PNAS Nexus, 2024, 3, pgae359
https://doi.org/10.1093/pnasnexus/pgae359
Advance access publication 17 September 2024
Research Report

Leveraging body-worn camera footage to assess the effects of training on officer communication during traffic stops

Nicholas P. Camp [a,*,1], Rob Voigt [b,c,1], MarYam G. Hamedani [d], Dan Jurafsky [e,f] and Jennifer L. Eberhardt [d,g,h,*,1]

[a]Department of Organizational Studies, The University of Michigan, Ann Arbor, MI 48109, USA
[b]Department of Linguistics, Northwestern University, Evanston, IL 60208, USA
[c]Department of Computer Science, Northwestern University, Evanston, IL 60208, USA
[d]Stanford SPARQ, Stanford University, Stanford, CA 94305, USA
[e]Department of Linguistics, Stanford University, Stanford, CA 94305, USA
[f]Department of Computer Science, Stanford University, Stanford, CA 94305, USA
[g]Stanford Graduate School of Business, Stanford University, Stanford, CA 94305, USA
[h]Department of Psychology, Stanford University, Stanford, CA 94305, USA
*To whom correspondence should be addressed: Email: npcamp@umich.edu, jleberhardt@stanford.edu
[1]Connotes joint authorship.
Edited By Michael Muthukrishna

Abstract

Can training police officers on how to best interact with the public actually improve their interactions with community members? This has been a challenging question to answer. Interpersonal aspects of policing are consequential but largely invisible in administrative records commonly used for evaluation. In this study, we offer a solution: body-worn camera footage captures police–community interactions and how they might change as a function of training. Using this *footage-as-data* approach, we consider changes in officers' communication following procedural justice training in Oakland, CA, USA, one module of which sought to increase officer-communicated respect during traffic stops. We applied natural language processing tools and expert annotations of traffic stop recordings to detect whether officers enacted the five behaviors recommended in this module. Compared with recordings of stops that occurred prior to the training, we find that officers employed more of these techniques in posttraining stops; officers were more likely to express concern for drivers' safety, offer reassurance, and provide explicit reasons for the stop. These methods demonstrate the promise of a footage-as-data approach to capture and affect change in police–community interactions.

Keywords: police interventions, body camera footage, procedural justice, natural language processing, computational linguistics

Significance Statement

Police chiefs, city officials, and community leaders alike have highlighted the need to establish respect and restore trust in police–community interactions. Police departments across the nation have implemented trainings to change the tenor of officer communication. How can we tell whether these efforts are successful? We use body camera footage to observe how officers communicate with the public, and how these interactions might change in response to officer training. Using body camera footage as data, we find that officers communicated more respectfully with drivers after being trained in procedural justice. Our findings illustrate how trainings might improve police–community conversations, and, more broadly, how body camera footage can be used to affect and measure change.

Introduction

Video recordings have long played a role in bringing the hidden realities of policing to public light. From camcorder footage of the beating of Rodney King in 1991 to the cell phone videos and body-worn camera recordings that captured the killings of Oscar Grant, George Floyd, Tyre Nichols, and Sonya Massey decades later, footage of police violence has sparked mass protests and calls to reform, reimagine, or abolish the police. Critical incidents such as these are the tip of the iceberg in the long and complex history of police–community relations, the vast majority of which take shape out of the public eye. Nearly 18.7 million people are stopped each year by law enforcement (1). While few of these encounters garner public attention, these interactions also play an important role in community members' perceptions of the police. Those

Competing Interest: J.L.E. was invited by a federal judge and monitor to serve as a Subject Matter Expert for the Oakland Police Department's reform efforts. The assignment began prior to the studies reported here.
Received: October 22, 2023. **Accepted:** May 27, 2024

perceptions reveal a clear racial divide: less than half of all Americans, and just over 1 in 10 African Americans, express confidence in law enforcement to treat all community members equally (2).

Today, the bulk of police–community interactions remain out of sight, not necessarily because they are unrecorded, but because the footage goes unobserved. With the rapid adoption of body-worn cameras, contacts between police officers and members of the public are captured at an unprecedented scale. 47% of US law enforcement agencies and 80% of large departments had adopted body cameras as of 2016 (3), with many more acquiring them in the years since. Despite—or because of—the sheer number of video recordings agencies produce each day, only a small subset of these videos are reviewed for accountability or evidentiary purposes.

Yet body-worn camera footage has the potential to serve another end: a source of data. The proliferation of body cameras means that police interactions can be observed at the scale of broad agency-wide patterns rather than as single case studies of isolated incidents. Moreover, the constant stream of footage agencies accumulate provides a means of comparing police encounters over time. These features make body-worn camera footage a powerful, but heretofore underutilized, tool to detect changes in police behavior in response to supervision, policy intervention, or training.

Police trainings cover a variety of topics, from the tactical use of force, to the proper way to conduct traffic stops, to the importance of procedural justice—that is, treating community members with fairness, respect, and transparency (4, 5). While training programs seek to affect a wide range of actions, they are commonly assessed with administrative records, which provide only a coarse measure of police behavior. Traffic stops that appear identical in administrative data, for example, can vary widely in the extent to which officers interacted in a trustworthy or disrespectful manner with the public, factors that are known to influence community members' institutional trust (6, 7). As a result, behavioral interventions lack proximate measures of whether they affect change in interpersonal aspects of police–community relations.

Body cameras provide a clear window onto police–community interactions as they transpire, filling the gaps in administrative records. A recording can reveal whether an officer gave voice to a driver (8), spoke to them respectfully (9), or even used a respectful tone of voice (10). Until now, these aspects of police interactions would only be visible in police data if a trained observer accompanied officers on their shifts (11–13): an expensive and logistically challenging endeavor. The presence of a physical observer can also itself lead officers to adjust their behavior, posing a challenge for assessing the effects of any intervention (14). Body camera recordings are less costly to obtain, less obtrusive, and are more scalable than ride-along observations. Moreover, these recordings can be revisited, reviewed, and analyzed for as long as they are stored.

As departments accumulate footage over time, it is also possible to observe interactions that preceded an exogenous shock, intervention, or training. How can the possibilities of body camera footage be harnessed to assess efforts to shift police–community relations? Here, we employ a *footage-as-data approach* to examine changes in officer language following a training conducted by the Oakland Police Department (OPD) aimed at enacting principles of procedural justice in routine interactions with the public. As part of this training, officers were instructed in empirically backed ways of communicating respect during traffic stops. By applying computational linguistic techniques to transcripts of officers' body-worn camera recordings, we could identify whether officers used the communication strategies directly targeted by the training.

Language change following OPD's "Procedural Justice in Motion" training

The OPD has operated under a court-ordered negotiated settlement agreement since 2003, the longest federal oversight in the US history. As part of its negotiated settlement, the OPD was required to institute procedural justice training for all sworn personnel, starting in 2014. Oakland is not alone: training on procedural justice and implicit bias have been "core features" of reform agreements between the Department of Justice and law enforcement agencies (15). A growing number of agencies, under federal monitorship or not, have trained officers on procedurally just policing tactics in an effort to improve police–community relations.

There is a considerable gap between the prevalence of such training programs and the evidence for their efficacy, however, much of which relies on officer self-report (16, 17). Officers may demonstrate knowledge of procedural justice or implicit bias (18, 19) on surveys, and may even acknowledge their importance (16, 20) following training, but those survey responses may not translate to actual changes in how officers respond in the field. Where researchers have measured officer behavior in response to intervention, they have largely relied on distal administrative measures, such as search decisions (21), use of force (20), citizen complaints, or behavior in constrained training settings (20). To date, we know of a single published study (13) that observed police–community interactions to assess the impact of procedural justice training on these exchanges. While law enforcement, policymakers, and, above all, communities have an interest in fostering respectful and fair police–community interactions, determining what works is challenging, so long as the behaviors that trainings seek to change are difficult to capture.

Our past work conducted in Oakland has used body-worn camera footage to fill this gap. As part of its monitorship, OPD was required to identify outside researchers who could analyze officer-initiated contacts with community members and determine whether there were significant racial disparities in stop outcomes. Our research team proposed that the department treat body camera footage of traffic stops as data alongside traditional administrative data. We went on to conduct an analysis of body camera transcripts to identify racial gaps in officer-communicated respect. Using a combination of thin-slice judgments and computational linguistic techniques, we identified aspects of speech that corresponded to perceptions of respect and disrespect, and found that OPD officers employed more respectful language with White vs. Black drivers in traffic stops (9).

In response to both community and officer reactions following the release of this work, OPD leadership sought to address these disparities during a planned refresher course on procedural justice (the department had been trained on the topic several years prior). The agency-developed, officer-led training was administered over an 8-month period and consisted of five modules that focused on applying the principles of procedural justice on the job, in the field, and in the local community. One module of the training centered on officer communication in routine traffic stops, which featured the results of our analysis of body camera transcripts. This module incorporated group discussion about those results, role-play dialogues illustrating more- and less-respectful stops, and a recorded interview featuring a member of our research team and OPD leadership. Our research team assisted the agency in developing the materials for this module.

Table 1. Linguistic features for detecting training recommendations.

Recommendation	Evaluation metric	Examples
Initiate with a greeting	Did officer greet the driver or introduce themselves?	"How's it going? Officer [last name], Oakland Police" "Hello, good afternoon. Officer [last name] with the Oakland Police"
State reason for stop early on	Did officer provide an explicit reason for stop?	"The reason I pulled you over is that left turn back there. Okay? There's a no left turn between 4 and 6 PM" "I stopped you because you- you ran that stop sign"
Offer reassurance	Did officer use reassuring language at least once in the interaction?	"That's okay. Do you happen to have your ID with you?" "It's no problem. I'll give you a copy"
Express concern for safety	Did officer express concern for the driver's safety at least once in the interaction?	"Get home to your family safely. Okay?" "So anyway, drive safe and, uh, be mindful of where you're going"
Use formal rather than informal titles	Did officer address the driver exclusively with formal titles?	"All right, Ms. [last name]. I'm going to go ahead and give you a warning" "That's exactly why we do operations for distracted driving, sir"

Specifically, the module identified five concrete actions officers could take to communicate respectfully with drivers (see Fig. S1): (i) beginning a stop by greeting the driver and introducing themselves; (ii) explicitly stating the legal justification for the stop early in the encounter; (iii) expressing concern for the driver's safety; (iv) reassuring the driver during the interaction; and (v) using formal rather than informal titles when addressing the driver (see Table 1, for examples). These particular features were targeted for two reasons. First, these aspects of communication are based on the theories of procedural justice and empirically derived from community impressions of OPD officer language (9). Explaining the justification for a stop, offering reassurance, or expressing concern are ways in which officers signal transparency, trustworthiness, and esteem for the communities they serve. Second, they also represent discrete speech acts: linguistic behaviors that perform a social action in the world (22, 23). Speech acts are readily accessible to speakers, who can discuss and acknowledge their significance (24, 25). These specific acts were chosen so officers could easily identify, remember, and implement them over the course of their shift.

As a source of data, body-worn camera footage captures the most proximate measure of whether OPD officers put these recommendations into practice: officers' language use during actual encounters. Here, we analyze footage transcripts—using a combination of natural language processing techniques and manual coding—to understand which specific communication techniques targeted by the training occur in officers' language (see Table 1). As officers were unaware that their recordings were being analyzed, let alone for the purposes of assessment, we are able to observe these elements of police communication as they naturally unfold.

Thus, we can answer two critical questions. First, were officers more likely to use recommended communication techniques in posttraining stops? Second, how and for which drivers did encounters change? Given the racial disparities that motivated the development of the training, did the intervention influence officer communication toward Black community members, or did the training only improve already-respectful interactions with White drivers? To find out, we compare the relative prevalence of the speech acts targeted by the respectful communication module in footage of pre- and posttreatment traffic stops.

Results

To test whether officers put these recommendations into practice, we sampled among traffic stops conducted in the period surrounding the intervention. For each stop, we identified the date in which the conducting officer (i.e. the officer who interacted with the driver during the stop) participated in the Procedural Justice in Motion training. We then sampled among pretreatment stops that occurred in a 4-week window prior to the conducting officer's training date ($n = 313$, mean time from training 14.6 days, SD = 7.7 days) and posttreatment stops that occurred up to 4 weeks following the conducting officer's training date ($n = 302$, mean time from training 13.8 days, SD = 8.7 days). To avoid having a small number of officers dominate our sample, we limited sampling to 15 stops pre-/posttreatment from each officer. Our sample includes traffic stops conducted by 122 officers, with an average of 5 stops per officer (SD = 7 stops).

On average, traffic stops were about 10 min long ($M = 10.02$ min, SD = 6.64). Almost half of the stops in our sample were of Black drivers (as recorded by the officer, $n = 290$), followed by Hispanic ($n = 137$), White ($n = 93$), and Asian ($n = 45$) drivers, with 50 drivers marked as "some other race." Stops sampled in the pretraining period did not significantly differ from stops in the posttraining period in duration or outcome (whether the driver received a warning, citation, or arrest), nor in the demographic compositions of drivers (age, race, or gender) or conducting officers (years' experience, number of stops per officer).

Sampled stops were professionally transcribed, diarized, and timestamped at the utterance level. Based on these stop transcripts, we assess whether officers applied the five communication strategies recommended in the training module at some point in the interaction. We detected three of the recommended acts—*express concern for safety*, *provide reassurance*, and *use formal titles*—automatically using pattern-based lexicons of relevant words used in prior research (9). Trained coders manually annotated the beginning of each interaction to extract the additional two recommended acts, *greet the driver* and *state the reason for the stop*.

Effects of training on communication in traffic stops

We estimate the prevalence of these communication techniques using generalized estimating equations that account for variance at the officer level. Specifically, we compare the number of recommended acts occurring between pre- and posttraining stops (ranging from none of the techniques to all five of them), as well as the likelihood of officers employing each specific technique, the estimated marginal probabilities of which are displayed in Fig. 1.

Officers used at least one of these techniques in the vast majority of stops: in over 99% of interactions throughout the study period, including over 98% of pretraining stops. Yet, even still, posttraining we find a statistically significant increase in the number of these communication techniques used, Wald $\chi^2 = 7.40$,

$\beta = 0.27$, 95% CI (0.08–0.48), $P = 0.007$. These findings persist controlling for the race and age of the driver, the outcome of the stop, the patrol area in which the stop was conducted, and whether the race of the officer was the same as the race of the driver, Wald $\chi^2 = 7.62$, $\beta = 0.27$ (0.10–0.44), $P = 0.002$.

This represents an increase from about three of these acts per interaction in the period prior to training to 3.4 acts following it. This difference was largely driven by an increase in the number of interactions in which the officer engaged in all five of the targeted speech acts: posttraining stops were over twice as likely to contain all five $b = 0.81$, odds ratio (OR) = 2.25 (1.14–4.44), Wald $\chi^2 = 5.48$, $P = 0.019$.

Beyond these aggregate results, our approach affords a fine-grained analysis of how officer conversations changed, the results of which are displayed in Fig. 1. The speech acts targeted by the training varied in both their general prevalence and in their frequency following the training. In posttraining stops, officers were more likely to *offer reassurance to the driver* ($b = 0.43$, OR = 1.47 [1.02–2.13], Wald $\chi^2 = 4.15$, $P = 0.042$), *express concern for the driver's safety* ($b = 0.59$, OR = 1.81 [1.13–2.90], Wald $\chi^2 = 6.15$, $P = 0.013$), and, marginally, *to explicitly state the reason for the stop* ($b = 0.47$, OR = 1.60 [0.98–2.60], Wald $\chi^2 = 3.49$, $P = 0.062$). Officers were no more likely, however, to *initiate the stop with a greeting* ($b = 0.15$, OR = 1.16 [0.73–1.84], Wald χ^2=0.38, $P = 0.54$) or to *use formal titles with drivers* ($b = 0.06$, OR = 1.06 [0.69–1.64], Wald χ^2=0.08, $P = 0.78$).

Given the racial disparities that motivated the training, did Procedural Justice in Motion improve officer interactions with Black community members? Or did it merely improve already-respectful interactions with White drivers? Our inferences on this question are constrained by both the number of stops in our sampling frame and the demographic composition of the drivers stopped. Only 15% of stops in the training period were of White drivers, reflecting the distribution of stops from which we sampled. As a result, we are limited in our power to estimate heterogeneity in training effects across demographic groups.

With this in mind, there is no indication that White drivers experienced more positive change overall than Black drivers ($\beta =$ 0.03 [−0.32, 0.37], Wald $\chi^2 = 0.03$, $P = 0.874$). Indeed, the aggregate treatment effects above were obtained in a sample largely consisting of stops of Black drivers, and the benefits were significant among this subgroup of community members ($\beta = 0.28$ [0.04–0.52], Wald $\chi^2 = 5.13$, $P = 0.024$). Our findings suggest broad benefits of the training across demographic groups, but we note one exception to this pattern: officers were no more likely to use formal titles throughout posttraining stops of White drivers ($b = 0.81$, OR = 2.26 [0.64–6.33], Wald $\chi^2 = 2.39$, $P = 0.122$) but were marginally less likely to do so among posttraining stops of Black drivers ($b = -0.47$, OR = 0.63 [0.36–1.08], Wald $\chi^2 = 2.86$, $P = 0.091$; interaction $b = 1.28$, Wald $\chi^2 = 5.22$, $P = 0.022$), a point we revisit in the general discussion.

Taken together, these findings identify concrete changes in police–community interactions following procedural justice training. Using body-worn camera footage to analyze officers' language, we can specify aspects of these conversations that did and did not differ. We note that, while modest, these findings were not transitory: encounters that occurred at the end of the 4-week window were just as likely to show improvement as stops that occurred the same day as the training.

While these findings speak to changes in traffic stop interactions, the dispersion of stops across officers, coupled with our sampling limits, constrains our ability to infer change for the average OPD officer. Of the 122 officers in the evaluation sample, 45 only conducted a single stop, and 68 were only observed in the pre- or post-training window. As such, a within-officer fixed effects model does not demonstrate sufficient evidence of change $b = 0.14$ (−0.03, 0.31), $t = 1.56$, $P = 0.118$.

Discussion

Video recordings of the police have the power to galvanize protest, shift public opinion, and affect policy change. Law enforcement has adopted body-worn cameras, largely in accordance with the theory of change that they will hold officers accountable and serve as evidence to judge the appropriateness of their behavior. This footage can also be powerful in another respect. It can be used as data to diagnose law enforcement practices that erode community trust, inform efforts to address those practices, and help gauge the effectiveness of those efforts.

The current work demonstrates how body camera footage can be used to observe and assess changes in policing in response to intervention, including officer behaviors that are difficult to measure by other means. Indeed, a question facing efforts to improve policing and address long-standing racial disparities is how to measure their efficacy. Applying computational linguistic techniques to body camera footage, we could capture nuances of officer communication that can matter for community respect and trust, but are inaccessible in administrative records.

The Procedural Justice in Motion training was particularly well-suited to this investigation, as it drew on analysis of footage from the very department it sought to change. Moreover, it addressed concrete speech acts in officers' communication, rather than their upstream attitudes or beliefs. We could thus determine whether officers employed the specific techniques targeted in the traffic stop module. Our data suggest that they did so: in stops following the training, officers were more likely to explicitly provide drivers with the reason for the stop, and expressed greater reassurance and concern toward the driver during the encounter. These improvements are meaningful in that they occurred in a dataset composed largely of interactions with Black drivers and in an agency with a particularly troubled history of police–community conflict. However, we also acknowledge that these inferences are limited by the observational nature of the data and the distribution of stops across officers; more work is needed to establish the causal effects of procedural justice training for individual officers.

The particular outcomes we measured were tailored to the recommendations of the training, which were themselves informed by past studies of officer respect in traffic stop recordings (9). However, there are any number of feature evaluators may be interested in observing, some of which may arise late in the assessment process. A further benefit of these recordings is that they can be reanalyzed so long as they are stored. In the context of the Procedural Justice in Motion training, for instance, one might ask whether OPD officers enacted other aspects of procedural justice in their traffic stops beyond those specified in the language module, such as giving citizens a chance to ask questions (as "giving voice" to community members is a key tenet of procedural justice). As an exploratory analysis, we revisited stop transcripts to automatically label whether the officer elicited questions from the driver they stopped (e.g. "Do you have any questions for me?"). Controlling for the same covariates as in the model above, officers were over two and a half times as likely to do so in posttraining stops, $b = 0.50$, OR = 1.65 (1.05–2.59), Wald $\chi^2 = 4.15$, $P = 0.030$. This suggests that the training influenced police interactions beyond the five recommendations of the language module; more broadly, it illustrates the opportunities body-worn camera recordings provide for revisiting interactions that other methods may lack.

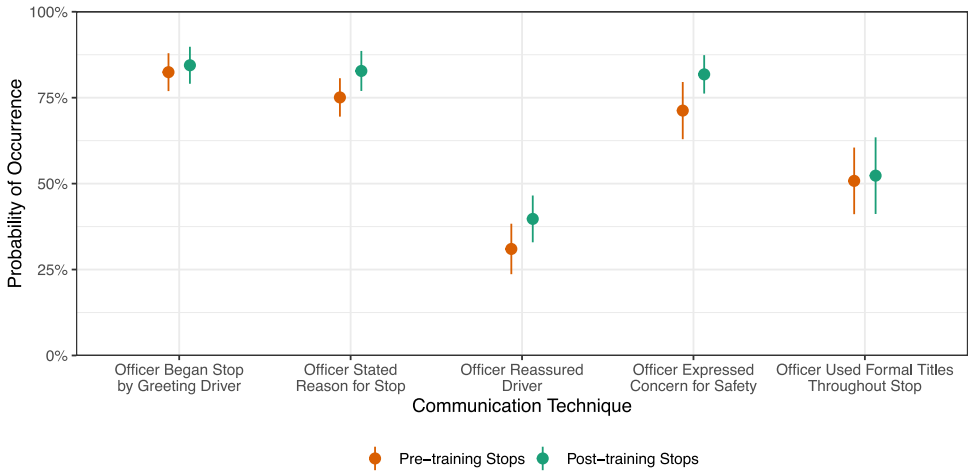

Fig. 1. Comparison of OPD officers' use of recommended communication strategies in stops occurring within 28 days prior to Procedural Justice in Motion training or 28 days following. Bars represent estimated marginal probabilities with 95% CIs.

Chief among these is access to a simple, critical, but elusive aspect of police encounters: how law enforcement communicates with the public. Officers speak for the state; their words carry legal consequences and their language matters for building or eroding institutional trust. Computational analyses of language have already been used to understand disparities and develop actionable insights in other domains as varied as counseling conversations (26), medical patient survey responses (27), and judicial decisions (28). Here, we demonstrate how such an approach can inform policing practices as well. The full richness of body camera footage can be used to examine not only broad constructs such as "respectfulness" in interactions, but also changes in linguistic behaviors that constitute them.

These methods also add nuance by identifying where the training had less impact. For instance, officers began stops by introducing themselves over 80% of the time in pretraining stops, and they were no more likely to do so following the training. The training also may have had unintended effects on officers' use of informal titles with drivers, which appeared to increase in stops of Black drivers. The complexity of the *titles* feature was mirrored in posttraining feedback, where some officers commented that they saw informal titles as a means to build rapport with Black drivers rather than a form of disrespect. While prior research found that community members saw informal language as less respectful (9), the reality may be more complex. Whether "dude" or "bro" are interpreted as rapport or condescension likely depends on context, which could include the identities of the officer and driver in the interaction, the circumstances in which those words are uttered, and whether informality is reciprocated by the driver (a question we consider in the Supplementary Material). This question foregrounds the need for future work to discern the subjective impact of these titles and other linguistic features, a need particularly acute given the role of race in the interpretation of language (29), and language in the construction of race (30). A formidable but worthwhile challenge for linguistic models is to incorporate the nuances of linguistic behaviors in context, beyond their mere presence or absence in speech.

Another open task is testing whether the changes in police–community interactions we observed, which were durable in our sampling period, persist beyond it. Will these changes last? We know that reductions in complaints from the mere introduction of body cameras have been shown to persist over time (31); yet, we also know from existing research on implicit bias interventions, for example, that the impacts of training can rapidly decay (18, 32). For our particular use case, we believe a close examination of decay offers rich opportunities: Does decay happen at the same rate for all of the linguistic gains we report here? Which gains prove to be especially robust? Could training be modified to improve robustness? Are follow-up trainings needed to boost effectiveness? And if so, at what cadence?

Such questions are relevant beyond the specific training we focus on here; body camera footage could be leveraged to inform and evaluate a wide variety of interventions, from policy changes for conducting traffic stops to trainings for community calls for service. As speech recognition and text-to-speech technology advances, we expect that these analyses can be made less expensive and more accessible to practitioners in the future. Future work could apply the natural language processing approach we employed to randomized controlled trials in policing, or tailored to specific units, for example. For police departments and the public to capitalize on these benefits, however, they must address barriers in implementation, logistics, and access to body camera footage. The usefulness of body camera data is contingent on agencies setting consistent policies for camera activation, officers following those procedures, and an internal process for linking footage to administrative records.

Beyond the challenges of accessing recordings, analysts must attend to the limitations of the footage itself. Body-worn camera footage can speak to what transpires over the course of police–community interactions, but it says little about the decisions or events that precede those exchanges. There are also considerations in how language is mapped onto outcomes of interest, such as respect or procedural fairness. Proxies by nature introduce challenges in inference (33), and the extent to which officers' language corresponds with drivers' impressions is an open question. Some studies have observed correspondence between officers' communication and citizens' impressions of police encounters (34) but others have found little evidence of generalization (21). Body camera recordings can contribute to these debates as a naturalistic source of data.

Body camera footage affords stakeholders a chance to better understand moments that matter in policing: actual interactions between officers and the communities they serve. An emerging body of research has used body-worn camera footage as data to uncover the dynamics of police–community encounters in traffic stops (8, 35) disparities therein (9), and their consequences (36). We add to this work, demonstrating how it can be used as data to

assess change. The question of how and when police–community interactions change—or do not change—is more than an academic one. In recent years, through protest, reform, or training, many have sought to change the relationship between law enforcement and the communities they serve. In light of these efforts, body camera footage holds promise as a tool to capture gaps in police interactions, to orient training toward those processes, and to measure progress in improving them.

Materials and methods

Our studies comply with all relevant ethical regulations and were approved by the Stanford University Institutional Review Board. Analyses were conducted on recordings that had been collected by the OPD in the course of officers' routine operations. The research reported in this manuscript was fully funded by the John D. and Catherine T. MacArthur Foundation, at no cost to the City of Oakland. The research team entered a data usage and research collaboration agreement with the City of Oakland and the OPD. The agreement included terms for research independence, publication rights, and data integrity. It also ensured a secure data pipeline to transfer, store, and analyze data, as well as to anonymize and safeguard confidential data.

Training program

As part of a department-wide initiative, all sworn and nonsworn staff of the OPD participated in a 4 hour procedural justice training, Procedural Justice in Motion. The training was deployed over 41 sessions conducted over an 8-month period (a pilot period from November to December 2017, followed by the full deployment from March to July 2018 on which our evaluation centers). To minimize the impact on department operations, each session included a cross section of personnel across roles (officers, nonsworn staff, and department leadership) and division (i.e. within the same squad). While officers were randomly assigned to training dates, only 35.8% of officers attended the session to which they were originally assigned, in part due to the cancellation and rescheduling of sessions.

Each training session was led by a pair of OPD trainers. Trainers were part of a larger team of officer trainers, ranging in rank and experience level. The five modules of the course consisted of a review of the tenets of procedural justice; how to apply them on the job, in the field, and in the local context; and actions that the agency and officers were taking to build connections with the Oakland community. The training was interactive and scenario based, highlighting local examples and efforts.

One module specifically focused on respectful officer communication when interacting with community members as an element of procedural justice. During this part of the training, participants viewed scenarios of traffic stop interactions differing in respectful communication, watched videos of a deputy chief and a member of our team discussing the research and officers' questions about the findings, and had group discussions. Officers were given five concrete tips for respectful communication several times throughout the module: to introduce themselves, state the reason for the stop early in the stop, reassure the driver, use formal titles, and express concern for the driver's safety (see Fig. S1).

Sampling and matching stops

We sought to transcribe stops that took place up to 4 weeks prior to the conducting officer participating in the training and stops that occurred no more than 4 weeks following the conducting officer's training date. Most officers conducted a small number of traffic stops, while a small number of officers conducted a large number of stops, differences that could be based in an officer's role (e.g. traffic enforcement versus patrol), assignment (e.g. to an area with a high number of speeding drivers), or use of discretion (e.g. leniency for minor offenses). To prevent high-stop officers from dominating our sample, we set a limit of 15 stops per officer in our sampling regime, a constraint which applied to 6 of the 122 officers who conducted traffic stops in this window.

We matched training rosters to stop data via to identify a sampling window for each officer. We then cross-referenced this information with body-worn camera metadata to find recordings associated with each stop, using the stop incident number as a common identifier. If there was no matching incident number, we identified the officer's camera recordings closest in time to the traffic stop data then manually matched videos to stop data records. For stops with multiple recordings (i.e. stops with more than one officer), we manually identified the training status of the officer who was the primary contact with the driver.

Stops were excluded from transcription for the following reasons: the recording began after the officer's initial contact with the driver, the driver and/or officer spoke in Spanish, or the stop was not conducted for a traffic violation. For officers with >15 stops, we sampled additional stops at random until we ran out of recordings or reached the limit of 15 stops per officer. The remaining recordings were professionally transcribed, diarized, and timestamped at the utterance level.

Feature extraction and annotation of stop transcripts

For each stop transcript, we coded the presence or absence of the five training recommendations using a combination of computational annotation and hand coding of body-worn camera transcripts. For *Greet the driver*, *express concern for safety*, and *offer reassurance*, we identified whether the act occurred at any point in the stop; for *state the reason for the stop*, we coded whether the officer explicitly stated the reason for stopping driver; and for *use formal titles*, we detected whether the officer exclusively used formal titles to address the driver.

We extracted three of these acts—*express concern for safety*, *offer reassurance*, and *use formal titles*—using pattern-based approaches based on lexicons of relevant words used in prior research (see Table S1). The two remaining acts—*greet the driver* and *state the reason for the stop*—were extracted manually, given the substantial linguistic diversity in how greetings could be expressed and how the reason for a stop could enter conversation. For example, an officer could directly state the reason for a stop (e.g. "I'm stopping you because you were speeding"), or obliquely refer to an offense (e.g. "Do you know how fast you were driving?" or "What's the hurry?") using a known-answer question to solicit a driver account (37). We therefore hand coded all interactions for the presence of explicit reasons, implicit reasons, and the rare cases in which community members preemptively admit fault.

Trained graduate student coders (blind to whether interaction was a pre- or posttraining stop) read transcripts up until the point at which the reason for the stop entered the common ground; that is, when both parties shared mutual knowledge of why the stop has occurred (38). Coders annotated any turns in which *greetings* occurred before the reason is given, and whether the *reasons* were explicitly stated by the officer, provided implicitly, or spontaneously mentioned by the driver (e.g. "Sorry, I think I missed the stop sign there.").

To confirm the robustness of our computational and manual coding, we performed an additional round of hand checks for all features using a different set of research assistants. The computational features received F1 scores of 0.99 for use formal titles, 0.89 for express concern for safety, and 0.75 for offer reassurance, the latter largely due to annotator interpretive flexibility in annotating reassurance, identifying cases like "I hear you" and "you just can't do that, okay?" as reassurance that our model did not. The human-annotated features obtained Cohen's Kappa agreement of 0.98 for greet the driver and 0.87 for state the reason for the stop.

Computational and hand-coded measurements of the five recommendations were then aggregated at the interaction level. For *Greet the driver*, *express concern for safety*, and *offer reassurance*, we coded whether the act occurred at any point in the stop. For *state the reason for the stop*, we coded whether the reason entered the common ground explicitly from the officer or not. For *use formal titles*, we counted how many formal and informal titles occurred in the stop and coded whether the officer exclusively used formal titles to address the driver.

Acknowledgments

The authors thank the OPD and past Chief LeRonne Armstrong. 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. The authors thank the researchers at Stanford SPARQ for their feedback, and Rebecca C. Hetey for contributions to the project. They also thank Lee Hickman and Prisha Srivastava for their assistance with transcript annotation for model validation.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

This research was supported by grants from the John D. and Catherine T. MacArthur Foundation (|G-1512-150464 and G-1805-153038).

Author Contributions

N.P.C. and R.V. contributed equally to this work, analyzed the data, and drafted the paper. N.P.C., R.V., M.G.H., J.L.E., and D.J. designed the research. R.V. performed the text processing and analysis. N.P.C. and M.G.H. were largely responsible for the project administration. M.G.H. and J.L.E. were centrally involved in the development of study materials. N.P.C., M.G.H., and J.L.E. secured camera footage. M.G.H., J.L.E., and D.J. were involved in the writing process and provided critical revisions. J.L.E., D.J., and M.G.H. secured the financial support for the project leading to this publication.

Data Availability

Data and code for this manuscript are accessible on the Open Science Framework at https://osf.io/9bqky/. Our data usage agreement with the OPD prohibits us from releasing raw or edited transcripts of body camera footage. However, we make available interaction-level feature labels and stop metadata needed to reproduce analyses in this manuscript.

References

1 Davis E, Harrell E. 2020. *Contacts between police and the public, 2018*. Bureau of Justice Statistics.
2 Pew Research Center. 2016. *On views of race and inequality, Blacks and Whites are worlds apart*. Pew Research Center.
3 Hyland SS. 2018. *Body-worn cameras in law enforcement agencies, 2016*. Bureau of Justice Statistics.
4 President's Task Force on 21st Century Policing. 2015. *Final report of the president's Task force on 21st century policing*. US Department of Justice, Office of Community Oriented Policing Services.
5 Tyler TR. 1990. *Why people obey the law: procedural justice, legitimacy, and compliance*. New Haven (CT): Yale University Press.
6 Sunshine J, Tyler TR. 2003. The role of procedural justice and legitimacy in shaping public support for policing. *Law Soc Rev*. 37: 513–548.
7 Tyler TR, Huo YJ. 2002. *Trust in the law: encouraging public cooperation with the police and courts*. Russell Sage Foundation.
8 Prabhakaran V, *et al.* 2018. Detecting institutional dialog acts in police traffic stops. *Trans Assoc Comput Linguist*. 6:467–481.
9 Voigt R, *et al.* 2017. Language from police body camera footage shows racial disparities in officer respect. *Proc Natl Acad Sci U S A*. 114:6521–6526.
10 Camp NP, Voigt R, Jurafsky D, Eberhardt JL. 2021. The thin blue waveform: racial disparities in officer prosody undermine institutional trust in the police. *J Pers Soc Psychol*. 121:1157–1171.
11 Antrobus E, Thompson I, Ariel B. 2019. Procedural justice training for police recruits: results of a randomized controlled trial. *J Exp Criminol*. 15:29–53.
12 McCluskey JD, Reisig M. 2017. Explaining procedural justice during police-suspect encounters: a systematic social observation study. *Polic Int J*. 40:574–586.
13 Weisburd D, *et al.* 2022. Reforming the police through procedural justice training: a multicity randomized trial at crime hot spots. *Proc Natl Acad Sci U S A*. 119:e2118780119.
14 Spano R. 2007. How does reactivity affect police behavior? Describing and quantifying the impact of reactivity as behavioral change in a large-scale observational study of police. *J Crim Justice*. 35:453–465.
15 US Department of Justice. 2017. *The civil rights division's pattern and practice police reform work: 1994-present*. Civil Rights Division, U.S. Department of Justice.
16 Skogan WG, Van Craen M, Hennessy C. 2015. Training police for procedural justice. *J Exp Criminol*. 11:319–334.
17 Schaefer B, Thomas H. 2016. *Honing interpersonal necessary tactics (H.I.N.T.): an evaluation of procedural justice training*. Southern Police Institute.
18 Lai CK, Lisnek JA. 2023. The impact of implicit-bias-oriented diversity training on police officers' beliefs, motivations, and actions. *Psychol Sci*. 34:424–434.
19 Worden RE, *et al.* 2024. Implicit bias training for police: evaluating impacts on enforcement disparities. *Law Hum Behav*.
20 Miller J, Quinton P, Alexandrou B, Packham D. 2020. Can police training reduce ethnic/racial disparities in stop and search? Evidence from a multisite UK trial. *Criminol Public Policy*. 19: 1259–1287.
21 Worden RE, McLean SJ. 2017. *Mirage of police reform: procedural justice and police legitimacy*. University of California Press.
22 Austin JL. 1975. *How to do things with words*. Oxford University Press.
23 Searle JR. 1969. *Speech acts: an essay in the philosophy of language*. Cambridge University Press.

24 Culpeper J, Haugh M, Johnson DE. 2017. (Im)politeness: metalinguistic labels and concepts in English. In: Giora R, Haugh M, editors. *Doing pragmatics interculturally. cognitive, philosophical and sociopragmatic perspectives*. Berlin, DE: de Gruyter. p. 135–147.
25 Jaworski A. 2004. Metalanguage: social and ideological perspectives; 3rd Cardiff round table in sociolinguistics, University of Wales Conference Centre, Gregynog Hall, Newtown, Hall, Mid Wales, June 1–3, 1998. de Gruyter.
26 Althoff T, Clark K, Leskovec J. 2016. Large-scale analysis of counseling conversations: an application of natural language processing to mental health. *Trans Assoc Comput Linguist*. 4:463–476.
27 Cammel SA, *et al.* 2020. How to automatically turn patient experience free-text responses into actionable insights: a natural language programming (NLP) approach. *BMC Med Inform Decis Mak*. 20:97.
28 Hausladen CI, Schubert MH, Ash E. 2020. Text classification of ideological direction in judicial opinions. *Int Rev Law Econ*. 62: 105903.
29 Rickford JR, King S. 2016. Language and linguistics on trial: hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language*. 92:948–988.
30 Alim HS, Rickford JR, Ball AF. 2016. *Raciolinguistics: how language shapes our ideas about race*. Oxford University Press.
31 Sutherland A, Ariel B, Farrar W, De Anda R. 2017. Post-experimental follow-ups—fade-out versus persistence effects: the Rialto police body-worn camera experiment four years on. *J Crim Justice*. 53:110–116.
32 Lai CK, *et al.* 2016. Reducing implicit racial preferences: II. Intervention effectiveness across time. *J Exp Psychol Gen*. 145: 1001–1016.
33 Knox D, Lucas C, Cho WKT. 2022. Testing causal theories with learned proxies. *Annu Rev Polit Sci*. 25:419–441.
34 Dobson K, Dittmann A, Yeager D. under review. A transparency statement improves community-police interactions. *PsyArXiv*. https://doi.org/10.21203/rs.3.rs-2239643/v1.
35 Rho EH, *et al.* 2023. Escalated police stops of Black men are linguistically and psychologically distinct in their earliest moments. *Proc Natl Acad Sci U S A*. 120:e2216162120.
36 Camp NP, Voigt R, Jurafsky D, Eberhardt JL. 2021. The thin blue waveform: racial disparities in officer prosody undermine institutional trust in the police. *J Pers Soc Psychol*. 121:1157.
37 Raymond CW, Stivers T. 2016. Off-record account solicitations. In: Robinson JD, editor. *Accountability in social interactions*. Oxford, UK: Oxford University Press. p. 321–353.
38 Clark HH. 1996. *Using language*. Cambridge, UK: Cambridge University Press.