brown noise, it is possible that participants formed their impressions from the censored driver speech and not officer prosody per se. For example, a listener could interpret a long segment of brown noise as an officer providing a citizen a chance to speak, or a driver protesting unfair treatment. To rule out this possibility, we conducted a conservative test of disparities in officer prosody among single uninterrupted officer utterances. We further constrained our sampling to the two most common speech acts officers perform in traffic stops: requesting documents (e.g., "Can I see your license and registration?") and giving reason for the stop (e.g., "I stopped you for running the stop sign"; Bayley, 1986; Prabhakaran et al., 2018).

Fifty-one undergraduate participants rated these content-filtered utterances (50 Black-directed/50 White-directed for each act) along the same dimensions as in Study 1 ($\alpha = .79$). A linear mixed-effects model with cross-specified random intercepts for

participants and stimuli and an additional fixed effect term for act (effects coded, −1 = *requesting documents*, 1 = *providing reason*) revealed that officer prosody toward White (vs. Black) male drivers communicated more positive interpersonal treatment, β = .09 [.00, .17], $t(197.1) = 1.94$, $p = .05$. This study is further detailed in the online supplemental materials.

Even when considering severely constrained clips of officer communication, then, our studies converge on the same finding: police officers communicated more respect, ease, and warmth toward the White men they stopped than they did toward Black men. Are these differences superficial, or can they influence police-community relations? We now consider the effects of officers' interpersonal communication.

From Interpersonal Treatment to Institutional Trust

Our findings in Studies 1A–1C are significant because interpersonal aspects of routine police encounters are theorized to be central to citizens' trust in law enforcement (Lind & Tyler, 1988; Sunshine & Tyler, 2003b; Tyler & Huo, 2002). Thus, racial disparities in officers' communication could feed gaps in police-community trust, and reductions in trust can, in turn lead to more negative evaluations of officers' communication. However, the experimental support for the link between officers' interpersonal communication and citizens' institutional trust is difficult to evaluate, since most tests take the form of field experiments that change multiple aspects of police behavior simultaneously (Mazerolle et al., 2013; Wood et al., 2020).

In Studies 2 and 3, we use our thin slices of officer communication to experimentally test the consequences of prosodic cues. Instead of judging audio clips in isolation, participants considered a police department after listening to multiple clips from our stimulus set to simulate repeated interactions with the police. By manipulating the subset of interactions from which these clips were sampled, we are able to test the causal effect of these interpersonal cues for citizens' views of policing institutions.

In Study 2, we test the general relationship between officers' paraverbal behavior and citizens' representations of the police. Using reverse correlation (Dotsch & Todorov, 2012), we measured participants' mental image of the typical police officer in a city police department, by randomly exposing participants to the either most positively rated or most negatively rated interactions, we could test how these interpersonal cues informed these images. We predicted that participants who heard more positive prosody would represent officers in an agency as more procedurally fair than those exposed to more negative tones of voice.

In Study 3, we directly test whether racial disparities in officer prosody cause reduced trust in police departments by manipulating whether community participants were exposed to either Black-directed or White-directed officer prosody. In addition to testing whether perceptions of officer prosody mediated the relationship between race and institutional trust, Study 3 served as a conceptual replication of Study 1 with a between-subjects design.

Study 2: Prosody and Representations of Institutional Agents

When you picture a police officer in your town, do you see an honest protector, or an untrustworthy agent to be avoided? In Study 2, we used reverse correlation (Dotsch & Todorov, 2012) to test whether one's answer to this question, in part, depends on how officers communicate with the public. Reverse correlation uses participants' judgments of visual stimuli to construct a classification image capturing their mental representation of a particular group or dimension of interest. Since participants can use whatever criteria they wish in these judgments, this approach is thought to measure the spontaneous use of social information, without directly probing the dimension of interest (Brinkman et al., 2017). This technique provided a conservative test of whether interpersonal cues as subtle as prosody could influence citizens' representations of the police.

We assessed participants' representations of a typical police officer from a department consisting of either the 40 most positively rated clips from Study 1A, or the 40 most negatively rated ones. Black-directed clips were more likely to be included in the negative-prosody department (OR = 2.16, [1.06, 4.53], $p = .04$), and White-directed clips slightly more likely to be included in the positive-prosody department (OR = 1.89, [.94, 3.92], $p = .08$).

Method

Data collection for Study 2 proceeded in two stages. First, in the *image-generation phase*, participants alternated between listening to audio clips of officer prosody and selecting officer images they thought belonged to the same department. From their responses, we generated a separate classification image for each condition. In the *image-rating phase*, a separate set of participants compared these images along several dimensions relating to trust and fairness.

Image-Generation Phase

Data for the image generation phase was collected over the course of a single academic quarter. Ultimately, 110 participants completed the image-generation phase of Study 2 for payment or course credit ($M_{age} = 21.8$, $SD = 7.1$; $N = 56$ female). The racial composition of the sample was 41.8% White, 21.8% Asian, 15.5% Latinx, 14.5% Black, 1.8% Native American, and 5% multiracial or some other race. Due to a computer error, incomplete data for the reverse correlation task was recorded for one participant.

Participants were instructed that they would be forming impressions of police officers in a medium-sized city, "Westmore." Participants alternated between a listening task, in which they listened to content-filtered clips of traffic stops ostensibly conducted in Westmore, and a reverse correlation task, in which they chose images that matched their mental image of the typical Westmore police officer. In each block, participants listened to 10 content-filtered clips, followed by 100 trials of the reverse correlation task. Participants completed four blocks of the task (i.e., listened to 40 clips and completed 400 trials of the reverse correlation task). As a manipulation check, participants provided their impressions of the clips they heard, then provided demographic information before being paid and debriefed. These items included the same trust items used in Study 1 (α = .89).

Listening Task. In the listening phase of each block, participants listened to 10 content-filtered clips, ostensibly of traffic stops conducted in Westmore. We randomly assigned participants to one of two conditions. In the positive prosody condition, participants listened to the 40 clips rated most positively in Study 1A, in

the negative prosody condition, participants listened to the 40 clips from the most negative quintile of Study 1A. While participants heard all 40 clips over the course of the experiment, the order in which they were presented was random. Each clip played once, followed by a chime. Upon hearing the chime, participants pressed a key to advance to the next clip in the block. After listening to ten clips, the participant switched to the image classification task.

Image Classification Task. In a reverse correlation design, participants classify many pairs of images along a dimension of interest, which are then used to create the classification image corresponding to that dimension. Each pair is derived from a single base image: a layer of random sinusoidal noise is superimposed on the base image to form one stimulus, and the inverse pattern is superimposed to form the other stimulus. The patterns participants select are averaged and superimposed on the base image to obtain the classification image (for an overview, see Brinkman et al., 2017).

We created a base image using personnel photos of officers in the same department from which the prosodic clips were sampled. We randomly sampled officers from the personnel list, choosing the first ten officers who had a neutral expression and who did not wear glasses. These 10 faces were then averaged together using photo-morphing software to create a single base image. We generated 400 pairs of stimuli for the image classification task using the *rcicr* R package (Dotsch, 2016).

On each trial of this task, participants viewed a pair of images, presented side by side. Participants were asked to choose which officer of the pair was in the Westmore Police Department. The order in which stimuli pairs were presented over the course of the study was randomized. Participants completed 100 trials of the classification task in each block, and then switched back to the listening task.

Judgments of Prosody. As a manipulation check, participants were asked to rate the extent to which the officers in the clips they heard *communicate respect toward drivers, talk down to drivers* (reverse-coded), *sound at ease with drivers*, *sound tense with drivers* (reverse-coded), *sound friendly toward drivers*, and *sound cold toward drivers* (reverse-coded) on a 1 (*almost never*) to 5 (*almost always*) scale ($\alpha = .88$). As anticipated, participants in the positive prosody condition viewed Westmore officers' interpersonal treatment more positively ($M = 3.61$, [3.43, 3.78]) than participants in the negative prosody condition ($M = 2.80$, [2.61, 2.97], $t(108) = 6.51$, $p < .001$, $d = 1.06$).

Participant Global Trust. As an ancillary analysis, we examined how participants' global trust in the police might relate to their perceptions of officer prosody. Since participants in the same condition were exposed to identical stimuli, we could measure the association between participants' global trust in the police and their perceptions of officer prosody, as in Studies 1A–1C. However, we also randomly assigned participants to different sets of stimuli, it could be the case that, more positive interactions would engender more global trust in the police (Maguire et al., 2017).

Global Trust and the Perception of Prosody. To determine if participants' global trust corresponded to their perception of prosody, we regressed judgments of prosody on participants' global trust in the police, controlling for condition (effects coded, $-1 = $ negative, $1 = $ positive). Participants with greater trust in the

police rendered more positive judgments of prosody for the same stimuli ($\beta = .35$ [.20, .50], $t(107) = 4.55$, $p < .001$).

Exposure to Positive/Negative Prosody and Global Trust. It is also possible that our manipulation actually influenced participants' global trust. Consistent with this account, participants randomly assigned to the positive prosody condition reported greater global trust in the police ($M = 4.61$, [4.27, 4.95]) than participants who were assigned to the negative prosody condition ($M = 4.02$, [3.68, 4.35]), $t(108) = 2.46$, $p = .02$, $d = .45$).

Image-Rating Phase

A second set of participants compared the classification images from the two prosody conditions on a range of dimensions relevant to trustworthiness. We created a classification image for each participant in the image-generation phase, based on the images they identified as Westmore police officers. This design let us incorporate variability among participant representations from the first phase of the experiment in the image-rating phase of the study. Comparing classification images aggregated at the participant level reduces the likelihood of Type I error relative to classification images created at the condition level (Cone et al., 2020). However, a comparison of condition-level classification images obtained similar results and are provided in the online supplemental materials for the interested reader.

We preregistered all methods and analyses for the image-rating portion of the study (preregistration materials are available at https://aspredicted.org/v5vu2.pdf). To have at least 80% power to detect a small difference between the images ($\beta = .15$), we set a recruitment goal of 150 usable participants. One hundred seventy-five participants were recruited, but 18 failed an attention check and were excluded before analysis, resulting in a final sample size of 157 ($M_{age} = 33.69$, $SD = 11.43$; $N = 76$ female). This sample was 7.6% Asian, 2.5% Black, 6.4% Latinx, 1.9% Native American, 74.5% White, and 7.0% multiracial or some other race.

Procedure

Participants were recruited from Prolific Academic for an online study on impressions of organizations, and told that they would compare composite images of officers from different police departments. Participants then compared five pairs of classification images. On each trial, participants viewed two classification images side-by-side, one classification image selected at random from a participant in the positive prosody condition, and one from a participant in the negative prosody condition. The order in which the two classification images were positioned was randomized for each trial.

Participants indicated which officer was more likely to engage in a range of behaviors, both procedurally fair (engaging in community policing, treating stopped drivers with respect) and unjust (treating citizens rudely, being accused of racial profiling). Participants used a slider to which composite was more likely to engage in the behaviors on a 7-point scale (1 = *definitely the left image*, 4 = *neither/equally likely*, 7 = *definitely the right image*).

Results

Figure 2 displays the results for all items, along with a sample of classification images from each condition. We used a linear-

Figure 2
A Sample of Subject Classification Images From the Image-Generation Phase of Study 2 Alongside Rating-Phase Participants' Judgments
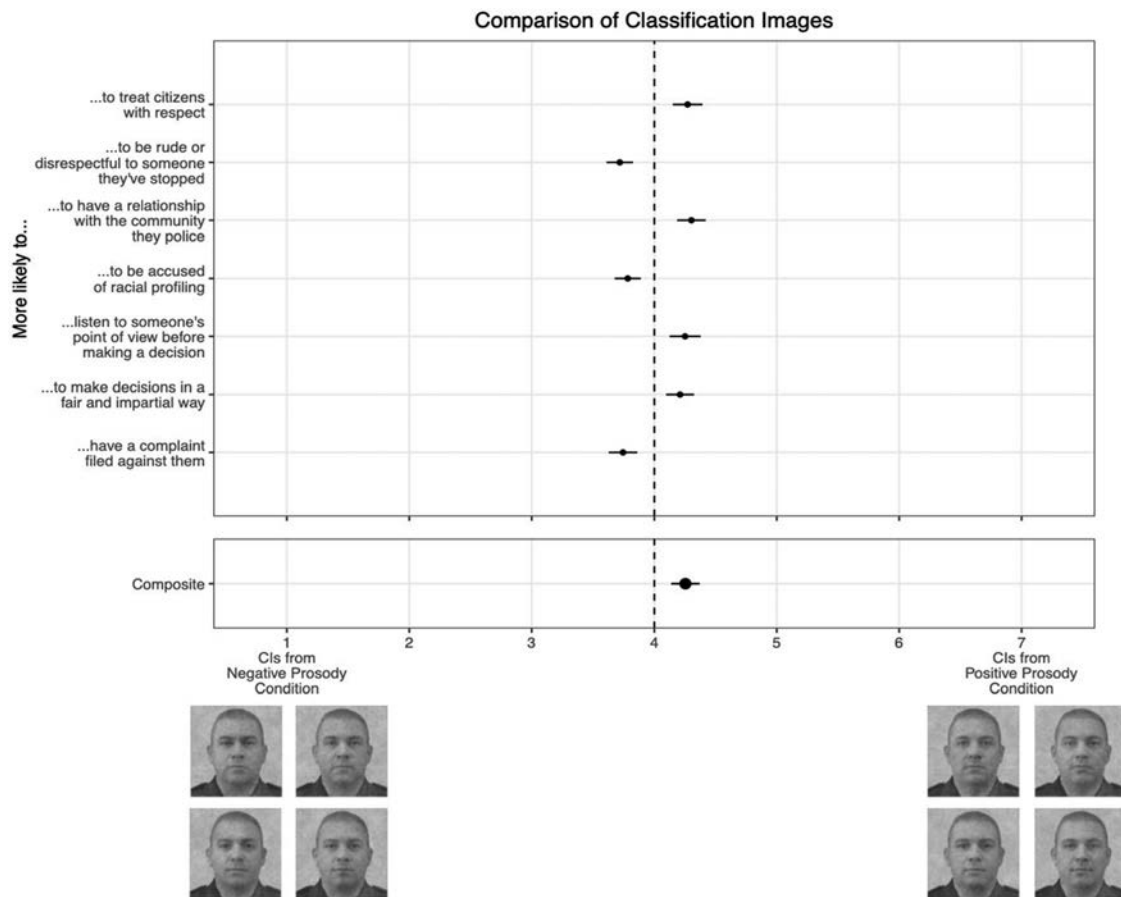
Note. Bars represent 95% confidence intervals of the estimated marginal mean, and the dotted line represents the point of indifference between the classification images.

mixed effects regression to test whether participants' ratings were significantly higher (classification image from positive prosody condition more likely) or lower (classification image from negative prosody condition more likely), that is, whether the intercept of this model significantly differed from the midpoint of the scale (neither classification image more likely). To account for the variation among our participants and stimuli, we included random intercept terms for stimuli and participants.

Since items were highly correlated with each other ($\alpha = .95$), they were combined into a single index of procedural fairness. Classification images generated by participants in the positive prosody condition were judged as more procedurally fair than classification images created by participants in the negative prosody condition ($\beta = .22$ [.18, .34], $t(50.3) = 4.36$, $p < .001$).

Discussion

The results of Study 2 demonstrate that interpersonal cues such as prosody can influence the way citizens represent institutional agents such as the police. Data from the image generation phase not only suggests that participants' global trust in the police was correlated with judgments of prosody as in Studies 1A–1C, but also that these attitudes themselves differed between experimental conditions.

Participants in the image-generation phase of Study 2 were free to use any criteria they wished when selecting images; nonetheless, those who were exposed to positive prosodic cues generated more trustworthy representations than those exposed to negative prosody. These findings provide an extension to procedural justice theory: Disrespectful or positive encounters not only affect social perceptions of legal actors, such encounters affect visual representations of these actors as well. Our data also build on recent work by Lloyd and colleagues (2020) comparing Black and White citizens' representations of the police using similar reverse correlation procedures. The authors found that Black participants generated representations of the police that were judged to be more dominant, biased, aggressive, and generally more negative than White subjects.

Together with the results of Study 1, Study 2 suggests that officers' communication may be one factor that contributes to these disparate representations. Indeed, while our prosody sampling was conditioned on treatment and not race, the disparities observed in

Study 1 meant that the positive prosody and negative prosody conditions in Study 2 differed in their racial composition. In Study 3, we directly test the consequences of Black-directed versus White-directed prosody for community trust.

Study 3: Racial Disparities in Prosody and Institutional Trust

Given the racial disparities in prosodic cues we observed in Study 1, the results of Study 2 would suggest that the divergent interactions Black and White men have with police officers engender different levels of institutional trust in the police. In Study 3, we tested this hypothesis in a diverse community sample of DMV patrons. We again exposed participants to repeated interactions, this time sampled from officer interactions with Black versus White men.

Procedure

Participants were recruited from the same DMV office as in Study 1B. To have at least 80% power to detect a moderate-sized effect ($d = .30$), we set a recruitment goal of 280 participants for the academic quarter in which the study was conducted. Two hundred eighty-one people participated in the study within this period, but 34 did not complete dependent measures, due to technical difficulties or being called for service, leaving a final sample size of 247 participants ($M_{age} = 35.8$, $SD = 14.3$). The demographic composition of the sample was as follows: 53.7% male, 43.4% female, 2.9% transgender/nonbinary/declined to answer, and 31.3% White, 33.1% Black, 12.5% Asian/Asian American, 9.3% Latinx, and 6.8% multiracial/other; 80% of participants had been stopped by the police at least once before.

Participants were recruited for a study on community interactions. Upon giving consent, participants were told that they would hear filtered excerpts of traffic stops from a single police department, and that their task was to form an impression of the department from those interactions. Participants proceeded to listen to 20 clips from the same stimulus set used in Studies 1A/1B and Study 2. We manipulated whether these clips were sampled from the pool of stops of Black men, or from the pool of stops of White men. After listening to the clips, participants advanced to the dependent measures of interest, then completed the same demographic measures as in Study 1B. These measures included the same scale items of participants' global trust in the police ($\alpha = .86$), as well as their perceptions of the fairness and outcome of their most recent police encounter.

Perceptions of Officer Prosody

First, participants provided their impressions of the set of clips that they heard. Using the same six-point bipolar scales as in Study 1, participants rated how generally *respectful*, *warm*, and *tense* the clips they heard were. As in Study 1, we averaged these judgments into a single composite ($\alpha = .68$).

Institutional Trust

Next, participants completed five items measuring their trust in the police department from which the clips came ($\alpha = .91$): for example, the extent to which the police department cared about

the community, the degree to which officers in that department tried to do right by the people they served, and the level of trust they would have in that department's officers to treat them fairly. For each item, participants responded on a 5-point scale, ranging from 1 = *not at all* to 5 = *a great deal*.

Results

Racial Disparities in Officer Prosody

In Study 1, we observed both racial disparities in officer tone and participant-level correlations between police trust and perceptions of these prosodic cues. We sought to conceptually replicate these findings in Study 3. Instead of judging individual thin slices of officer speech, however, participants gave their general impressions after listening to repeated officer interactions with either White men or Black men. Participants who were exposed to repeated White-directed prosody reported more positive officer interpersonal treatment ($M = 4.00$, $SD = 1.21$) than those who heard speech directed toward Black men ($M = 3.67$, $SD = 1.04$; $t(245) = 2.37$, $p = .02$, $d = .30$).

Perceiver Differences and Officer Prosody

Consistent with our previous findings, we found that participants' judgments of officer prosody were correlated with their global trust in the police and their personal experiences of fairness. Participants who had greater trust in the police judged the clips they heard more positively than those who were less trusting ($\beta = .35$ [.24, .27], $t(245) = 5.94$, $p < .001$). Also consistent with our findings in Study 1, among participants who had been stopped by the police in the past, the procedural fairness they experienced at the hands of the police predicted evaluations of officer prosody ($\beta = .27$ [.07, .46], $t(194) = 2.74$, $p = .01$), where the favorability of the outcome did not ($\beta = .02$ [−.17, .21], $t(194) = .25$, $p = .80$).

Given the potential effects of our manipulation in Study 2 on participants' global trust, and the racial disparities in our stimuli, it was possible that our perceivers' trust would differ as a function of experimental condition. However, exposure to Black versus White-directed speech had no effect on participants' global trust ($\beta = .20$ [−.05, .47], $t(246) = 1.60$, $p = .12$), or self-reported fairness ($\beta = .17$ [−.11, .45], $t(196) = 1.20$, $p = .23$) or outcome ($\beta = .07$ [−.21, .36], $t(196) = .52$, $p = .61$) in past encounters.

To determine whether differences in institutional trust were driven by perceptions of officer prosody, we performed a mediation analysis (see Figure 3). Participants' perceptions of prosody were correlated with their trust in the police department, $\beta = .63$ [.53, .73], $SE = .05$, $t(245) = 12.92$, $p < .001$. After accounting for these perceptions, the relationship between condition and institutional trust was reduced to nonsignificance, $b = .10$, $SE = .10$, $p = .32$. Following the recommendations of Shrout and Bolger (2002), we tested the indirect effect of perceptions of prosody on institutional trust with 10,000 bootstrapped samples using the *mediation* R package (Tingley et al., 2014). The indirect effect of condition on institutional trust through prosody was significant, $b = .19$ [.03, .34], $p = .02$, mediating 66% of the direct effect. In other words, disparities in institutional trust were driven by the differences in officers' prosodic cues in their interactions with White and Black men.

Figure 3

Mediation Model Depicting the Relationship Between Participant Condition in Study 3 (0 = Stops of Black Men, 1 = Stops of White Men) and Institutional Trust in the Police Department, as Mediated by Perceptions of Officer Tone
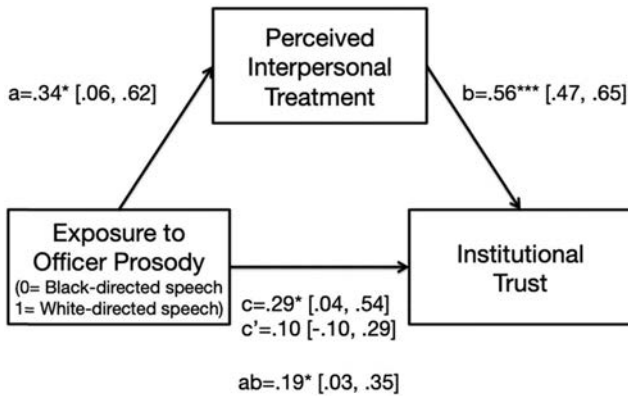
Note. Asterisks denote significant coefficients.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

Police officers interact with the public in a dual role. A police department cannot be friendly, deferential, unwelcoming, or condescending, but an officer's tone can communicate these qualities to citizens. At the same time, these interpersonal qualities carry unique consequences when they are expressed by individuals who are deputized by the state. The present findings contribute to our social psychological understanding of how such agents' interpersonal communication establishes or undermines citizens' institutional trust, and how racial disparities in these encounters hinder police-community relations. Focusing on one channel of communication (prosody) we demonstrate a cycle that contributes to these gaps.

First, police officers communicate with White and Black citizens in divergent ways. White men in our stimulus set experienced more friendly and respectful prosody than Black men. That is not to say that officers are invariably rude to Black citizens and unfailingly polite to Whites; participants judged most officer speech in our stimulus set, both Black and White-directed, to be neutral in nature. However, racial disparities in prosodic cues persisted after accounting for variability among stimuli and raters, controlling for a range of covariates, and blinding participants to both the driver's race and the content of the officer's words. Where previous research identified disparities in what officers say (Voigt et al., 2017), here we find that these differences extend to the manner in which those words are expressed, even at the level of single utterances.

Such findings are meaningful because interpersonal cues accumulate across citizens' interactions with law enforcement. Consistent with this perspective, while we observed relatively small racial disparities at the individual clip level in Studies 1A–1C, the magnitude of this difference was much greater in participants' judgments of multiple police-citizen interactions in Study 3. Moreover, even these paraverbal aspects of officer communication influence citizens' perceptions of the police. The present work provides the first experimental evidence of how community members

generalize their impressions of interpersonal cues to their perceptions of institutions: what typical institutional actors are like (Study 2) and whether those institutions deserve their trust (Study 3). Even in cases where Black and White citizens receive the same legal outcome, then, our results suggest that differences in the interpersonal aspects of these encounters can contribute to racial gaps in police-citizen trust.

While not the primary focus of this work, we also note relationships between perceivers' views of the police and their perceptions of officer prosody. Raters' personal experiences of officer fairness (but not positive legal outcomes), were associated with more positive perceptions of officers' tone. Likewise, perceivers who were generally more trusting of the police rated identical stimuli more favorably than those who expressed lower levels of trust. Our findings in Studies 2 and 3 offer mixed support for the possibility that interpersonal cues could themselves influence trust in the police generally as well as in specific institutions. We cannot adjudicate between these accounts in our data, but note that both patterns are consistent with prior research, some of which relate citizens' identification with the police to their interpretation of ambiguous police-citizen altercations (Granot et al., 2014) and others that observe changes in global views of the police in response to specific encounters (Maguire et al., 2017). Coupled with our experimental findings, these data suggest that a citizen's perception of officer communication in a particular interaction is influenced by their previous encounters and affect subsequent ones, an account which merits further study.

Our findings add to a small but emerging experimental literature on procedural justice in policing. Past experiments have taken one of two forms: randomized control trials of procedural justice policing (Mazerolle et al., 2013; Sahin et al., 2017) or vignette experiments that contrast procedurally just versus unjust interactions (D. Johnson et al., 2017; Maguire et al., 2017; Reisig et al., 2018). In many respects, we were able to combine the strong causal inferences afforded by lab experiments with the ecological validity of field observations. By content-filtering officer speech, we isolated a uniquely interpersonal cue in police-citizen interactions, where field experiments combine other elements of procedural justice with interpersonal respect (such as providing a legitimate reason for the stop; Mazerolle et al., 2013; Sahin et al., 2017). In contrast to experiments that have compared a small number of researcher-created scenarios (e.g., Maguire et al., 2017) we sampled a large set of stimuli from actual police encounters. This procedure allowed us to account for variability among stimuli; an important but often overlooked aspect of experimental design (Judd et al., 2012; Westfall et al., 2014; see also Monin & Oppenheimer, 2014).

Above all, this technique let us expose participants to encounters that they would not experience otherwise. A diverse set of participants detected similar disparities in officer tone, and were similarly affected by them. Some of the same design choices that provided experimental control came at the cost of mundane realism, however. We exposed participants to thinner slices (both in duration and in the available channels of information) of a larger number of police encounters in a shorter amount of time than they would experience outside of the experiment. In the context of the current work, this provided a well-powered demonstration of a minimal interpersonal cue. Capturing processes that unfold over one's lifetime in the duration of an experiment is always a challenge, but future work can consider whether fewer, "thicker"

slices—longer recordings, or unfiltered audio, for example—accumulate in the manner we observed. As a second point of departure, our participants were necessarily third party observers of BWC recordings of Black and White men; they were a step removed from the actual subjects of those stops. However, we note that studies in which participants are vicariously exposed to encounters (Reisig et al., 2018) render similar salutary effects of procedural justice as field experiments that survey citizens immediately after they interact with the police (Mazerolle et al., 2013).

Body Camera Footage as Data, Stimuli, and Treatment

Alongside these theoretical advancements, we present a novel methodology for examining police-community interactions more broadly. Police records tell us little about the interpersonal aspects of officer behavior, but body cameras capture the nuances of their conversations. Citizen self-reports cannot decouple citizens' recollections of police encounters from the exchanges themselves, but BWC footage lets us expose individuals to encounters outside of their own experience. It provides a bridge between rich literatures in psychology, such as social–cognitive work on biases in nonverbal communication and procedural justice research on policing, as well as a link to other disciplines, such as criminal justice and sociolinguistics.

BWC footage can serve as a "cultural snapshot" (Weisbuch et al., 2009, 2017) for police-citizen interactions: to reveal patterns in citizen interactions (as data), their relation to perceiver-level factors in their interpretation (as stimuli), and their causal effects on institutional trust (as treatment). Sampling thin slices from traffic stops, we not only extend research on the content of officers' communication with the public (Voigt et al., 2017), but also probe how participants' own beliefs correspond to their perceptions of these cues, and demonstrate their consequences for police-community relations. We are agnostic here as to the specific acoustic features underlying perceivers' judgments, but this question suggests a natural point of connection with linguistic research on particular aspects of intonation and their association with social judgments. Research on the linguistic features underlying evaluations in the policing context can further link these bodies of work in the future.

In addition to its ecological validity, an attractive feature of our approach is that it incorporates the heterogeneity present among police-citizen interactions, since stimuli themselves are sampled from a distribution of police encounters. We chose to sample our stimuli from traffic stops because they are the most common interaction citizens have with law enforcement (Davis et al., 2018). This decision was tailored to our focus on everyday interactions, and the availability of prosody throughout the behavioral stream (Ambady et al., 2000). However, with an adjusted sampling regime, BWC footage could be used to analyze low-frequency but high-impact events that disproportionately impact communities of color, such as escalation: one might sample encounters from a wider range of time, and upweight encounters that involved the use of force, for example. Although the sampling choices would necessarily vary based on the question being investigated, BWC recordings open up important aspects of policing to examination that were previously closed to researchers.

Our corpus had an insufficient number of stops to adequately test for racial disparities in officers' tone directed toward female drivers. This pattern is not unique to our particular sample; men are more generally likely than women to be stopped by the police (Davis et al., 2018; Davis & Harrell, 2020). However, future research ought to examine whether the disparities we observed here generalize to stops of female drivers. Indeed, past work offers competing predictions as to how race and gender might interact in the interpersonal aspects of routine police encounters. Consistent with a target male hypothesis (Sidanius & Pratto, 2001), some studies would suggest that Black men would be particularly singled out as threats (Wilson et al., 2017) or elicit officers' suspicions of criminality (Eberhardt et al., 2004; Glaser, 2015), however, Voigt et al. (2017) found that racial disparities in officers' language were of equivalent size for men and women. In light of conflicting predictions, the intersection of race and gender in police-citizen interactions is both a theoretically rich and socially important area for future study.

These considerations are important not just for contextualizing the current research, but for guiding the future use of BWC footage in social scientific research. In this work we have shown how body camera footage, as recorded daily in large quantities by departments across the country, can make the interpersonal dimensions of police interactions visible: both the content of officers' communication and their causal influence on community trust. New methods pose novel challenges, but they also provide an unprecedented means to understand the psychology of policing and being policed.

Breaking Cycles of Mistrust

Over 60 million Americans make contact with law enforcement each year (Davis & Harrell, 2020). Police officers are simultaneously representatives of the state and the human face of the law; as a result, these interpersonal interactions have institutional consequences. Racial disparities in cues as subtle as an officer's tone of voice can shape citizens' trust in the police and alter their interpretations of subsequent encounters. Considering the frequency of these encounters and the richness of these interactions, our findings attest to the role everyday interactions play in building or eroding police-community trust across race.

The cyclical nature of these relationships illustrates how patterns of racial inequality can be self-reinforcing. At the same time, we find that they are not set in stone, but sensitive to how officers interact with the public. Programs that intervene on routine interactions with the police may be particularly well-suited for breaking this cycle. The approach we outline here can use body-worn camera footage not only to shed light on the everyday experiences of race, policing, and being policed, but also to inform and assess efforts to change those realities.

References

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In M. P. Zanna (Ed.), *Advances in experimental social psychology*, (Vol. 32, pp. 201–271). Academic Press.

Ariel, B., Farrar, W. A., & Sutherland, A. (2015). The effect of police body-worn cameras on use of force and citizens' complaints against the police: A randomized controlled trial. *Journal of Quantitative Criminology*, *59*(4), 390–535. https://doi.org/10.1007/s10940-014-9236-3

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baumgartner, F. R., Epp, D. A., & Shoub, K. (2018). *Suspect citizens: What 20 million traffic stops tell us about policing and race*. Cambridge University Press.

Bayley, D. H. (1986). The tactical choices of police patrol officers. *Journal of Criminal Justice*, *14*(4), 329–348. https://doi.org/10.1016/0047-2352(86)90126-1

Belin, P., Bestelmeyer, P. E., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40*(2), 531–539. https://doi.org/10.3758/brm.40.2.531

Braga, A. A., Winship, C., Tyler, T. R., Fagan, J., & Meares, T. L. (2014). The salience of social contextual factors in appraisals of police interactions with citizens: A randomized factorial experiment. *Journal of Quantitative Criminology*, *30*(4), 599–627. https://doi.org/10.1007/s10940-014-9216-7

Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology. *European Review of Social Psychology*, *28*(1), 333–361. https://doi.org/10.1080/10463283.2017.1381469

Cheng, J. T., Tracy, J. L., Ho, S., & Henrich, J. (2016). Listen, follow me: Dynamic vocal signals of dominance predict emergent social rank in humans. *Journal of Experimental Psychology: General*, *145*(5), 536–547. https://doi.org/10.1037/xge0000166

Cone, J., Brown-Iannuzzi, J. L., Lei, R., & Dotsch, R. (2020). Type I error is inflated in the two-phase reverse correlation procedure. *Social Psychological & Personality Science*. Advance online publication. osf.io/bvtzg

Culpeper, J. (2011). It's not what you said, it's how you said it!": Prosody and impoliteness. *Discursive Approaches to Politeness*, *8*, 57–83.

Depaulo, B. M., & Friedman, H. S. (1998). Nonverbal communication. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 3–40). McGraw-Hill.

Davis, E., & Harrell, E. (2020). *Contacts between police and the public, 2018*. Bureau of Justice Statistics.

Davis, E., Whyde, A., & Langton, L. (2018). *Contacts between police and the public, 2015*. Bureau of Justice Statistics.

Dotsch, R. (2016). rcicr: Reverse correlation image classification toolbox. R package version 0.3.4.1 [Computer software]. https://cran.r-project.org/package=rcicr

Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological & Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68. https://doi.org/10.1037/0022-3514.82.1.62

Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*(5), 510–540. https://doi.org/10.1006/jesp.1997.1331

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology*, *87*(6), 876–893. https://doi.org/10.1037/0022-3514.87.6.876

Eith, C. A., & Durose, M. R. (2011). *Contacts between police and the public, 2008*. Washington, DC: US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Epp, C. R., Maynard-Moody, S., & Haider-Markel, D. P. (2014). *Pulled over: How police stops define race and citizenship*. University of Chicago Press.

Gelman, A., Fagan, J., & Kiss, A. (2007). An analysis of the New York City police department's "stop-and-frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association*, *102*(479), 813–823. https://doi.org/10.1198/016214506000001040

Glaser, J. (2015). *Suspect race: Causes and consequences of racial profiling*. Oxford University Press.

Goel, S., Rao, J. M., & Shroff, R. (2016). Precinct or prejudice? Understanding racial disparities in New York City's Stop-and-Frisk Policy. *The Annals of Applied Statistics*, *10*(1), 365–394. https://doi.org/10.1214/15-AOAS897

Granot, Y., Balcetis, E., Schneider, K. E., & Tyler, T. R. (2014). Justice is not blind: Visual attention exaggerates effects of group identification on legal punishment. *Journal of Experimental Psychology: General*, *143*(6), 2196–2208. https://doi.org/10.1037/a0037893

Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition*, *25*(1–2), 135–155. https://doi.org/10.1016/0010-0277(87)90007-2

Hall, J. A., Coats, E. J., & LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin*, *131*(6), 898–924. https://doi.org/10.1037/0033-2909.131.6.898

Hebl, M. R., & Dovidio, J. F. (2005). Promoting the "social" in the examination of social stigmas. *Personality and Social Psychology Review*, *9*(2), 156–182. https://doi.org/10.1207/s15327957pspr0902_4

Hyland, S. (2018). *Body-worn cameras in law enforcement agencies, 2016*. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice.

Jeong, S. (2016). Conventions in prosody for affective meanings: Non-canonical terminal contours in English polar interrogatives. *Speech Prosody*, 907–911.

Jeong, S. (2017). Iconization of sociolinguistic variables: The case of archetypal female characters in classic Hollywood cinema. In A. Zirker, M. Bauer, O. Fischer, & C. Ljungberg (Eds.), *Dimensions of iconicity* (pp. 263–286). John Benjamins Publishing Company.

Johnson, D., Wilson, D. B., Maguire, E. R., & Lowrey-Kinberg, B. V. (2017). Race and perceptions of police: Experimental results on the impact of procedural (in)justice. *Justice Quarterly*, *34*(7), 1184–1212. https://doi.org/10.1080/07418825.2017.1343862

Johnson, R. L., Roter, D., Powe, N. R., & Cooper, L. A. (2004). Patient race/ethnicity and quality of patient–physician communication during medical visits. *American Journal of Public Health*, *94*(12), 2084–2090. https://doi.org/10.2105/AJPH.94.12.2084

Jones, K. A., Crozier, W. E., & Strange, D. (2017). Believing is seeing: Biased viewing of body-worn camera footage. *Journal of Applied Research in Memory & Cognition*, *6*(4), 460–474. https://doi.org/10.1016/j.jarmac.2017.07.007

Jones, K. A., Crozier, W. E., & Strange, D. (2019). Look there! The effect of perspective, attention, and instructions on how people understand recorded police encounters. *Behavioral Sciences & the Law*, *37*(6), 711–731. https://doi.org/10.1002/bsl.2441

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

Knowles, J., Persico, N., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, *109*(1), 203–229. https://doi.org/10.1086/318603

Ko, S. J., Sadler, M. S., & Galinsky, A. D. (2015). The sound of power: Conveying and detecting hierarchical rank through voice. *Psychological Science*, *26*(1), 3–14. https://doi.org/10.1177/0956797614553009

Kraus, M. W., Park, J. W., & Tan, J. J. X. (2017). Signs of social class: The experience of economic inequality in everyday life. *Perspectives on Psychological Science*, *12*(3), 422–435. https://doi.org/10.1177/1745691616673192

Krogstad, J. M. (2014). *Latino confidence in local police lower than among whites*. Pew Research Center.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. Springer Science & Business Media.

Liscombe, J., Venditti, J., & Hirschberg, J. (2003). *Classifying subject ratings of emotional speech using acoustic features*. Eighth European Conference on Speech Communication and Technology.

Lloyd, E. P., Sim, M., Smalley, E., Bernstein, M. J., & Hugenberg, K. (2020). Good cop, bad cop: Race-based differences in mental representations of police. *Personality and Social Psychology Bulletin*, *46*(8), 1205–1218. https://doi.org/10.1177/0146167219898562

Maguire, E. R., Lowrey, B. V., & Johnson, D. (2017). Evaluating the relative impact of positive and negative encounters with police: A randomized experiment. *Journal of Experimental Criminology*, *13*(3), 367–391. https://doi.org/10.1007/s11292-016-9276-9

Major Cities Chiefs Major County Sheriffs. (2015). *Survey of technology needs–Body worn cameras*. Author.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mazerolle, L., Antrobus, E., Bennett, S., & Tyler, T. R. (2013). Shaping citizen perceptions of police legitimacy: A randomized field trial of procedural justice. *Criminology*, *51*(1), 33–63. https://doi.org/10.1111/j.1745-9125.2012.00289.x

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PLoS ONE*, *9*(3), e90779. https://doi.org/10.1371/journal.pone.0090779

Monin, B., & Oppenheimer, D. M. (2014). The limits of direct replications and the virtues of stimulus sampling. *Social Psychology*, *45*(4), 299–300.

Morris, A. C. (2002). *An information theoretic measure of sequence recognition performance*. IDIAP.

Morris, A. C., Maier, V., & Green, P. (2004). *From WER and RIL to MER and WIL: Improved evaluation measures for connected speech recognition*. Eighth International Conference on Spoken Language Processing.

Pew Research Center. (2016). *On views of race and inequality. Blacks and Whites are worlds apart*. Author.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, *4*(7), 736–745. https://doi.org/10.1038/s41562-020-0858-1

Podesva, R. J. (2011). Salience and the social meaning of declarative contours: Three case studies of gay professionals. *Journal of English Linguistics*, *39*(3), 233–264. https://doi.org/10.1177/0075424211405161

Prabhakaran, V., Griffiths, C., Su, H., Verma, P., Morgan, N., Eberhardt, J. L., & Jurafsky, D. (2018). Detecting Institutional Dialog Acts in Police Traffic Stops. *Transactions of the Association for Computational Linguistics*, *6*, 467–481. https://doi.org/10.1162/tacl_a_00031

Reisig, M. D., Mays, R. D., & Telep, C. W. (2018). The effects of procedural injustice during police–citizen encounters: A factorial vignette study. *Journal of Experimental Criminology*, *14*(1), 49–58. https://doi.org/10.1007/s11292-017-9307-1

Rogers, P. L., Scherer, K. R., & Rosenthal, R. (1971). Content filtering human speech: A simple electronic system. *Behavior Research Methods and Instrumentation*, *3*(1), 16–18. https://doi.org/10.3758/BF03208115

Sahin, N., Braga, A. A., Apel, R., & Brunson, R. K. (2017). The impact of procedurally-just policing on citizen perceptions of police during traffic stops: The Adana randomized controlled trial. *Journal of Quantitative Criminology*, *33*(4), 701–726. https://doi.org/10.1007/s10940-016-9308-7

Scherer, K. R., Johnstone, T., & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433–456). Oxford University Press.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, *7*(4), 422–445.

Sidanius, J., & Pratto, F. (2001). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

Simpson, A. W., & Erickson, M. T. (1983). Teachers' verbal and nonverbal communication patterns as a function of teacher race, student gender, and student race. *American Educational Research Journal*, *20*(2), 183–198. https://doi.org/10.3102/00028312020002183

Sunshine, J., & Tyler, T. (2003a). Moral solidarity, identification with the community, and the importance of procedural justice: The police as prototypical representatives of a group's moral values. *Social Psychology Quarterly*, *66*(2), 153–165. https://doi.org/10.2307/1519845

Sunshine, J., & Tyler, T. R. (2003b). The role of procedural justice and legitimacy in shaping public support for policing. *Law & Society Review*, *37*(3), 513–548. https://doi.org/10.1111/1540-5893.3703002

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). *Mediation: R package for causal mediation analysis*. https://doi.org/10.18637/jss.v059.i05

Turner, B. L., Caruso, E. M., Dilich, M. A., & Roese, N. J. (2019). Body camera footage leads to lower judgments of intent than dash camera footage. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(4), 1201–1206. https://doi.org/10.1073/pnas.1805928116

Tyler, T. R. (1990). *Why people obey the law: Procedural justice, legitimacy, and compliance*. Yale University Press.

Tyler, T. R. (2005). Policing in black and white: Ethnic group differences in trust and confidence in the police. *Police Quarterly*, *8*(3), 322–342. https://doi.org/10.1177/1098611104271105

Tyler, T. R., & Blader, S. L. (2003). The group engagement model: Procedural justice, social identity, and cooperative behavior. *Personality and Social Psychology Review*, *7*(4), 349–361. https://doi.org/10.1207/S15327957PSPR0704_07

Tyler, T. R., & Huo, Y. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. Russell Sage Foundation Foundation.

Tyler, T. R., & Lind, E. A. (1992). A relational model of authority in groups. *Advances in Experimental Social Psychology*, *25*, 115–191.

Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., Jurgens, D., Jurafsky, D., & Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(25), 6521–6526. https://doi.org/10.1073/pnas.1702413114

Weisbuch, M., & Ambady, N. (2008). Affective divergence: Automatic responses to others' emotions depend on group membership. *Journal of Personality and Social Psychology*, *95*(5), 1063–1079. https://doi.org/10.1037/a0011993

Weisbuch, M., Lamer, S. A., Treinen, E., & Pauker, K. (2017). Cultural snapshots: Theory and method. *Social and Personality Psychology Compass*, *11*(9), e12334. https://doi.org/10.1111/spc3.12334

Weisbuch, M., Pauker, K., & Ambady, N. (2009). The subtle transmission of race bias via televised nonverbal behavior. *Science*, *326*(5960), 1711–1714. https://doi.org/10.1126/science.1178358

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Wilson, J. P., Hugenberg, K., & Rule, N. O. (2017). Racial bias in judgments of physical size and formidability: From size to threat. *Journal of Personality and Social Psychology*, *113*(1), 59–80. https://doi.org/10.1037/pspi0000092

Wood, G., Tyler, T. R., & Papachristos, A. V. (2020). Procedural justice training reduces police use of force and complaints against officers. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(18), 9815–9821. https://doi.org/10.1073/pnas.1920671117

Yokum, D., Ravishankar, A., & Coppock, A. (2019). A randomized control trial evaluating the effects of police body-worn cameras. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(21), 10329–10332. https://doi.org/10.1073/pnas.1814773116