



A field experiment on community policing and police legitimacy

Kyle Peyton^{a,b,1}, Michael Sierra-Arévalo^c, and David G. Rand^{d,e}

^aDepartment of Political Science, Yale University, New Haven, CT 06511; ^bInstitution for Social and Policy Studies, Yale University, New Haven, CT 06511;

^cSchool of Criminal Justice, Rutgers University, Newark, NJ 07102; ^dSloan School, Massachusetts Institute of Technology, Cambridge, MA 02142;

and ^eDepartment of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02142

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved August 12, 2019 (received for review June 12, 2019)

Despite decades of declining crime rates, longstanding tensions between police and the public continue to frustrate the formation of cooperative relationships necessary for the function of the police and the provision of public safety. In response, policy makers continue to promote community-oriented policing (COP) and its emphasis on positive, nonenforcement contact with the public as an effective strategy for enhancing public trust and police legitimacy. Prior research designs, however, have not leveraged the random assignment of police–public contact to identify the causal effect of such interactions on individual-level attitudes toward the police. Therefore, the question remains: Do positive, nonenforcement interactions with uniformed patrol officers actually cause meaningful improvements in attitudes toward the police? Here, we report on a randomized field experiment conducted in New Haven, CT, that sheds light on this question and identifies the individual-level consequences of positive, nonenforcement contact between police and the public. Findings indicate that a single instance of positive contact with a uniformed police officer can substantially improve public attitudes toward police, including legitimacy and willingness to cooperate. These effects persisted for up to 21 d and were not limited to individuals inclined to trust and cooperate with the police prior to the intervention. This study demonstrates that positive nonenforcement contact can improve public attitudes toward police and suggests that police departments would benefit from an increased focus on strategies that promote positive police–public interactions.

community policing | field experiment | legitimacy | intergroup contact

Following nationwide political unrest after Michael Brown was shot and killed by police in Ferguson, MO in 2014, President Barack Obama's Task Force on 21st Century Policing identified building trust and legitimacy as a foundational goal of effective and just policing (1). Legitimacy—the belief that an individual, group, or institution has the authority to dictate an individual's behavior and demand their cooperation—is vital to the effective function of police as a social institution (2, 3). When police lack legitimacy, residents are less likely to contact police or cooperate with their investigations (4, 5). Worse, police–public interactions charged by distrust are more likely to escalate into contests for dominance and status that can lead to the injury or death of police and the public alike (6). Such interactions fuel a cycle of mutual antipathy that further erodes police–public relations and frustrates public safety (5, 7).

In light of the contemporary crisis of police legitimacy and recognition of the damage caused by aggressive policing and mass incarceration (8), community-oriented policing (COP) has re-emerged as a potential policy tool for improving police–community relations. In contrast to policing that emphasizes punitive enforcement (9), COP encourages the formation of cooperative relationships through a variety of nonenforcement interactions, such as community meetings and neighborhood watch programs (1, 10). Despite decades of scholarship suggesting the necessity of cooperative police–community relations for the successful function of police organizations (11), prior research designs have not leveraged the random assignment of

police–public contact to identify the causal effect of such interactions on individual-level attitudes toward the police (12–14). Therefore, the question remains: Do positive, nonenforcement interactions with uniformed patrol officers actually cause meaningful improvements in attitudes toward the police?

This question has direct implications for police policy and public health in the United States, particularly in minority communities where police–community relations are often characterized by longstanding conflict and distrust (4, 15). Interpersonal contact between the public and government officials is a fundamental part of democratic political socialization, with negative experiences undermining trust and political efficacy (16), particularly in the domain of criminal justice (17–20). Police officers come into frequent contact with the public and exercise wide discretion in the implementation of criminal justice policies (21). This discretion covers not only what kinds of offenses they choose to formally sanction through citation or arrest (22) but also, how they choose to interact with the public. The limited studies focused on COP and nonenforcement interactions suggest that, under certain circumstances, police contact may improve police–public relations (3, 14, 20). This is further supported by decades of research on the social psychology of intergroup relations more broadly: positive interpersonal contact can have a powerful effect on attitudes (23–25), even in the case of a single brief interaction (26). However, positive interpersonal contact does not necessarily lead to attitude change in contexts where intergroup relations are charged by a history of

Significance

Repeated instances of police violence against unarmed civilians have drawn worldwide attention to the contemporary crisis of police legitimacy. Community-oriented policing (COP), which encourages positive, nonenforcement contact between police officers and the public, has been widely promoted as a policy intervention for building public trust and enhancing police legitimacy. To date, however, there is little evidence that COP actually leads to changes in attitudes toward the police. We conducted a randomized trial with a large urban police department. We found that positive contact with police—delivered via brief door-to-door nonenforcement community policing visits—substantially improved residents' attitudes toward police, including legitimacy and willingness to cooperate. These effects remained large in a 21-d follow-up and were largest among nonwhite respondents.

Author contributions: K.P., M.S.-A., and D.G.R. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹To whom correspondence may be addressed. Email: kyle.peyton@yale.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1910157116/-DCSupplemental.

violent conflict (27), and field interventions demonstrating the efficacy of brief instances of positive contact for durable attitude change are still rare (25, 28). Given the distrust engendered by repeated high-profile instances of police violence against unarmed civilians in the United States, skepticism that the palliative effects of a brief positive interaction can be applied to police–public interactions is warranted, particularly for members of minority groups (1, 4, 15). A recent report by the National Academy of Sciences, for example, characterized the few empirical studies in this domain as insufficient, concluding that “research is needed that tests the ability of a single interaction to shape general views about police legitimacy” (29).

Here, we report a field experiment that provides such a test. Our intervention, which focused on the individual-level consequences of positive, nonenforcement contact between police and the public, represents a significant departure from prior studies on police–community interactions. Prior work has drawn largely on aggregate administrative records, cross-sectional surveys, case studies, and metaanalyses of various observational research designs (12–14, 20). Observed differences in community-level outcomes across environments that use COP approaches, however, are not necessarily attributable to individual-level attitude change or the mechanism of nonenforcement contact. Take a widely cited study conducted in Houston, Texas (1983 to 1984), which found that a contact program increased residents’ satisfaction with police service and decreased reported crime victimization and fear (30). In this study, 1 small geographic area was, for a period of ~10 mo, targeted with a variety of initiatives, including positive interpersonal contact, more frequent patrols and traffic stops, and newsletters delivered to residents by mail. Prepost interview data from residents in this targeted area were then compared with responses from a sample of residents in another area. As with other studies using such designs, differences in outcomes between control areas and those using a variety of COP interventions might be explained by differences across the communities in which COP was implemented rather than COP itself (14). The only field experiments related to COP we are aware of have used cluster-randomized designs that administered treatments across census block groups in the United States (31) and villages in post-conflict Liberia (32), preventing assessment of the individual-level effects of police-public contact on attitudes toward police.

Methods

The intervention described here, conversely, involved uniformed patrol officers in New Haven, CT, making unannounced visits to randomly assigned homes across the city’s 10 police districts. Prior to 2017, the police department used dedicated walking beats in each district, and officers would regularly introduce themselves and interact with residents; however, officer shortages have since undermined their capacity to maintain regular walking beats, and there has been no systematic study of the effect of such interpersonal contact on public attitudes. In the implementation described here, officers first knocked on the door of a home, initiated a formal greeting with anyone who answered, and immediately explained that they were making a community policing visit in a nonenforcement capacity (e.g., “Everything is okay. No one is in trouble and everyone is safe.”). They then asked to speak with the resident(s) living at the home and engaged them in a brief 10-min conversation using a series of strategies shown in prior studies to encourage positive intergroup contact (23–25). Officers were trained to anticipate suspicion and immediately disarm anxiety within the first 20 s of contact. During the interaction, officers communicated respect by initiating a formal greeting, emphasizing that the visit was an equal status engagement with the goal of improving their shared community, and encouraging residents to provide feedback about policing and neighborhood issues. Officers ended the interaction by giving residents a personalized business card with their work-issued cell phone number handwritten in a designated area. Additional details about the intervention’s implementation are provided in *SI Appendix, section 1.5*.

We measured the effects of these visits by combining the randomized experiment with parallel survey measurement. Following best practice in the design and implementation of field experiments with survey outcomes (26, 33), we first contacted registered voters ($n = 49,757$) via mail to participate in an

ostensibly unrelated survey that was presented as the first wave of a longitudinal university-sponsored public opinion study of city residents. The survey attempted to conceal the connection between the survey and intervention by including unrelated questions about city government, local politics, and national politics in each survey wave (*SI Appendix, section 1* has additional details). Of those contacted by mail, 2,013 individuals nested in 1,852 households completed the baseline survey and provided their contact details to participate in follow-up surveys. We then randomly assigned 926 households (1,007 individuals) to the treatment (i.e., to receive a COP visit) and 926 households (1,006 individuals) to the control (i.e., to receive no COP visit).

Random assignment took place at the household level so that all individuals living in the same household received the same treatment assignment. Of the 1,007 individuals in the treatment group, 412 were successfully reached at the door and received treatment. Compliance was measured in the field using an iPhone app installed on the department-issued cell phones of all participating officers. After the intervention, all 2,013 individuals who participated in the baseline survey were invited via email to complete 2 follow-up surveys. These additional online surveys, presented as a continuation of the baseline survey, were sent 3 d after the intervention ($n = 1,484$ respondents) and 21 d after the intervention ($n = 1,069$ respondents). We did not have sufficient research funds to conduct more than 2 follow-up surveys. Our study was reviewed by the Human Subjects Committee (HSC) at Yale University, and the HSC determined that it presented “minimal risk to research subjects” (Institutional Review Board Protocol ID 2000023097). *SI Appendix, section 1* provides additional information about recruitment, design, and sample characteristics. *SI Appendix, Fig. S1* shows an overview.

As specified in our preanalysis plan (PAP) (*SI Appendix, Appendix C*), we included both primary (confirmatory) and secondary (exploratory) outcomes in the online panel survey. The primary outcome measures, derived from widely used survey batteries on civilian attitudes toward police (3, 6, 34), tap 4 attitudinal dimensions: legitimacy, perceived effectiveness, cooperation, and compliance. Conceptually, the first 2 measures capture values-based beliefs about the normative appropriateness of the police, whereas the latter 2 measures capture behavioral legitimacy, or the willingness to act in a manner that aids law enforcement (35). Following prior work, the “legitimacy” dimension included measures that tap the interdependent concepts of trust in the police and judgements about the normative alignment of police–public values (3, 6, 34–36). Secondary outcomes included an index of respondents’ judgements about “the police” as a group (e.g., whether police officers are “compassionate” or “cold hearted”), an index of questions about respondents’ confidence in the police (e.g., “The police make me feel safer in my neighborhood”), and support for specific policies (e.g., a funding increase to hire more patrol officers). *SI Appendix, section 2* provides additional details about outcome measurement, and *SI Appendix, Appendix D* presents all survey instruments.

Prior to the intervention, individuals in treatment and control scored similarly on all primary outcome indices (*SI Appendix, Fig. S10*). Background characteristics remained balanced across treatment and control for all subsequent survey waves (*SI Appendix, Tables S7–S10*), and differential attrition was not detected in any wave (*SI Appendix, Tables S11 and S12*). We assess the effect of the intervention by estimating the “average treatment effect on the treated” (ATT) and the “intent-to-treat effect” (ITT). The ATT measures the effect of the treatment on the treated. The estimand is the average causal effect among “compliers,” the subset of individuals who would be treated if assigned to the treatment group. We use the term ATT rather than “complier average causal effect” since none of the units assigned to control were treated (see *SI Appendix, section 4* for additional details). The ITT represents the overall effect of the intervention (comparing all individuals assigned to treatment with all individuals assigned to control regardless of whether they were successfully contacted) and is relevant to the practical implementation of policing strategy, since no intervention that assigns unannounced police contact in a field setting can guarantee that all attempted visits will result in contact. All outcomes are transformed to a scale ranging from 0 to 100 in order to facilitate interpretation and comparisons across measures. Additional details are provided in *SI Appendix, section 4*. The replication data and statistical code for this study have been deposited at the Institution for Social and Policy Studies (ISPS) Data Archive (<https://isps.yale.edu/research/data>).

Results

The intervention had a significant positive effect on overall attitudes toward the police as measured by an index of all primary outcome measures 3 d after the intervention (ITT: $t = 6.94$; $P < 0.001$; ATT: $t = 7.15$; $P < 0.001$) and 21 d after the intervention (ITT: $t = 3.83$; $P < 0.001$; ATT: $t = 3.85$; $P < 0.001$). These differences are substantively large: among those who were treated, attitudes

toward the police increased by about 7 points on a 0 to 100 scale in the first follow-up survey. To provide some context for this effect size, we note that this increase is larger than the gap between white and black respondents observed at baseline (5 points). Similarly, the visits increased generalized positivity toward police by 9.5 points on 0 to 100 point “feeling thermometer” scale. For additional context, this is comparable with the 10-point increase in positivity toward transgender people reported in a recent field experiment that showed that brief door-to-door interactions emphasizing active perspective taking reduced transgender prejudice (26).

The effect of the intervention was evident across all 4 primary outcome measures (Fig. 1). The immediate effect was strongest on perceptions of police performance and legitimacy, and these effects were also evident in the 21-d follow-up. The intervention also increased self-reported willingness to cooperate with police and willingness to comply with police directives. Although these effects were comparatively smaller, the effect on willingness to cooperate is particularly noteworthy given how responses were clustered near the upper bound of this index at baseline (*SI Appendix*, Fig. S10). The intervention was also broadly effective across all preregistered subgroups (Fig. 2). Among black respondents who were successfully contacted, the intervention had an initial effect of ~11 scale points—nearly twice as large as the 6-point effect among white respondents (*SI Appendix*, section 5.1). Furthermore, the visits had the strongest effect among individuals who held the most negative views toward police prior to the intervention as measured in the baseline survey (*SI Appendix*, section

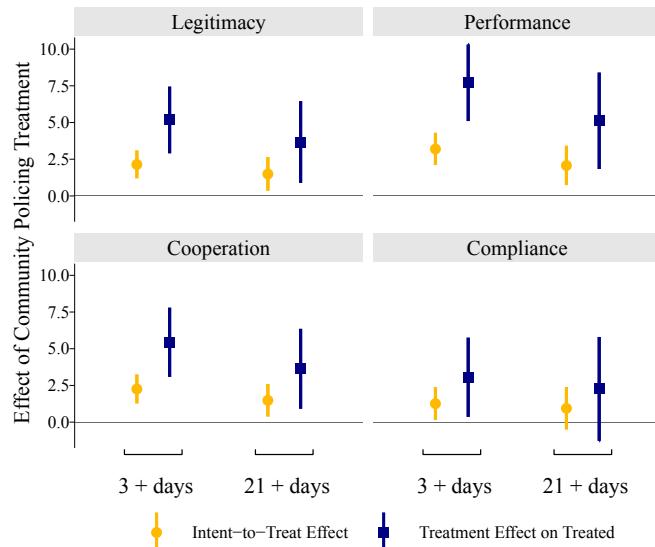


Fig. 1. Effect of community policing treatment on primary outcome measures. Primary outcome measures are indices of multiple questions and scaled to range from 0 to 100. Legitimacy is an index of responses to 8 statements about police legitimacy (e.g., “They make fair and impartial decisions”). Performance is an index of responses to 4 statements about police effectiveness (e.g., “I have confidence that the police in New Haven can do their job well”). Cooperation is an index of responses to 4 questions about willingness to assist police (e.g., “If the police were looking for a suspect who was hiding, and you knew where that person was, how likely would you be to provide the police with information?”). Compliance is an index of responses to 4 questions about willingness to comply with police directives (e.g., “If the police tell you to do something, you should do it”). Covariate-adjusted point estimates and 95% confidence intervals are constructed using the prespecified levels and estimation procedures described in the PAP (*SI Appendix*, Appendix C). Point estimates and standard errors are presented in tabular form in *SI Appendix*, Table S13. *SI Appendix*, section 2 has all individual questions.

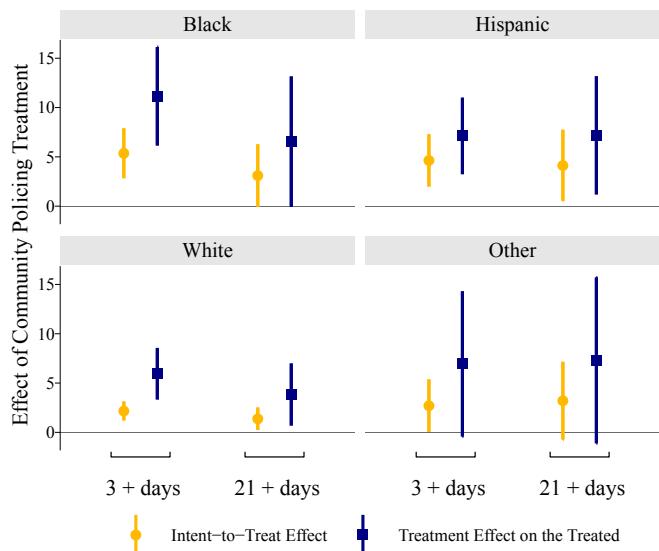


Fig. 2. Effect of community policing treatment on the index of primary outcomes by race/ethnicity. The index of primary outcomes includes all questions underlying the performance, legitimacy, cooperation, and compliance measures and is scaled to range from 0 to 100. Race/ethnicity was measured with the combined race and ethnicity question format used in the 2010 US Census: “Other” includes Asian, Native American, Middle Eastern, multiracial, or another race or ethnicity. Covariate-adjusted point estimates and 95% confidence intervals are estimated separately for each of the 4 major subgroups of subjects in the study. Point estimates and standard errors are presented in tabular form in *SI Appendix*, Table S14. *SI Appendix*, section 2 shows all individual questions.

5.2). Thus, efficacy was not limited to those subgroups inclined to have positive attitudes toward police prior to the intervention.

Finally, the intervention also improved attitudes toward police on our secondary outcome measures (Fig. 3). Of particular note is the substantial reduction in generalized negative attitudes toward police as captured by a reduction in negative beliefs about police officers as a group and an increase in perceived warmth toward police on the feeling thermometer. We also examined the effect on support for 2 specific policy initiatives: the department’s use of body-worn cameras and a 10% funding increase to hire more patrol officers. Body-worn cameras were implemented in 2017, and this initiative was supported by 95% of respondents in the baseline survey. We find evidence that the intervention caused a small reduction in support for body-worn cameras in the 3-d follow-up survey, but these effects were not statistically distinguishable from 0 in the 21-d follow-up, suggesting that the officer visits did not have a durable effect on residents’ support for this popular initiative. During the intervention period, the department had roughly 20% fewer uniformed patrol officers than in 2017, and starting salaries were \$44,400, the lowest level in the state of Connecticut. We also found that the intervention raised support for a policy to increase the number of police on the street by 10%, a measure of public support that prior research shows to be negatively affected by militarized policing practices (37). Relatedly, the visits had a large positive effect on an index of questions about respondents’ confidence in the police (e.g., “The police make me feel safer in my neighborhood”). Conversely, we found little evidence that the visits improved diffuse support for local government institutions beyond the police as measured with a battery of trust in government questions that regularly appear in national public opinion surveys. We note there are long-standing concerns about the validity of these “trust in government” measures (38), and as

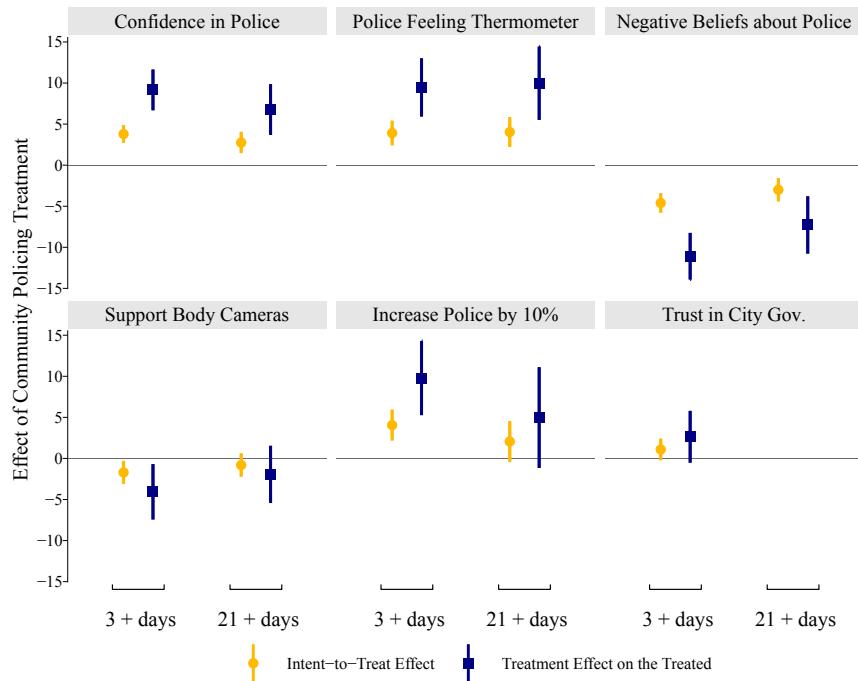


Fig. 3. Effect of community policing treatment on secondary outcome measures. All secondary outcomes are scaled to range from 0 to 100. Confidence in police is an index of responses from “strongly disagree” to “strongly agree” to 6 statements (e.g., “The police make me feel safer in my neighborhood”). Police feeling thermometer is a sliding scale from “cold” (0) to “warm” (100). The negative beliefs about police measure is an index of responses to 5 judgements about the police as a group measured using a sliding scale (e.g., the police as are rated from compassionate to cold hearted). Support for body cameras and increase police by 10% are both single-item measures ranging from “strongly oppose” to “strongly support.” Trust in city government is an index of responses to 4 items (e.g., “How often do you think you can trust government in the City of New Haven to do what is right?”). Covariate-adjusted point estimates and confidence intervals (95% for support body cameras and 92% for all others) are constructed using the prespecified levels and estimation procedures described in the PAP (*SI Appendix, Appendix C*). Point estimates and standard errors are presented in tabular form in *SI Appendix, Table S15*. *SI Appendix, section 2* shows all individual questions.

described in our preanalysis plans, they were not included in the 21-day survey to make room for other questions.

Discussion

The findings reported here have both theoretical and applied importance. A growing body of empirical evidence suggests that brief interactions focused on promoting positive contact can play a powerful role in attitude change. This foundational component of intergroup relations theory, proposed more than 50 y ago (23), has been subjected to only a handful of tests using randomized experiments conducted in the field rather than the laboratory (25). For example, a recent study using a field experiment design similar to the one reported here showed that a single interaction with an activist canvasser could substantially reduce antitransgender prejudice, especially when the activist also identified as transgender (26). The intervention reported here provides evidence in support of the power of positive intergroup contact, extending these insights to interactions between uniformed patrol officers and the individuals they police. The broad effects of positive, non-enforcement police–public interactions reported here are especially noteworthy in light of the well-documented tensions between police and the public, including within minority communities, where one might expect longstanding distrust of police to engender decidedly negative interactions (39). The observation that the largest attitudinal improvements in this field experiment occurred among racial minorities and those who held the most negative views toward police at baseline underscores the power of positive contact in communities most in need of less punitive, more cooperative policing.

At an applied level, prior research on the effectiveness of neighborhood-level community policing has been hampered by the amorphous operationalization of this concept. Indeed, over the

past 30 y, COP has been defined as community meetings, neighborhood watches, community newsletters, and door-to-door visits (10), ultimately restricting claims of effectiveness to the particular bundle of COP activities implemented by a given department. The lack of experimental designs in past evaluations further compounds difficulties in assessing the causal effect of COP on public attitudes toward police. Although police legitimacy is widely acknowledged as a necessary condition for lower crime and increased public safety (40, 41), the lack of rigorous evaluations of well-defined COP strategies has left many law enforcement professionals skeptical of their value. As a result, many police departments in the United States eschew COP and instead invest limited time, money, and personnel in enforcement activities that a growing body of research links to public distrust and damaged police legitimacy (4, 37, 39).

The intervention described here provides an example of how a relatively simple change to police behavior can have a substantial positive effect on measures of both values-based and behavioral legitimacy (35). Of course, it is premature to definitively conclude that such an intervention could be successfully replicated in another jurisdiction, much less the full range of nearly 18,000 state and local law enforcement agencies spread across the United States (42). Even in police departments with an organizational commitment to COP, the reality of shrinking budgets and long-standing issues in the hiring and retention of officers will likely pose significant barriers to the implementation of such programs. While the intervention assessed here improved public attitudes toward police, positive, nonenforcement police contact is no panacea for longstanding issues in policing that include police brutality, corruption, and racial bias (1). In short, it is our view that improved police–public relations is a necessary but not sufficient condition to achieve more just, effective policing and should not be

pursued to the exclusion of other vital reforms. These challenges and cautions notwithstanding, evaluation of similar interventions in other municipalities as well as long-term longitudinal analyses of downstream effects on outcomes like crime rates, crime reporting, and neighborhood violence, are clear avenues for future research. The results reported here provide clear empirical support for the efficacy of policing strategies aimed at improving attitudes toward the police via positive nonenforcement contact between officers and the communities they serve.

Data and Materials Availability. An initial PAP, dated 15 September 2018, was uploaded to the Open Science Framework website on 16 September 2018. This PAP was uploaded after baseline data were collected in the T0 Survey but before data collection was completed in the posttreatment surveys. Two supplements to the initial PAP were filed during the intervention period as described in *SI Appendix*. These are available at <https://osf.io/zhuqm/>. This study was approved by the Human Subjects Committee (HSC) at Yale University (IRB Protocol ID 2000023097). Consent to participate in the 3 wave panel survey was obtained online during the first survey, and the HSC waived written informed consent for the

home visit portion of the study per federal regulation 45 CFR 46.117 (c)(2). All participants were debriefed after the third survey wave. All replication data and code are deposited at the ISPS Data Archive (<https://isps.yale.edu/research/data>).

ACKNOWLEDGMENTS. We thank M. Stagnaro for his contributions to conceptualizing the project and M. Stagnaro and A. Coppock for their contributions to a pilot survey experiment conducted on Amazon Mechanical Turk in 2017 that was used to facilitate outcome selection and survey design for the field study. This project would not have been possible without the support and enthusiasm of many individuals at the New Haven Police Department past and present. We give special thanks to Chief A. Campbell, Chief L. Casanova, Lt. M. Colon, N. Pastore, Chief O. Reyes, Lt. H. Sharp, S. Spell, and Lt. D. Zannelli. We thank J. Dovidio, G. Grossman, J. Hacker, G. Huber, H. Jefferson, R. Johnston, J. Kalla, A. Papachristos, L. Peer, F. Sávje, and T. Tyler for helpful comments and feedback. This research was presented and benefited from feedback at the University of Melbourne, the University of Sydney, the New York University Experimental Political Science Conference, the Harvard Experimental Political Science Conference, the 77th Midwest Political Science Association Annual Meeting, and the Law and Society Association 2019 Annual Meeting. K.P. acknowledges funding from the Field Experiments Initiative at the Institution for Social and Policy Studies and the Justice Collaboratory at Yale Law School.

1. Office of Community Oriented Policing Services, *Community Policing Defined* (Office of Community Oriented Policing Services, 2012).
2. T. R. Tyler, Psychological perspectives on legitimacy and legitimization. *Annu. Rev. Psychol.* **57**, 375–400 (2006).
3. T. R. Tyler, J. Jackson, Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. *Psychol. Public Policy Law* **20**, 78–95 (2014).
4. P. J. Carr, L. Napolitano, J. Keating, We never call the cops and here is why: A qualitative examination of legal cynicism in three Philadelphia neighborhoods. *Criminology* **45**, 445–480 (2007).
5. M. Desmond, A. V. Papachristos, D. S. Kirk, Police violence and citizen crime reporting in the black community. *Am. Sociol. Rev.* **81**, 857–876 (2016).
6. J. Sunshine, T. R. Tyler, The role of procedural justice and legitimacy in shaping public support for policing. *Law Soc. Rev.* **37**, 513–548 (2003).
7. J. A. Shjorback, J. Nix, S. E. Wolfe, The ecological structuring of police officers' perceptions of citizen cooperation. *Crime Delinq.* **64**, 1143–1170 (2017).
8. M. S. Phelps, D. Pager, Inequality and punishment: A turning point for mass incarceration? *Ann. Am. Acad. Pol. Soc. Sci.* **663**, 185–203 (2016).
9. H. Goldstein, Toward community-oriented policing: Potential, basic requirements, and threshold questions. *Crime Delinq.* **33**, 6–30 (1987).
10. C. M. Lum, C. S. Koper, *An Evidence-Assessment of the Recommendations of the President's Task Force on 21st Century Policing* (International Association of Chiefs of Police, 2016).
11. O. W. Wilson, *Police Administration* (McGraw-Hill, 1950).
12. J. Hawdon, J. Ryan, Police-resident interactions and satisfaction with police: An empirical test of community policing assertions. *Crim. Justice Policy Rev.* **14**, 55–74 (2003).
13. P. K. Manning, Community policing. *Am. J. Police* **3**, 205–227 (1984).
14. C. Gill, D. Weisburd, C. W. Telep, Z. Vitter, T. Bennett, Community-oriented policing to reduce crime, disorder and fear and increase satisfaction and legitimacy among citizens: A systematic review. *J. Exp. Criminol.* **10**, 399–428 (2014).
15. R. K. Brunson, "Police don't like black people": African-American young men's accumulated police experience. *Criminol. Publ. Policy* **6**, 71–101 (2007).
16. J. Soss, Lessons of welfare: Policy design, political learning, and political action. *Am. Polit. Sci. Rev.* **93**, 363–380 (1999).
17. V. M. Weaver, A. E. Lerman, Political consequences of the carceral state. *Am. Polit. Sci. Rev.* **104**, 817–833 (2010).
18. A. White, Misdemeanor disenfranchisement? The demobilizing effects of brief jail spells on potential voters. *Am. Polit. Sci. Rev.* **113**, 1–14.
19. A. E. Lerman, V. Weaver, Staying out of sight? Concentrated policing and local political action. *Ann. Am. Acad. Pol. Soc. Sci.* **651**, 202–219 (2014).
20. J. Fagan, T. R. Tyler, Legal socialization of children and adolescents. *Soc. Justice Res.* **18**, 217–241 (2005).
21. J. H. Skolnick, *Justice Without Trial: Law Enforcement in Democratic Society* (Wiley, 1975).
22. J. Goldstein, Police discretion not to invoke the criminal process: Low-visibility decisions in the administration of justice. *Yale Law J.* **69**, 543–594 (1960).
23. G. W. Allport, *The Nature of Prejudice* (Basic Books, 1954).
24. T. F. Pettigrew, L. R. Tropp, A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.* **90**, 751–783 (2006).
25. E. L. Paluck, S. A. Green, D. P. Green, The contact hypothesis re-evaluated. *Behav. Publ. Policy*, 1–30 (2018).
26. D. Broockman, J. Kalla, Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science* **352**, 220–224 (2016).
27. A. Scacco, S. S. Warren, Can social contact reduce prejudice and discrimination? *Am. Polit. Sci. Rev.* **112**, 654–677 (2018).
28. E. L. Paluck, D. P. Green, Prejudice reduction: What works? A review and assessment of research and practice. *Annu. Rev. Psychol.* **60**, 339–367 (2009).
29. National Academies of Sciences Engineering, Medicine, *Proactive Policing: Effects on Crime and Communities* (The National Academies Press, Washington, DC, 2018).
30. M. A. Wycoff, W. G. Skogan, M. Pate Antony, L. W. Sherman, A. Sampson, *Citizen Contact Patrol: The Houston Field Test, Executive Summary* (Police Foundation, Washington, DC, 1985).
31. D. Weisburd, N. A. Morris, J. Ready, Risk-focused policing at places: An experimental evaluation. *Justice Q.* **25**, 163–200 (2008).
32. R. Blair, S. Karim, B. Morse, Establishing the rule of law in weak and war-torn states: Evidence from a field experiment with the Liberian National Police *Am. Polit. Sci. Rev.* **113**, 641–657 (2019).
33. D. Broockman, J. Kalla, J. Sekhon, The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Polit. Anal.* **25**, 435–464 (2017).
34. T. R. Tyler, J. Fagan, A. Geller, Street stops and police legitimacy: Teachable moments in young urban men's legal socialization. *J. Empir. Leg. Stud.* **11**, 751–785 (2014).
35. M. Levi, A. Sacks, T. Tyler, Conceptualizing legitimacy, measuring legitimizing beliefs. *Am. Behav. Sci.* **53**, 354–375 (2009).
36. J. Jackson, J. M. Gau, "Carving up concepts? Differentiating between trust and legitimacy in public attitudes towards legal authority" in *Interdisciplinary Perspectives on Trust*, E. Shockley, T. Neal, L. PytlkZillig, B. Bornstein, Eds. (Springer, 2016), pp. 49–69.
37. J. Mummolo, Militarization fails to enhance police safety or reduce crime but may harm police reputation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9181–9186 (2018).
38. M. Levi, L. Stoker, Political trust and trustworthiness. *Annu. Rev. Polit. Sci.* **3**, 475–507 (2000).
39. J. M. Gau, R. K. Brunson, Procedural justice and order maintenance policing. *Justice Q.* **27**, 255–279 (2010).
40. T. R. Tyler, J. Fagan, Legitimacy and cooperation: Why do people help the police fight crime in their communities? *Ohio State J. Crim. Law* **6**, 231–275 (2008).
41. L. Mazerolle, S. Bennett, J. Davis, E. Sargeant, M. Manning, Procedural justice and police legitimacy: A systematic review of the research evidence. *J. Exp. Criminol.* **9**, 245–274 (2013).
42. B. A. Reaves, *Census of State and Local Law Enforcement Agencies, 2008* (Bureau of Justice Statistics, Office of Justice Programs, US Department of Justice, 2011).

Supporting Information for ‘A Field Experiment on Community Policing and Police Legitimacy’

Contents

1 Recruitment Procedures and Experimental Design	4
1.1 Random Assignment of Households	5
1.2 Reinterview Procedures	6
1.3 Sample Characteristics	7
1.3.1 Attitudes toward police	8
1.4 Pre-Analysis Plans	9
1.5 Intervention Details	10
1.5.1 Officer Training	11
1.5.2 Intervention Procedure	12
1.5.3 Background Characteristics of Participating Officers	12
1.5.4 Implementation and Field Observations	13
1.5.5 Monitoring for Unanticipated Problems	15
1.5.6 Suspicion about Survey and Intervention Connection	15
2 Outcome Measurement and Scaling	17
2.1 Primary outcomes	17
2.1.1 Cooperation	17
2.1.2 Compliance	17
2.1.3 Performance	18
2.1.4 Legitimacy	18
2.1.5 Index of all Primary Outcomes	18
2.2 Secondary Outcomes	19
2.2.1 Confidence in Police	19
2.2.2 Police Feeling Thermometer	19
2.2.3 Police Stereotypes	19
2.2.4 Police Policy Questions	20
2.2.5 Trust in Local Government	20
3 Design Diagnostics	22
3.1 Covariate Balance	22
3.2 Differential Attrition	23

4 Estimation and Inference	24
4.1 Treatment Non-Compliance	24
4.2 Results: Primary Outcomes	26
4.3 Results: Secondary Outcomes	26
5 Treatment Effect Heterogeneity	27
5.1 Race/Ethnicity of Respondents	27
5.2 Quantiles of Baseline Support	27
6 Figures S1 to S7	27
7 Tables S1 to S16	39
A Appendix A: Statistical Code	46
A.1 Functions used for outcome construction	46
A.2 Randomization inference procedures	47
A.2.1 Covariate imbalance	47
A.2.2 Sharp null on response behavior	49
A.2.3 Treatment-covariate interactions on response behavior	49
A.3 Effects on primary/secondary outcomes	50
A.3.1 Primary outcomes	50
A.3.2 Secondary outcomes	52
A.4 Heterogenous treatment effects	55
A.4.1 Race/Ethnicity	55
A.4.2 Quantiles of Baseline Support	57
B Appendix B: Pre-Analysis Plan and Supplements	60
C Appendix C: Survey Instruments	77
C.1 Baseline (T0 Survey)	77
C.2 First Followup (T1 Survey)	109
C.3 Second Followup (T2 Survey)	134

List of Figures

S1	Summary of Experimental Design	29
S2	Field Experimental Manipulation	30
S3	Officer Ratings of Interactions with Residents	31
S4	Bilingual Mailer (English Side)	32
S5	Example of Additional Incentive Payment Email (English Version)	33
S6	Example of Final Survey Expiration Reminder (English Version)	34
S7	Example of Slider Question Format in Online Survey (English Version)	35
S8	Institutional Review Board Letter to Suspicious Participants	36
S9	Post-Study Debrief (English Version)	37
S10	Distribution of Primary Outcome Measures in Baseline Survey	38

List of Tables

S1	Survey Respondents by Experimental Condition and Wave	39
S2	Characteristics of sample at each stage of study	39
S3	Empirical validation of race/ethnicity classification	39
S4	Reported confidence in the police department by level and group	40
S5	Feeling thermometer ratings of police by level and group	40
S6	Uniformed patrol officers by sex and race/ethnicity	41
S7	Background Covariates by Treatment Assignment in T0 Survey	41
S8	Background Covariates by Treatment Assignment in T1 Survey	42
S9	Background Covariates by Treatment Assignment in T2 Survey	43
S10	Randomization Inference (RI) for covariate balance	43
S11	Randomization Inference (RI) for the sharp null hypothesis of no effect on response for any unit	44
S12	Randomization Inference (RI) for differential attrition	44
S13	Estimated treatment effects on primary outcome measures	44
S14	Estimated treatment effects by Race/Ethnicity	45
S15	Estimated treatment effects on secondary outcome measures	45
S16	Estimated treatment effects by Quantile of Baseline Support	45

1 Recruitment Procedures and Experimental Design

We used a mail to online panel design to establish a baseline sample of eligible participants and measure outcomes before and after treatment (26, 33). We started by recruiting subjects in [CITY NAME] to participate in an online panel survey called the “[CITY NAME] Opinion Study 2018”. On August 15th 2018, approximately 50,000 [CITY NAME] residents were sent an invitation to the mailing address listed in their voter registration record (see Fig. S4).

The mailers directed recipients to an online survey via a landing page at [www.\[CITYNAME\]opinion.com](http://www.[CITYNAME]opinion.com). Each respondent was provided with a unique login code in their recruitment letter, which they used to access the survey. The survey was available in English and Spanish. A dedicated local phone number and university email address were created to field respondent inquiries during the initial recruitment period, which ended on September 6th 2018.

2,013 individuals nested in 1,852 households completed the baseline survey (T0 Survey) and provided their contact details to participate in followup surveys. On 9 September 2018 we randomly assigned 926 households (1,007 individuals) to receive treatment and 926 households (1,006 individuals) to receive control. Block random assignment took place at the household level using matched quartets (more details below).

Our partner organization, a large urban police department, subsequently canvassed all 926 households in the treatment group and delivered a short (approximately 10 minute) conversation. Individual officers provided their contact information via department issued business cards. Canvassing officers were assisted by a mobile application that was installed on the department issued cell phones of all participating officers.

Outcomes were measured in two post-treatment survey waves. Treated units (and the control units in their quartet) were invited to participate in the first post-treatment survey (T1 Survey) via email 3 days after treatment. Treated units (and the control units in their quartet) were invited to participate in the second post-treatment survey (T2 Survey) via email 21 days after treatment was delivered.

All surveys were administered using Qualtrics Survey Software. The survey instruments for each wave are provided in full in Appendix C. These documents are survey item exports and should only be used to reference the question wording of items, not their appearance in the online survey environment. In particular, the slider items for the feeling thermometer and stereotype questions do not render properly in PDF format. In the online survey, sliders were presented with the neutral midpoint as the starting position, with groups presented in random order (see Fig. S7). Numeric scores would also appear on screen as survey respondents dragged the bars.

1.1 Random Assignment of Households

Randomization was conducted at the household level with the assistance of the `blockTools` and `randomizr` packages in R (43-44). Matched quartets of households were formed using the following blocking variables derived from a combination of administrative voterfile data and responses to the baseline survey:

- The proportion of residents above the median income category of \$50,000-59,999.
- The average age in the household.
- The proportion of females in the household.
- The proportion of residents in each of 4 major race/ethnicity categories (black, white, Hispanic, or other).
- The household average on the first factor from a PCA on all the items underlying the 4 primary outcome measures: cooperation, legitimacy, compliance, and performance.
- The proportion of residents who indicated any prior arrest by the police department.
- The proportion of residents who indicated having any face to face contact with the police department in the last 12 months.
- The proportion of residents who indicated any unfair treatment by the police department in the past.
- The proportion of spanish speaking residents in the household
- The household size
- The police district

1,852 households (clusters) of 2,013 individuals were randomly assigned to treatment or control across 463 blocks (quartets of households) using complete random assignment. The probabilities of assignment were constant across households. 926 households (1,007 individuals) were assigned treatment and 926 households (1,006 individuals) were assigned control.

1.2 Reinterview Procedures

All respondents that completed the T0 Survey were re-contacted to participate in the T1 Survey, and all respondents that completed the T1 Survey were re-contacted to participate in the T2 Survey. Three days after receiving treatment, treated households and the control households in their matched quartet were invited to participate in the T1 Survey. Subjects were then invited to the T2 Survey 3 weeks after treatment (18 days after the T1 invitation). These procedures are detailed below.

All individuals nested in homes (the unit of assignment) became eligible for a T1 Survey invitation when the following criteria were satisfied:

- A response to the T0 Survey has been recorded, and an invitation to the T1 Survey had not yet been sent.
 - For units assigned treatment:
 - (1a) Three days have passed since treatment was delivered.
 - For units assigned control:
 - (1b) The unit belongs to a quartet where at least 1 unit assigned treatment satisfies condition (1a).

All individuals nested in homes (the unit of assignment) became eligible for a T2 Survey invitation when the following criteria were satisfied:

- A response to the T1 Survey has been recorded, and an invitation to the T2 Survey has not yet been sent.
 - For units assigned treatment:
 - (2a) Three weeks (21 days) have passed since treatment (18 days since the T1 Survey invitation).
 - For units assigned control:
 - (2b) The unit belongs to a quartet where at least 1 unit assigned treatment satisfies condition (2a).

All T0 Survey respondents were enrolled in a raffle to win one of 13 \$75.00 prizes. This initial raffle was held on 6 September 2018, and participants were notified by email prior to any further interview solicitations. For both the T1 and T2 Surveys, individuals were given a choice between receiving a \$5.00 payment, or again entering into a raffle to win one of the \$75 prizes. 54% of T1 Survey respondents elected to enroll in the raffle, and 47% of T2 Survey respondents elected to enroll in the raffle. All payments were delivered via email link using an E-Gift Card merchant.

All reinterview attempts followed best practices from similar experimental designs: 1) all initial email solicitations coincided with a text message reminder (to those that provided their cell phone number); 2) multiple email reminders were sent to encourage prompt responses; and 3) the T1 and T2 Surveys were each left open for 2 weeks after the initial solicitation (26). In addition, we offered all individuals who had not responded within a week of the initial reinterview solicitation an additional \$5.00 payment (see Fig. S5), and sent email reminders when the survey period was

about to expire (see Fig. S6). We applied these recruitment efforts equally to all subjects regardless of whether they were in treatment or control. Table S1 shows the total number of individuals responding to each survey wave, by treatment and control groups.

1.3 Sample Characteristics

Our sample frame was the list of voters registered at residential addresses within the city. Most voter files in the United States contain some limited demographic information, such as sex, and date of birth. Only a few U.S. states record race/ethnicity in their registration data. In the state where this study was conducted, the voter registration database did not include race/ethnicity.

In an attempt to benchmark the race/ethnicity of respondents against the voter file we predicted the race/ethnicity of all individuals in the voter file using the `wru` package in R, which draws on U.S. census data to make probabilistic predictions of membership in 5 categories (white, black, Hispanic, asian, and other), conditional on an individual's surname, geolocation, sex, and age (45). Table S2 reports these basic descriptive statistics, alongside the number of respondents at each stage of the study. The response rates at each wave were on par with those that have been obtained in previous mail-to-online panel designs (26, 33). However, these comparisons suggest female voters were over-represented in the baseline survey, and black and Hispanic voters were under-represented.

The accuracy of the race/ethnicity predictions reported in Table S2 is generally unknown since the state voter file does not collect these data. However, we can evaluate a subset of these predictions against the 2,013 self-reported responses in the T0 Surey. We assess performance by calculating the groupwise error rate, which is the proportion of voters in each race/ethnicity group who were incorrectly classified by the algorithm. We also compute the proportion of false positive and false negatives across each group. A “false positive”, for example, occurs when a non-black voter is classified as black; whereas a “false negative”, occurs when a black voter is not clasified as black.

These results, along with the true prevalence of each group (based on self-reported race in the T0 Survey), are presented in Table S3. The algoritm performed well for classifying Hispanic voters. However, it performed considerably worse for non-Hispanic voters: approximately 22% of white voters, and 20% of black voters were incorrectly classified. Further, the algorithm performed very poorly at detecting voters who were not a member of the three major race/ethnicity groups: although only 9% of voters in the “Other” category were incorrectly classified, the false negative rate was 79%.

These results suggest the algorithm overestimates the proportion of black and Hispanic voters, and underestimates the proportion of white and other voters. These consequences are apparent when comparing the predicted race/ethnicity classifications in column 2 of Table S2 with the true prevalence rates in Table S3. For example, although 27% of survey respondents were classified as black, the true prevalence rate was 19%. Nonetheless, it appears that black and Hispanic voters were less likely to respond to the initial recruitment mailer, and are therefore under-represented in the study.

Importantly, we did not find evidence of differences in background characteristics across treatment and control groups at any wave of the survey (see Design Diagnostics Section). Predicted race/ethnicity are only reported here for exploratory purposes. The self-reported race/ethnicity data in the T0 Survey were used for blocking in the random assignment procedure, and all data analyses.

1.3.1 Attitudes toward police

We can benchmark attitudes toward the police against recent nationally representative survey data published by Pew Research Center for two questions: confidence in the police and the police feeling thermometer. Table S4 compares confidence in the police recorded at baseline (T0 Survey) with the most recent Pew data¹. Overall, study participants were less likely to report “A lot” of confidence in the police, relative to Pew respondents. However, about 77% of the respondents in both surveys still reported “A lot” or “Some” confidence in the police. Ratings were similar by race/ethnicity: black respondents reported lower levels of confidence in the police in both surveys. The Pew data only facilitate comparisons for black and white survey respondents.

Table S5 compares police feeling thermometer ratings recorded at baseline (T0 Survey) with the most recent Pew data². For comparison purposes, the feeling thermometer ratings (0-100) are binned into the categories reported by Pew: very cold (0-24), somewhat cold (25-49), neutral (50), somewhat warm (51-75), very warm (76-100). Overall, study participants felt less warm toward police than survey respondents in the Pew survey. Black and Hispanic study participants were *more likely* to view police officers warmly, compared to black and Hispanic respondents in the Pew survey. White study participants were *less likely* to view police officers warmly, compared to white respondents in the Pew survey.

The relatively cool ratings provided by whites may partly be explained by the fact that the city’s white voters are very liberal: 84% of white respondents reported a Democratic party affiliation, with 54% of those reporting a “very strong” attachment to the Democratic party. Approximately 90% of black respondents were Democrats (47% of those reported a “very strong” attachment), 83% of Hispanic respondents were Democrats (36% of those reported “very strong” attachment). Of those in another racial/ethnic group, 84% were Democrats (37% of those reported a “very strong” attachment).

¹See <http://www.pewresearch.org/fact-tank/2017/09/15/deep-racial-partisan-divisions-in-americans-views-of-police-officers/>

²See <http://www.pewresearch.org/fact-tank/2017/09/15/deep-racial-partisan-divisions-in-americans-views-of-police-officers/>

1.4 Pre-Analysis Plans

An initial pre-analysis plan (PAP), dated 2018-09-15, was uploaded to the Open Science Framework (OSF) website (<https://osf.io/>) on 16 September 2018. This PAP was uploaded after baseline data were collected in the T0 Survey, and the random assignment procedure had been performed, but before data collection was completed in the post-treatment surveys. Prior to the upload, a visit had been attempted at 86 of 926 households in the treatment group.

After this PAP was written and uploaded to OSF, we decided to offer additional \$5 payments to those who had not responded to followup surveys within a week of their initial reinterview invitations with the goal of increasing response rates. We did not randomly assign these additional incentives in an attempt to discern their effectiveness. A supplement to the PAP (dated 2018-10-02) was uploaded on 4 October 2018 to document this. The supplement also documented delays in treatment delivery, and corrected two minor errors related to the construction of a secondary outcome measure described in the PAP.

We later discovered an error in our description of the reinterview procedure for the followup surveys, which stated that T2 survey invitations were sent out 3 weeks after a respondent had completed the T1 survey. This was incorrect, as the procedure for T2 survey invitations was analogous to the procedure for T1 invitations: the reference point (for units in treatment and control units in their quartet) was the timing of *treatment delivery*, rather than the timing of T1 survey completion. We also realized that we neglected to mention in the previous supplement that we had included a 7-item Cognitive Reflection Test (CRT) at the end of T2 Survey, rather than the 12-item questionnaire about trust in government that appeared at the end of the T1 Survey. The CRT data were collected for exploratory purposes, and are not part of any analyses presented here. A second supplement (dated 2018-10-28) uploaded to OSF on 31 October 2018 clarified these issues.

The PAP and both supplements (available at: <https://osf.io/afer7/>) are attached in Appendix B. Note that the dates on these documents reflect the time their final version was compiled with a text editor, and not the date they were subsequently uploaded to OSF.

1.5 Intervention Details

The design of the intervention was motivated by the research literature on intergroup contact, participant observation research conducted in the field with patrol officers, and unstructured interviews with current and former members of the police department who were vocal proponents of community policing. The shared insights of these individuals drew heavily on past experiences as patrol officers, and stressed the importance of 1) establishing common connections (e.g. “we [the police] are part of the community”); 2) emphasizing cooperation toward common goals (e.g. “we [the police] need your help to make this community safe”); and 3) personalized interactions (e.g. “Here’s my phone number. Somebody playing the music too loud at 2am and you don’t want to call down to the station? Give me a call: I’m your champion in the neighborhood.”).

Individual officers interviewed would frequently acknowledge that although “community policing” via resident centered interactions was not “traditional policework”, these interactions established an important foundation for trust and reciprocity. In conversations, this foundation was often cited as the key ingredient for generating outcomes (e.g. solving crimes) that matter most for professional advancement within a police department. For example, an excerpt from the field notes of an interview conducted with a former homicide detective (Officer names are pseudonyms) is illustrative:

Leon gave a personal example [of the benefits of community policing] from his time on the force. In 2007 a cop was shot in the face [Robert told me this same story during our recent patrol shift] and Leon was the one who solved the crime. He attributed this to the rapport he had established with the residents in his district. They held him in high regard and trusted him. They trusted he would protect their identities. Someone called the station after the shooting and left a number saying they had information but would only speak to “Big Leon” [...] The caller later informed on the shooter and Leon was able to solve the crime.

Early participant observation research conducted in 2016 revealed that some patrol officers would voluntarily provide their personal cell phone numbers to residents in order to facilitate direct contact, typically with the express purpose of reporting crimes; but also for police assistance on a variety of other matters. A field notes excerpt from a patrol shift with “Heather” on a Wednesday night in 2016 is illustrative:

Near 9.00pm we went to pick up one of the rookies on the walking beat to take him to get food downtown. As he got in the car, Heather received a text from the mother of a 13 year old girl who had a restraining order on [a 20 year old male recently arrested for sex with minors]. The mother heard from a neighbor that the male was trying to visit the girl. The rookie entered the car and said “I’ll arrest his ass right now” as we rushed to the scene. [...] This was a personal call that Heather received on her cell phone. It didn’t go out over the radio, and therefore wouldn’t be captured in the “calls for service” database. Heather explained that many people in the community would make similar calls for assistance to her personal phone. This woman, a single mother with 5 kids, worked as a home care nurse, which often required overnight stays with patients. Heather had developed a personal relationship with the family, and would often check in on the kids while their mother worked overnights.

In 2017 all police officers in the department were given department issued cell phones in order to

promote non-enforcement contact with residents. This was explicitly promoted as a “community policing” policy, and officers were encouraged to engage in this behavior prior to the intervention period. However, there was no attempt to systematically record this behavior and measure the consequences prior to this intervention.

1.5.1 Officer Training

All participating officers attended a 1 hour training session at police headquarters led by the researchers. Training sessions emphasized the four tenants of promoting positive intergroup contact (23-25):

- Equal status between groups in the situation.
- Cooperation toward a common goal.
- Support of authorities, law or custom.
- Personal interactions.

Along with three³ components of procedural justice in police-civilian interactions (39-41):

- Encouraging citizen participation (“providing voice”) during the interaction.
- Communication of dignity and respect during the interaction.
- Communication of trustworthy motives during the interaction.

These foundational concepts for promoting procedural justice and positive intergroup contact are closely related, and were communicated during the training sessions during each of four different steps in the interaction process:

- **The 20 Second Rule.** Officers were trained to immediately disarm any potential anxiety within the first 20 seconds of contact by signaling they were not there in any law enforcement capacity. Rather, the interaction was an equal status engagement between community members with the superordinate goal of improving the community. Officers would signal respect by initiating a formal greeting (e.g. extending a handshake) with whoever answered the door. Officers were trained to anticipate nervous or suspicious reactions, and **not** to interpret this as suspicious behavior.
- **Providing voice.** Officer’s were trained to solicit feedback from the resident, and emphasize their common connection toward the pursuit of a superordinate goal. Officers would identify and emphasize common connections, using “we” and “our community” to signal a common connection in pursuit of a superordinate goal: working together to improve the community. For example, “How can **we** work together to improve **our** community?”
- **Address concerns.** Officers were trained to anticipate negative reactions, to allow the residents to express their concerns or complaints without judgement, and then acknowledge their perspective. Officers would communicate that both the individual Officer, and the larger Police Department, are committed to community policing because they care about community members and their well being. Thus communicating trustworthy motives.

³The fourth element “explanation and neutrality” was not emphasized directly. This dimension of procedural justice concerns communicating impartiality in an officer’s decision making process. For example, that an officer’s decision to arrest an individual is motivated by objective criteria such as noncompliance with the law, and not the result of the officer’s individual biases, such as racial prejudice.

- **Make it personal.** Officer's were trained to personalize the interaction by shaking hands with the resident, and providing their personal business cards. The business cards contained the name of the officer and the contact details for the police department. Officers provided their work cell phone number on the cards by hand writing it in a designated area.

1.5.2 Intervention Procedure

Uniformed patrol officers made unannounced visits to the homes of randomly assigned residents in their police district (see Figure S2). If the individual(s) who came to the door matched the named individual(s) in our records, the officer marked them as contacted using a mobile application installed on their department issued cell phones. In households with multiple residents, only those who the officer made direct contact with were marked as contacted.

The training sessions emphasized that the content of interactions should be personalized, rather than scripted or robotic, in their delivery. Officers were trained to follow a 5 step process when delivering the intervention:

1. **Disarm Anxiety.** “Hi, I’m Officer [officer’s Surname] with the [City Name] Police Department. I am one of the officers in this area. Everything’s OK. No one is in trouble and everyone’s safe. I’m here because I care about our community, and today I’m personally introducing myself to our community members.”
2. **Identify Resident(s).** If the individual(s) who came to the door matched the voter file record, then the officer marked them as contacted in their mobile application. In households with multiple residents, only those who the officer made direct contact with were marked as contacted.
3. **Provide Voice.** “A lot of times we only see people when there’s a problem, or because someone is being arrested, but we also care about helping people and working together to improve our community. Your input is a valuable part of this partnership. Today I’m here to ask about how you’re doing, and learn about your views on how we can work together to keep our community safe.”
4. **Address Concerns.** “Earlier you mentioned [resident concern] as a concern. What do you think should be done?” Responses should be resident centered. Remain in a “learning orientation”. Listen and acknowledge their concerns. Demonstrate trustworthiness and respect for the resident.
5. **Make it Personal.** “Thank you for taking the time to speak with me today. Before I go, please take this card. These are my personal contact details. We are committed to building a cooperative relationship with our community, and we depend on you to help us keep our community safe. If you have any problems, please don’t hesitate to reach out to me directly. Thanks for your time, and have a nice day.”

1.5.3 Background Characteristics of Participating Officers

The senior leadership at the police department (PD) initially selected 20 uniformed patrol officers from the B-Squad (patrol shifts between 3.00pm to 11.00pm or 4.00pm to 12.00am) as potential

participants in the study. All selected officers, if they agreed, were then enrolled to participate in the study. None of the selected officers declined to participate.

Two of the officers (1 white male, 1 white female) initially selected to participate were placed on a special crime suppression assignment in response to a pattern of drug overdoses in a city park that occurred during the study period. These two officers were unavailable for the entire duration of the study. Another officer (white male) selected to participate was subsequently moved to an administrative assignment after just one shift of canvassing. One additional officer (white male) was added to the study after the initial selection period to take the place of this officer.

Therefore, 21 officers received training to participate in the study, but only 19 officers delivered one or more treatments. The median age of these 19 officers was 32 (with a minimum of 25 and maximum of 49), and the median years of service on patrol was 3 (with a minimum of 2 and a maximum of 19). Five of the 21 selected officers were black (4 male, 1 female), four were Hispanic (all male), and ten were white (8 male, 2 female).

1.5.4 Implementation and Field Observations

The intervention period began on Tuesday 11 September 2018 and ended on Wednesday 17 October 2018. During the intervention period, one home visit was made to all households assigned to the treatment group, and no home visits were made to households assigned to the control group. We initially anticipated all visits would be conducted within 10-20 days, whereas the intervention period lasted for 37 calendar days. This extended time-frame was subject to several factors beyond the control of the researchers.

First, all home visits were restricted to the 6.00-8.00pm time frame during the weekdays, and prior to 8.00pm on the weekends. These time slots were selected to maximize the number of successful conversations, subject to operational and resource constraints at the PD. For example, most participating officers usually began their weekday shifts at 3.00pm or 4.00pm. During the first 1-2 hours of each new shift, the norm among incoming patrol officers is to “clear the board” of any outstanding calls for service that the previous squad could not finish. Therefore, participating officers were typically occupied with responding to backlogged calls for service prior to 6.00pm.

Second, call volume was unusually high during the intervention period, and the department was facing a significant staffing shortage, at approximately 72% of budgeted patrol officers. These included several “Priority One” calls for service that completely diverted participating officers from canvassing activity. For example, a shooting on 22 September 2018 interrupted a large weekend canvass that was scheduled because participating officers were responders to this incident.

Third, severe thunderstorms and flooding prevented scheduled knocks on several days during the first two weeks of the study. For example, it rained on 16 of the 37 days while the study was in the field, with at least one severe thunderstorm (two of which delivered more than 2 inches of rainfall in a 24 hour period) occurring each week in September 2018.

Fourth, staffing constraints restricted the number of officers available during the initial phase of the intervention. These restrictions were particularly acute during the second week of the intervention period. Two officers sustained work related injuries and were out for approximately one week each during the second week. An additional officer was placed on a special assignment during the second

week, and was replaced by another trained officer approximately two weeks later. Finally, five participating officers were on vacation for at least one week each during the second week of the intervention period.

During the intervention period, no staff member at the PD (including participating officers and the senior leadership) was made aware of the details of the survey design or content for any of the surveys administered to city residents. No staff member at the PD had access to, or any influence over, the survey instruments, question wordings, or pre-analysis plans. Participating officers did not indicate to residents that the community policing intervention was in any way connected to the University funded survey, which was described to residents as a public opinion study.

The PI was embedded with each of the participating officers at least once (for at least 2 hours of shift time). During all police-civilian interactions, the PI waited in the patrol car while officers canvassed homes in order to stay out of view of the residents. The PI never accompanied an officer to the door of a home, and therefore the content of the police-civilian interactions was never directly monitored or overheard by the PI or any other researcher.

The modal interaction observed in the field was unsuccessful (resident was not home or did not come to the door). The shortest successful interaction (resident came to the door and script was delivered) observed in the field lasted for approximately 5 minutes and the longest interaction lasted approximately 40 minutes. The modal successful interaction observed in the field lasted approximately 10 minutes.

All interactions were recorded by officers using the application installed on their work issued cell phones. The primary goal here was to know exactly which respondents were contacted in order to facilitate estimation of causal effects among contacted subjects (more below). Immediately after each visit, participating officers recorded whether the individual(s) listed at the address were reached, and rated the quality of their interactions with residents on a 5 point scale from “very negative” to “very positive”.⁴ Officers could also provide (optional) feedback as an open-ended text response after each interaction.

In total, 412 of the 1,007 residents assigned to the treatment group were contacted by officers, for a contact rate of 41%. Figure S3 shows the distribution of officer ratings across all 412 interactions. The vast majority of interactions ($368/412 = 89\%$) were rated as either “positive” or “very positive”. About 10% ($42/412$) of the interactions were rated as “neutral”. One interaction was rated as “very negative” and one interaction was rated as “negative”.

The “very negative” interaction took place near the end of the intervention period. The officer reported the interaction to the PI via phone immediately following the visit. The officer (a white female, approximately 5'2" tall) knocked on the door of a home in an affluent neighborhood. The resident (a white female of approximately the same age) came to the door, and the officer proceeded to introduce herself and deliver the treatment. The officer reported that the entire script was followed (starting with “the 20 second rule”, etc.), and when asked if she had any feedback about community policing, the resident then replied: “The only thing I’ll say is that this has been the most terrifying experience of my life. I’ll excuse myself now.” The resident then ended the interaction by shutting the door on the officer before she could respond, or provide her contact card.

⁴Officers were asked “Overall, how would you rate the quality of this interaction?” [Very Negative (1), Negative (2), Neutral (3), Positive (4), Very Positive (5)]

The “negative” interaction took place near the beginning of the intervention period. The officer noted in the open ended responses following the interaction that the resident (a male of unknown race⁵ approximately the same age) said he was “dissapointed with the cops in his area”. The officer (a Hispanic male) did not immediately report the interaction, but mentioned it in a conversation with the PI during the line-up period for one of his subsequent shifts. The officer explained the script was delivered in full, and when the resident was asked if he had any feedback, he explained how he was “dissapointed” about slow police response times to past calls for service regarding cars parked in front of the driveway for his apartment building, with individuals loitering in the area. The officer reported he spoke with the resident on a subsequent day, and responded to a report of individuals loitering outside the building by parking his patrol car across the street. These individuals soon left the area, without making contact with the officer.

1.5.5 Monitoring for Unanticipated Problems

Researchers were in frequent communication with the PD leadership, as well as all participating officers conducting the home visits in the field. The PD agreed to promptly report any unanticipated events (e.g. arrest or confrontation) that arose as a result of a visit. Unanticipated Problems Involving Risks to Subjects or Others (UPIRSOs) are defined by the IRB as any incident, experience or outcome that meets the following criteria:

1. Is unexpected (in terms of nature, specificity, severity, or frequency) given (a) the research procedures described in the protocol-related documents, such as the IRB-approved protocol and informed consent document and (b) the characteristics of the subject population being studied; AND
2. Is related or possibly related to participation in the research (possibly related means there is a reasonable possibility that the incident, experience, or outcome may have been caused by the procedures involved in the research); AND
3. Suggests that the research places subjects or others at greater risk of harm (including physical, psychological, economic, legal, or social harm) than was previously known or recognized.

No UPIRSOs occurred during the course of the intervention period. No UPIRSOs, or any other unanticipated events, were reported to the researchers by the PD. The most common feedback reported by participating officers was that the interactions with residents were positive, and that having non-enforcement related interactions with the public provided a morale boost that improved the overall quality of their shifts.

1.5.6 Suspicion about Survey and Intervention Connection

Following best practice recommendations for the design and analysis of field experiments with survey outcomes, we attempted to conceal the survey’s connection with the intervention by 1) presenting it as a university-sponsored public opinion survey of city residents; 2) including a variety of unrelated questions in each wave about city government, as well as local and national politics (see Appendix C

⁵He was described by the officer as "light skinned", possibly Hispanic or black. He listed "Middle Eastern" for race/ethnicity in the T0 Survey.

for all survey instruments); 3) waited for a period of three days after the officer visits were conducted before sending our first request for re-interview.

In addition, officers were trained to make a note in their cell phone app if anyone they visited raised suspicion about the connection between the community policing visit and the survey during the course of their interaction. In the event that a resident did raise such a concern, officers were trained to explain that 1) they were instructed by their District Manager to visit residents in their district and introduce themselves as part of the Department's community policing efforts; 2) that the Department did not fund or create the survey; and 3) that they did not have access to any of the survey data.

Of the 412 people who received a visit, five persons raised a concern about the connection between the survey and the intervention. Two of these cases were raised in the field and subsequently reported to the researchers indirectly via the officers (and also noted in the cell phone app). Neither of these individuals directly contacted the researchers, and both completed the subsequent follow-up survey waves.

The remaining three cases were raised directly with the researchers following the visit and prior to completing any followup surveys. These individuals explicitly raised concerns about the confidentiality of their survey responses. The researchers, in response, addressed these concerns directly by re-iterating the confidentiality assurances provided in the informed consent document: 1) that protections were in place to ensure respondent confidentiality and the protection of all the data collected as part of the study; 2) once the study is complete (before the end of this year), all of the information will be deleted from our records to ensure that nobody – including the researchers – can link your survey responses to your name, address, or any other personally identifying information; 3) these protections are part of the research policy we adhere to at our university and our Institutional Review Board (IRB) approves all research studies before they begin and monitors them while they are in progress to ensure compliance with these policies, and the protection of research participants.

Finally, these persons were again provided with the contact information for the IRB, along with the informed consent document they agreed to at the beginning of the study. Our main goal in responding to these inquiries was to reassure participants that the police department did not design the survey or have access to any of their survey responses. In two of these suspicious cases, the individuals were unsatisfied despite these assurances and were withdrawn from the study. In the other case, the individual's confidentiality concerns were satisfied, they continued to participate, and completed both follow-up waves. Communications with the two individuals who withdrew from the study further escalated and they were not satisfied with the researchers' assurances that their survey responses were not shared with the police department. This would be a violation of IRB protocol, and we therefore reported these individuals' concerns to the IRB.

The IRB then investigated the circumstances, concluded the researchers were following protocol and, importantly, verified that no survey data were shared with the police department and only known to the PI and study researchers. A formal notification letter was then sent to these two individuals directly from the IRB notifying them of the result of their investigation (see Fig. S8). After the study was complete, all study participants were then sent a debrief letter from the PI explaining that the researchers partnered with the police department to measure the efficacy of the community policing visits using questions in the opinion survey (see Fig. S9).

2 Outcome Measurement and Scaling

All primary outcome measures were derived from widely-used survey batteries that have appeared in the research literature on civilian attitudes toward the police (3,6,34). Minor modifications to existing batteries were made based on the results from a survey experiment conducted on Amazon Mechanical Turk (MTurk) in March 2017. Specifically, we used factor analysis to select 19 questions, from a larger universe of 26, that tapped four key dimensions: legitimacy, cooperation, compliance, and perceived effectiveness. For analysis purposes, we recode responses so that higher values indicate more positive attitudes toward the police. The primary and secondary outcomes, along with the procedures for combining and scaling them to create indices, were articulated in our pre-analysis plans filed with OSF.

2.1 Primary outcomes

Each of the primary outcome measures (**cooperation, compliance, performance, legitimacy**) were constructed by combining multiple individual items to create a scale. Scales were constructed with the `scale_sum()` function presented in Appendix A.1, which standardizes the individual items (to mean 0 and standard deviation 1) associated with each domain and then sums to create a single outcome measure for each of the four dimensions (which is standardized to mean 0 and standard deviation 1). We transform all outcome measures to a 0-100 point scale to facilitate interpretation and comparison across measures. Fig. S10 shows the (pre-treatment) baseline distribution of the four primary outcome measures in the treatment and control group.

2.1.1 Cooperation

For the next set of questions, we will ask you to estimate how likely you are to engage in certain kinds of behaviors that involve contact with **your local police department**.

- If the police were looking for a suspect who was hiding, and you knew where that person was, how likely would you be to provide the police with information?
- How likely would you be to call the police to report a crime?
- How likely would you be to report suspicious activity to the police?
- How likely would you be to attend a community meeting to discuss problems in your neighborhood with the police?

All items are recorded on a 7-point scale from Extremely Unlikely to Extremely Likely with a neutral midpoint.

2.1.2 Compliance

- If the police tell you to do something, you should do it.
- There are times when it is ok to ignore what the police tell you to do (reverse coded)

- It is your duty to accept the decisions made by the police even when you disagree with them.
- It is your duty to do what the police tell you to do, even if you do not like the way they treat you.

All items are recorded on a 7-point scale from Strongly Disagree to Strongly Agree with a neutral midpoint.

2.1.3 Performance

- I have confidence that the police in [City Name] can do their job well.
- The police in [City Name] are effective at fighting crime.
- The police in [City Name] are effective at helping people who ask for help.
- When people in your neighborhood call the police for help, they respond quickly.

All items are recorded on a 7-point scale from Strongly Disagree to Strongly Agree with a neutral midpoint.

2.1.4 Legitimacy

We'd like to hear more about how you feel about the police in [City Name]. Do you agree or disagree with the statements below?

- They care about the well-being of everyone they deal with
- They make fair and impartial decisions
- They act in ways consistent with your own ideas about what is right and wrong
- They have the same sense of right and wrong that you do
- They stand up for values that are important to you
- They behave according to the law when dealing with people
- They make decisions based on their biases or opinions (reverse coded)
- They are often dishonest (reverse coded)

All items are recorded on a 7-point scale from Strongly Disagree to Strongly Agree with a neutral midpoint.

2.1.5 Index of all Primary Outcomes

As specified in our pre-analysis plans, we examine the overall effect of the treatment on the primary outcomes by creating a single dependent variable. This Index of Primary Outcomes was constructed by extracting the first component from a principal components analysis (PCA) on all the primary outcome measures described above. This is the same dimension reduction procedure used for index

construction in related studies (26), except the underlying PCA calculations are performed using singular value decomposition, rather than spectral decomposition, to improve numerical accuracy. The `scale_outcome()` function in Appendix A.1 provides the R code.

2.2 Secondary Outcomes

2.2.1 Confidence in Police

The first five questions were developed from qualitative fieldwork and interviews with current and former officers in the [POLICE DEPARTMENT] and are not part of any established research literature. The last measure, based on a question asked by Gallup since 1965, was incorporated for reasons described in the first supplement to our pre-analysis plan dated 2018-10-02 (see Appendix B). In our pre-analysis plans this index was called “Index of Community Support”. One reviewer noted the ambiguity of this label. To address this concern, we re-labeled the index “Confidence in Police”.

- The police make me feel safer in my neighborhood.
- The police are trying to make my community better.
- The police respect the people in my community.
- The police treat all people equally.
- The police are a part of my community.
- I respect the police in my community.

All question prompts ask for specific evaluations about the local police department. Responses are recorded on a 7-point scale from “Strongly Disagree” to “Strongly Agree”. Responses to these questions were all highly correlated in the baseline survey ($\alpha = 0.90$). They were combined into a single scale using teh `scale_sum()` function.

2.2.2 Police Feeling Thermometer

The question prompts associated with all the primary and secondary outcome measures described above ask respondents to provide specific evaluations about their local police department, the [POLICE DEPARTMENT]. As a measure of general affect toward “police”, we also included a 100 point “feeling thermometer” toward police embedded among various other groups. The feeling thermometer is a purely affective measure that appears in various public opinion surveys across time. It is analyzed separately.

2.2.3 Police Stereotypes

We measured stereotypes about police on the following dimensions using a 7 point scale where 4 was a neutral midpoint (and the starting position of the survey instrument):

- Hardworking (1) to Lazy (7)
- Humble (1) to Arrogant (7)
- Intelligent (1) to Unintelligent (7)
- Trustworthy (1) to Untrustworthy (7)
- Compassionate (1) to Cold-hearted (7)

An example question prompt is: “A score of 1 means you think almost all people in that group are “hardworking” and a score of 7 means you think all people in that group are “lazy”. A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.”

As with the feeling thermometer, the stereotype questions were embedded into a battery with other groups: firefighters, politicians, doctors, school teachers, and religious leaders. Each individual police stereotype item was scaled such that higher values reflected more negative evaluations of the police. The five individual measures were combined into a single scale using the `scale_sum()` function.

2.2.4 Police Policy Questions

Finally, we include two additional outcomes of specific policy support:

- As you may know, the city of [CITY NAME] has recently required police to wear body cameras that record their activities while on duty. Do you support or oppose police body cameras?
- How would you feel about increasing the number of police on the street by 10 percent, even if it means fewer funds for other public services?

Both items are recorded on a 7-point scale from Strongly Oppose to Strongly Support, with a neutral midpoint. The question about body cameras was included out of general interest, and the final question was added as a more specific measure of support for police (37). These items are analyzed individually.

2.2.5 Trust in Local Government

We also included a measure of trust in local government derived from the trust in government questions that have appeared in the American National Election Studies (ANES) survey since the 1960s. We recoded values so that higher scores correspond to more positive evaluations of local government. We speculated in our pre-analysis plan that treatment (interacting with an agent of local government) may have a small positive effect on trust in local government.

- How often do you think you can trust the government in [CITY NAME] to do what is right? [Always (5); Most of the time (4); About half the time (3); Sometimes (2); Never (1)]
- Do you think that people in government in [CITY NAME] waste much of the money we pay in taxes? [Waste a lot (1); Waste some (2); Don’t waste very much (3)]

- How many of the people running the government in [CITY NAME] are corrupt? [All (1); Most (2); About half (3); A few (4); None (5)]
- Would you say the government in [CITY NAME] is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people? [Run by a few big interests (0); For the benefit of all the people (1)]

We create a single trust in local government measure by extracting the first component from a principal components analysis (PCA) on the four individual measures listed above via the `scale_outcome()` function.

3 Design Diagnostics

In this section, we examine key design assumptions, and conduct statistical tests to evaluate their plausibility. These diagnostic procedures reflect best practice standards in the design and analysis of field experiments in the social sciences (26,46,47), and were not specifically articulated in our pre-analysis plans.

First, we look for differences in background covariates between treatment and control groups across each survey wave. Covariate imbalance in the T0 Survey would raise questions about the random assignment procedure. Covariate imbalance in the T1 and T2 Surveys would provide evidence of differential attrition across groups. We do not find evidence of differences in background characteristics between treatment and control in any survey wave. We present both descriptive results and conduct statistical tests to examine covariate balance.

Next, we look for evidence of differential attrition more explicitly by 1) testing the sharp null hypothesis that treatment has no effect on attrition for any unit; 2) testing the null hypothesis that treatment-covariate interactions have no effect on attrition. These tests are performed independently for both the T1 and T2 survey waves. The first test allows for a powerful direct test that treatment caused subjects to respond at higher (or lower) rates. The second test examines whether treatment caused attrition to differ by covariates; for example, by encouraging individuals who support the police to respond and discouraging unsupportive individuals.

We rely on Randomization Inference (RI) to compute P-values for all test-statistics. This non-parametric approach makes use of the properties of the experimental design to simulate the sampling distribution of a test statistic under a null hypothesis. Unlike conventional parametric tests, inferences (about covariate balance, etc.) do not depend on assumptions about the shape of the sampling distribution (46). P-values below 0.05 are typically considered evidence of covariate imbalance, or differential attrition. None of the statistical tests conducted below produced RI P-values lower than 0.30.

We use all the background covariates that were used for blocking in the household level random assignment procedure, as well as a 7-point measure of party identification that ranges from “Strong Democrat” to “Strong Republican” that was pre-registered to be included for regression adjustment in our estimation procedures. The R code used for all procedures described below is presented in Appendix A.2. We use the `randomizr` package for assistance in all computations (44).

3.1 Covariate Balance

Tables S7, S8, and S9 show covariate balance between treatment and control groups across each survey wave. The Index of Primary Outcomes is scaled to range from 0-100 for presentation. Descriptively, background covariates are well balanced across each wave. We test the null hypothesis of covariate balance between treatment and control groups, across each survey wave, using RI. The null distribution of each test statistic is approximated based on 10,000 permutations of the experimental design. The RI P-value is the proportion of permuted F-statistics as extreme (or more extreme) as the one observed under the null hypothesis of covariate balance. The observed

F-statistic, along with the 0.025th and 0.975th quantiles of the distribution of permuted F-statistics and the RI P-value, are presented for each survey wave in Table [S10](#).

3.2 Differential Attrition

We use randomization inference to test the (sharp) null hypothesis that treatment had no effect on survey response for any unit. This test is performed by essentially estimating the “treatment effect” on a binary indicator for survey response. The RI P-value is the proportion of permuted estimates that are as extreme (or more extreme) than the one observed under the null hypothesis of covariate balance. The observed estimates, along with the 0.025th and 0.975th quantiles of the distribution of permuted estimates and the RI P-value, are presented for each survey wave in Table [S11](#).

We use linear regression of a response indicator on treatment, baseline covariates, and treatment-covariate interactions to test the null hypothesis that all the interaction coefficients are zero. The RI P-value is the proportion of permuted F-statistics that are as extreme (or more extreme) than the one observed under the null hypothesis of no differential attrition. The observed F-statistic, along with the 0.025th and 0.975th quantiles of the distribution of permuted F-statistics and the RI P-value, are presented for each survey wave in Table [S12](#).

4 Estimation and Inference

For estimation and inference, we follow the procedures outlined in our first pre-analysis plan filed with OSF on 15 September 2018 (see Appendix B) and focus on two target parameters: the ITT and the CACE. Unless otherwise indicated, all outcomes are standardized to a 0-100 point scale to facilitate interpretation and comparisons across measures. The replication code for all estimation procedures is provided in Appendix A.

We anticipated some level of non-compliance and therefore articulated two different target parameters for estimation and inference in our pre-analysis plan: the “Intent-to-Treat effect” (ITT) and the “Complier Average Causal Effect” (CACE). In the special case of perfect compliance the CACE is equivalent to the “Average Treatment Effect” (ATE). While the ATE is of theoretical interest, it can only be point-identified in field experiments on human subjects if compliance is forced, or the researcher makes additional assumptions. The CACE (also called the “Local Average Treatment Effect” or LATE) is the average causal effect among a (latent) subset of individuals who would be treated if assigned to the treatment group. In the case of one-sided non-compliance, this corresponds to the effect of treatment on the treated and is often called the “Average Treatment Effect on the Treated” (ATT)⁶. In the manuscript, we use the term ATT (rather than CACE or LATE), and describe the reported estimates as measures of the treatment effect on the treated. Our estimand is the effect among the latent class of compliers (the CACE), not just those who happened to be treated in the realized assignment. Since the expected ATT is equal to the CACE, this distinction is not of major consequence.

As specified in our pre-analysis plan, 1) we use linear regression to estimate ITTs, and Instrumental Variable (IV) regression to estimate CACEs; 2) adjusting for pre-specified background covariates; 3) and robust standard errors clustered at the level of random assignment (household), for all primary and secondary outcomes. Our goal is to test for the existence of treatment effects within the sample rather than generalize to a population or “super-population” of units that did not participate in the study. No weighting is used in any analyses. For textbook treatments of regression adjusted estimators for the ITT and CACE in the analysis of randomized experiments, see (46, 48-49).

We follow the pre-specified asymmetric hypothesis testing procedures for both ITT and ATT estimators for primary and secondary outcomes. Since all point estimates were in the expected direction we would use a one-tailed P -value with a rejection threshold of 0.04 (0.08 two-tailed). This still preserves the property that the rejection region contains 5% of the sampling distribution, but focuses on the bottom 1% and the top 4% instead of the typical bottom 2.5% and top 2.5% (26,50). The only exception is the question about police body cameras. We did not have a prior expectation about the direction of any treatment effect here, and therefore use the conventional symmetric hypothesis test with $\alpha = 0.05$.

4.1 Treatment Non-Compliance

To date, the vast majority of door-to-door canvassing experiments in the published literature have been aimed at testing theories about voter turnout and political persuasion (51-52). In these

⁶see Chapter 4.4.3 in (*48*)

experiments, most individuals assigned to the treatment group do not actually *receive* the treatment. This failure to treat or “one-sided non-compliance” arises for a variety of reasons. For example, canvassers (often volunteers) may fail to contact everyone in the treatment group due to resource or time constraints. More commonly, many individuals assigned to treatment do not come to the door when canvassed (e.g. not home, refuse to answer). A more complex situation, called “two-sided” non-compliance, occurs when individuals assigned to the control group also receive treatment. For example, enthusiastic volunteers might want to deliver a political campaign’s message to everyone.

In this study, none of the homes assigned to the control group received a visit, and a visit was *attempted* at all the homes assigned to the treatment group. Many individuals were not reached at the door, and the study therefore encountered one-sided non-compliance. Of the 1,007 individuals assigned to the treatment group, 412 were contacted by officers (for a 41% contact rate). This rate of non-compliance is much lower than that observed in the typical door-to-door canvassing experiment on voter turnout or political persuasion. As a benchmark, the contact rate is between 20-30% in the most rigorously-executed canvassing experiments that use professional canvassers (26). We conjecture that individuals are simply more likely to answer the door for a uniformed police officer than a political organizer.

As specified in our pre-analysis plan dated 15 September 2018, we classify individuals as “compliers” if they were reached by a police officer when a home visit was attempted and direct human contact was made, regardless of how long the contact lasted, or how the canvassing officer rated the quality of the contact. Thus, we *do not* reclassify treatment status or compliance status based on what happened at the door, nor do we exclude any subject that came to the door from analyses. This limits researcher discretion and is consistent with best practices in the design and analysis of field experiments (26, 46-47).

Our estimation approach for estimating the ATT (Instrumental Variables Regression) rests on two key assumptions to identify the causal effect of treatment among the treated. First, although taking the treatment (e.g. opening the door for the officer) may be confounded, the *assignment* to treatment is unconfounded. This assumption is satisfied by the experimental design. Second, is that the assignment does not have a direct effect on the dependent variable(s) of interest (the “exclusion restriction”). Evidence of differential attrition, for example, could be prognostic of such a violation. This assumption is also supported by the design: we did not tell subjects which group they were assigned to, that they were part of an experiment, or that the visit by the police was connected to the survey.

These two key assumptions are rarely credible in observational studies, where IV estimators are frequently applied. Alternative estimation approaches such as “Per-Protocol” or “As-Treated” analyses are often used in randomized experiments that encounter non-compliance rather than IV methods. Under one-sided non-compliance IV regression is a consistent estimator for the ATT. These alternative methods, however, are generally invalid for recovering causal estimates (see Section 23.9 (49)).

Door-to-door canvassing experiments sometimes use a “placebo” design where subjects in the control group are also contacted by canvassers, but receive a different message; for example, a conversation about recycling as opposed to prejudice reduction (26). The placebo design offers a substantial improvement in statistical precision, and greatly reduces survey costs, as researchers can choose to only reinterview the compliers – those who answered the door – in both groups (26, 33). Moreover,

the researcher can simply estimate the CACE using OLS.

However, one important assumption is that the placebo message does not affect the dependent variable. As specified in our pre-analysis plan dated 15 September 2018, we considered the placebo design but deemed a placebo canvass condition by the police (e.g. having police officers visit homes and deliver a message about something unrelated to policing) as unrealistic, and that no police contact was the appropriate counterfactual. Another important assumption is that the “compliers” are equivalent across the two groups, so that the types of people who open the door to receive a placebo message are the same as those who open the door to receive the treatment message. This rules out a design whereby the placebo message is delivered by an individual not wearing a police uniform (e.g. a paid canvasser or plain clothes officer), which could have significantly reduced compliance rates in the placebo group.

4.2 Results: Primary Outcomes

Estimates and standard errors from the pre-registered covariate-adjusted regression estimators of the ITT and ATT are presented in Table S13, for all primary outcomes, across the T1 and T2 survey waves.

4.3 Results: Secondary Outcomes

Estimates and standard errors from the pre-registered covariate-adjusted regression estimators of the ITT and ATT are presented in Table S15, for all primary outcomes, across the T1 and T2 survey waves.

5 Treatment Effect Heterogeneity

In our pre-analysis plan we stated an intent to examine effect heterogeneity by baseline support in the T0 Survey and race/ethnicity of the respondents. Here we use the index of primary outcome variables as our measure of baseline support and a 4-category factor of race/ethnicity (black, Hispanic, white, other), both of which were used for blocking in the random assignment procedure. As specified in our pre-analysis plan, we have no intention of investigating heterogeneity for any other specific sub-group of survey respondents unless required by a reviewer, in which case we will note the deviation from our pre-analysis plan. We use the default $\alpha = 0.05$ for statistical significance here rather than $\alpha = 0.08$, as we did not pre-register asymmetric hypothesis testing for heterogeneous treatment effects. No expectations about the direction of effects were pre-registered in advance. These analyses should be considered exploratory.

5.1 Race/Ethnicity of Respondents

Our pre-analysis plan dated 15 September 2019 stated we would also examine heterogeneity by race/ethnicity using “the 4-category race variable (black, Hispanic, white, other) with white as the reference group.” We instead estimate conditional causal effects among each sub-group. This is simply to facilitate interpretation. Thus, the point estimates for the “black” subgroup presented below represent causal effects among the population of black study participants, rather than the differences between black and white participants. The results, organized by the relative size of each group in the baseline sample, are presented in Table S14 with sample proportions at each wave.

5.2 Quantiles of Baseline Support

This analysis partitions respondents into 4 approximately equal sized groups, based on their pre-treatment levels of support on the index of primary outcomes. Respondents in the Q1 group are those who fell in the bottom 1/4 of the distribution of baseline support, and those in the Q4 group are those who fell in the top 1/4 of baseline support. Table S16 presents estimates of the ITT and ATT across both survey waves. Point estimates are scaled to range between 0-100 to facilitate interpretation. Thus, a 5 unit change corresponds to a 5 point shift on a 0-100 scale. Note that each group will not contain *exactly* 1/4 of the sample because the partitions are formed using baseline support measures, and not all T0 survey respondents are observed in the T1 and T2 waves. Significant deviations from 1/4 in any of these groups would suggest baseline support levels are predictive of non-response to subsequent surveys. Table S16 also presents the observed fraction of the sample in each of the 4 groups, and the observed deviations from 0.25 are trivial.

6 Figures S1 to S7

Figure S1: Summary of Experimental Design

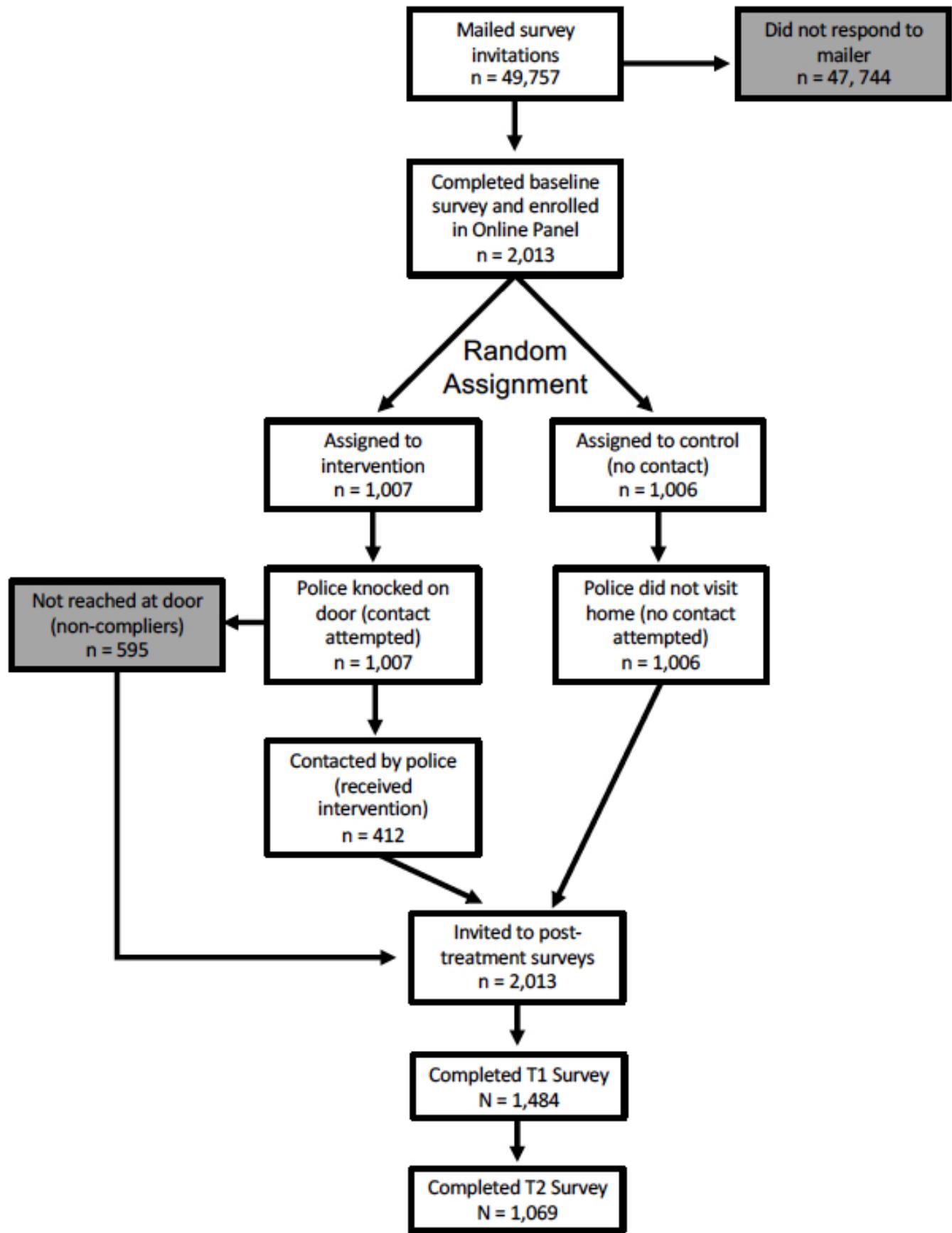


Figure S2: Field Experimental Manipulation

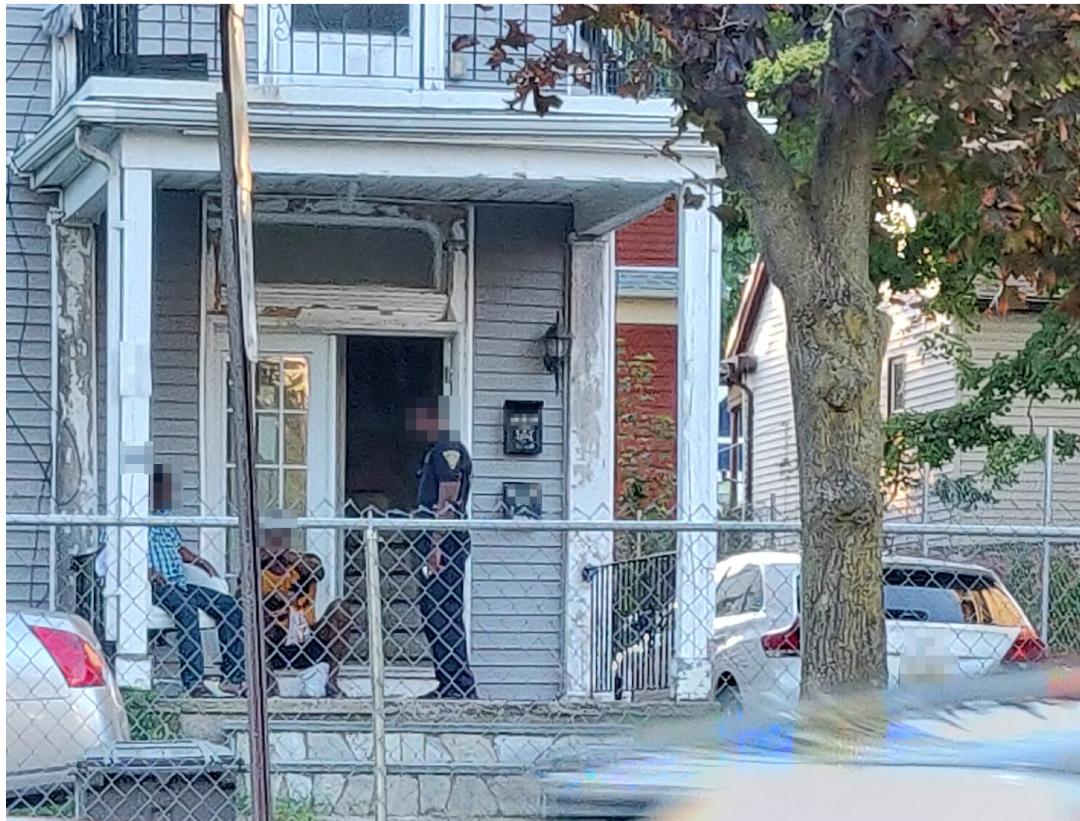


Figure S3: Officer Ratings of Interactions with Residents

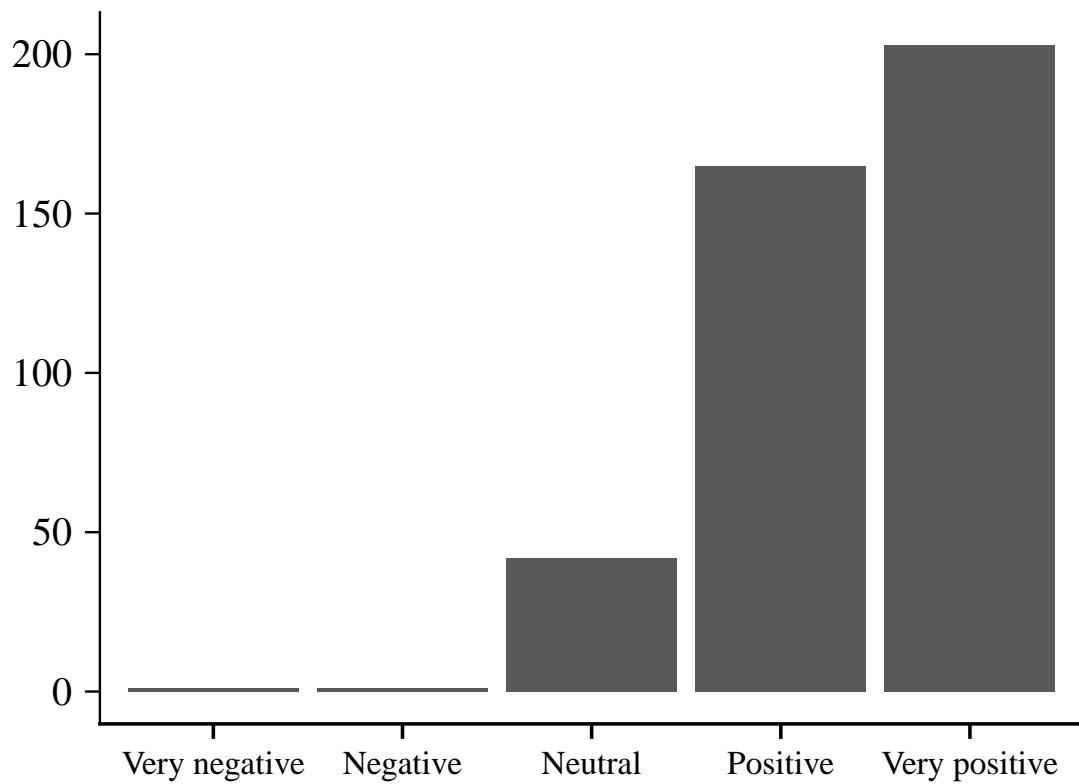


Figure S4: Bilingual Mailer (English Side)

**Your Invitation to the 2018 [REDACTED] Opinion Study
(para los españoles, consulte inversa)**

Dear [REDACTED]

You are one of a small number of people who has been chosen to participate in the 2018 [REDACTED] Opinion Study, a research study being conducted at [REDACTED] as part of a collaboration between researchers from [REDACTED]

You were randomly selected to complete a survey and represent the opinions and experiences of people in your area. The purpose of this study is to better understand people's views about their community, and how these views change over time. Understanding the people of [REDACTED] – their needs and opinions, and their hopes for the future – is vital. Your response would be invaluable to our research and to the community.

We value your privacy, and any information you provide will be kept secure and confidential. Your participation is voluntary, although we hope you will take this opportunity to have your voice heard.

After completing this short initial survey, you will be entered into a raffle to receive one of over a dozen \$75 gift cards that can be spent at your choice of 84 popular companies, including Amazon.com, or donated directly to one of 18 charities, such as the American Cancer Society. By completing the first survey, you will also become a member of the 2018 [REDACTED] Opinion Study Panel. Members of the panel are eligible to participate in follow-up surveys in the months ahead. **For each follow-up survey you take, you choose to either receive a \$5 gift card, or enroll in another raffle with even greater chances of winning a \$75 gift card.**

This survey will take about 10 minutes to complete. You can participate on your computer or smartphone by visiting [REDACTED] with the login information below:

For: [REDACTED] Login: [REDACTED]

The survey is available now. We would appreciate it if you were to **respond by September 5, 2018.**

If you have trouble accessing the survey, please email [REDACTED]
You can also call [REDACTED]

Thank you for being part of this study.

Sincerely,

Figure S5: Example of Additional Incentive Payment Email (English Version)

Dear \${m://FirstName},

We still haven't heard back from you, and your invitation to participate is set to expire soon. We know your time is valuable, and as a way of showing our thanks, **we are sending a \$5.00 payment to you now.** You will receive an email from TangoCard with instructions for accepting this payment within 1 business day.

Once you complete the survey below, you can choose to **receive an additional \$5.00 payment, or enroll in a new raffle with a chance to win one of more than a dozen \$75 prizes.**

Follow this link to the Survey:

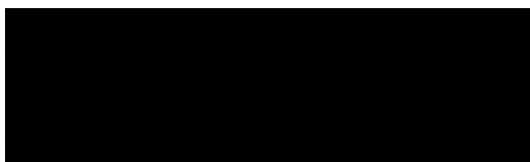
[\\${l://SurveyLink?d=Take the Survey}](#)

Or copy and paste the URL below into your internet browser:

[\\${l://SurveyURL}](#)

The survey will only take 5-10 minutes to complete. **If you are no longer interested in being part of the [REDACTED] Opinion Study Panel, please let us know by clicking the Opt-Out link at the bottom of this email.** If you have any questions or concerns, please email me directly at [REDACTED]

Sincerely,



Follow the link to opt out of future emails:

[\\${l://OptOutLink?d=Click here to unsubscribe}](#)

Figure S6: Example of Final Survey Expiration Reminder (English Version)

Dear \${m://FirstName},

We still haven't heard back from you, and **your invitation to participate is set to expire tomorrow.** We know your time is valuable, and as a way of showing our thanks, **we have sent you a \$5.00 payment from TangoCard.** Please let us know if you did not receive it.

Once you complete the survey below, **you can still choose to receive an additional \$5.00 payment, or enroll in a new raffle with a chance to win one of more than a dozen \$75 prizes.**

Follow this link to the Survey:

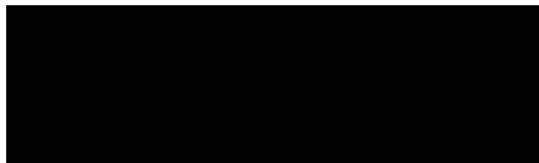
[\\${l://SurveyLink?d=Take the Survey}](#)

Or copy and paste the URL below into your internet browser:

[\\${l://SurveyURL}](#)

If you are no longer interested in being part of the [REDACTED] Opinion Study Panel, please let us know by clicking the Opt-Out link at the bottom of this email. If you have any questions or concerns, please email me directly at [REDACTED]

Sincerely,



Follow the link to opt out of future emails:

[\\${l://OptOutLink?d=Click here to unsubscribe}](#)

Figure S7: Example of Slider Question Format in Online Survey (English Version)

A score of 1 means you think almost all people in that group are "arrogant" and a score of 7 means you think all people in that group are "humble". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.

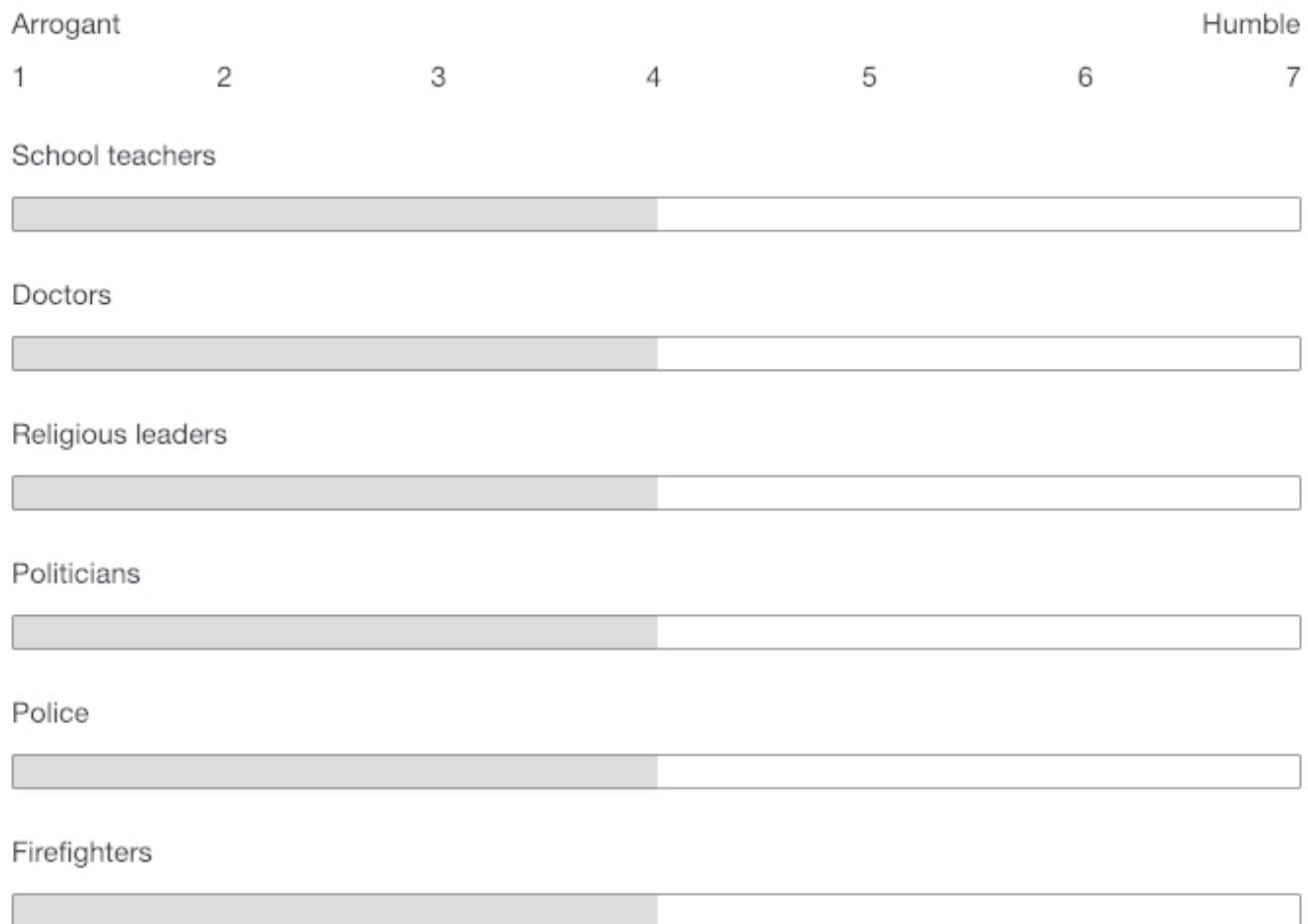


Figure S8: Institutional Review Board Letter to Suspicious Participants

[REDACTED]
[REDACTED]
November 27, 2018

Dear Research Participant,

You recently participated in [REDACTED] Opinion Survey. This study is being conducted by researchers at [REDACTED] with the goal to learn about the views and opinions of the residents [REDACTED]. The Principal Investigator (PI) is [REDACTED]

The Institutional Review Board (IRB) is an oversight committee that reviews and approves research at [REDACTED], and is responsible for ensuring the rights and welfare of research participants are respected and protected. As part of this oversight, PIs are required to report to the IRB when participants have complaints about the research, explain the nature of what caused the complaint, and describe actions taken to address any participant complaints.

The PI of this study reported to the IRB that participants expressed concern that it was possible their private survey data provided to the study team was shared with [REDACTED]. [REDACTED] The IRB has investigated the circumstances of this complaint and has required this letter be provided to the participants that have expressed this concern. Please know that the IRB takes all concerns expressed by study participants very seriously.

If you received a recent visit to your residence by [REDACTED], please be assured the IRB has verified it was not related to your survey responses in [REDACTED] Opinion Survey research study. Further, the IRB has verified that your survey data were not shared with [REDACTED]. Individual survey responses were only known to the PI and study researchers. All survey responses have been anonymized, that is all identifying data have been removed. The IRB also seeks to reassure you if you requested to no longer be enrolled you have been withdrawn from the study. Please note that all participants in the [REDACTED] Opinion Survey study will receive a full debriefing when the study is complete.

Thank you for expressing your concerns. If you would like to further discuss your concerns please feel free to reach out to the [REDACTED] Institutional Review Board at [REDACTED]. Concerns can be confidential or anonymous.

[REDACTED]

Figure S9: Post-Study Debrief (English Version)

Dear \${m://FirstName},

Thank you for being a member of the [REDACTED] Opinion Study! Data collection is now complete, and we would like to provide you more details about the study. The purpose of the study was to investigate the effectiveness of community policing for improving relationships between the police and the citizens of [REDACTED]. In order to do so, the surveys in [REDACTED] Opinion Study were timed to coincide with community policing visits conducted by the [REDACTED]

[REDACTED] officers routinely interact with residents in the neighborhoods they patrol to introduce themselves and encourage cooperation on community issues. To evaluate whether these types of community policing initiatives are effective, we collaborated with the [REDACTED] and randomly assigned half of the homes in [REDACTED] Opinion Panel to receive such a visit. The last visits were completed on 17 October 2018.

That means that, purely by chance, your home may have been selected to receive a community policing visit. We did not tell you about this in advance because we didn't want the possibility of a community policing visit to influence how you responded to the survey questions about policing.

Your individual survey responses have not, and will not, be shared with [REDACTED]. The survey responses you provided as part of [REDACTED] Opinion Study have been anonymized (that is all identifying data has been removed) so no individual survey responses can be connected to you.

We hope that the results of this study will provide knowledge about how to improve relations between police and community members. The results of the study will be shared after it has undergone peer review. This process takes at least 6 to 12 months to complete. If you would like to receive a notification when the results are available for release, please sign up at the link below. Thank you for your valuable contribution to this study!

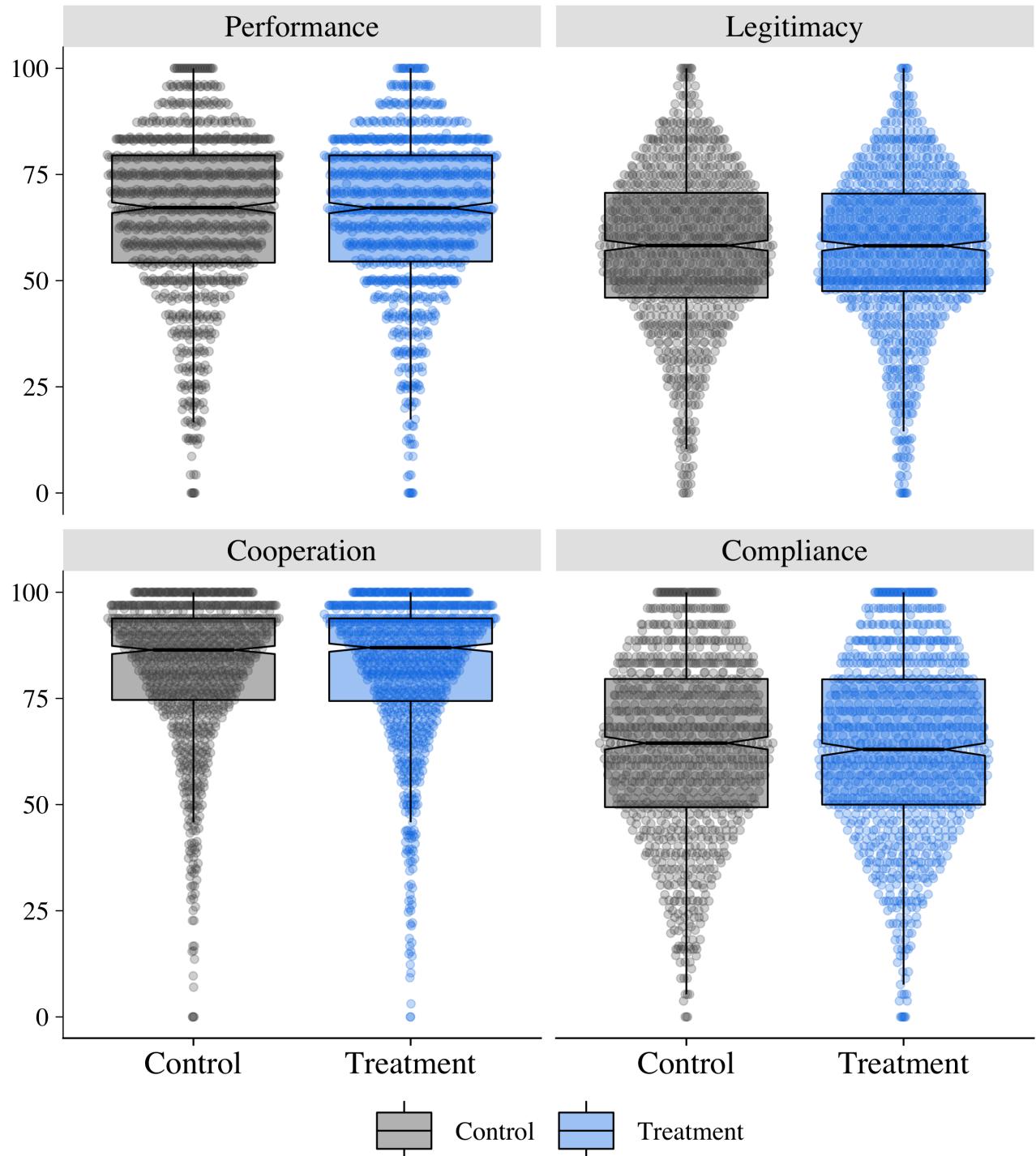
\${l://SurveyLink?d=Follow this link to sign up for updates}

Or copy and paste the URL below into your internet browser:
\${l://SurveyURL}

If you have any questions or concerns, you may contact the researchers directly at [REDACTED]. If you would like to talk with someone else to discuss problems or concerns, or to discuss your rights as a research participant, you may contact the [REDACTED]
Additional information is available at [REDACTED]

\${l://OptOutLink?d=Click here to unsubscribe}

Figure S10: Distribution of Primary Outcome Measures in Baseline Survey



Responses plotted for all individuals that participated in the baseline survey ($n = 2,013$), across each primary outcome measure: performance, legitimacy, cooperation, and compliance. Blue dots denote the individuals nested in treated households ($n = 1,007$) and grey dots denote individuals nested in control households ($n = 1,006$). Points are jittered to reduce over-plotting and notched boxplots are overlaid to summarize the distribution within each group.

7 Tables S1 to S16

Table S1: Survey Respondents by Experimental Condition and Wave

Condition	<i>Survey Wave</i>		
	Baseline (T0)	3 Day (T1)	3 Week (T2)
Control	1006	735	523
Treatment	1007	749	546

Table S2: Characteristics of sample at each stage of study

	Invited	T0 Survey	T1 Survey	T2 Survey
Median Age	41	43	42	42
Pct. Female	0.56	0.61	0.61	0.62
Pct. White	0.36	0.48	0.5	0.52
Pct. Black	0.32	0.27	0.26	0.25
Pct. Hispanic	0.25	0.17	0.16	0.14
Pct. Other	0.07	0.08	0.08	0.09
N	49757	2013	1484	1069

Note:

Race/ethnicity estimates are based on predictions.

Table S3: Empirical validation of race/ethnicity classification

Group	Prevalence	Groupwise error	False positive	False negative
White	0.57	0.22	0.23	0.21
Black	0.20	0.20	0.18	0.26
Hispanic	0.12	0.04	0.03	0.15
Other	0.11	0.09	0.01	0.79

Note:

Predictions validated against self-reported race/ethnicity in T0 Survey (N = 2,013)

Table S4: Reported confidence in the police department by level and group

Group	A Lot	Some	Only a little	None at all
Baseline (T0 Survey)				
Overall	0.26	0.50	0.17	0.07
White	0.28	0.54	0.14	0.04
Black	0.18	0.49	0.24	0.09
Hispanic	0.31	0.43	0.17	0.09
Other	0.26	0.43	0.20	0.11
Pew Survey (2016)				
Overall	0.36	0.41	0.13	0.09
White	0.42	0.39	0.12	0.06
Black	0.14	0.41	0.20	0.24

Note:

Pew data from Aug.16-Sept. 12, 2016 Survey of U.S. adults.

Table S5: Feeling thermometer ratings of police by level and group

Group	Very cold	Somewhat cold	Neutral	Somewhat warm	Very warm
Baseline (T0 Survey)					
Overall	0.07	0.14	0.13	0.25	0.38
White	0.07	0.13	0.11	0.29	0.40
Black	0.08	0.19	0.17	0.22	0.31
Hispanic	0.05	0.12	0.16	0.22	0.43
Other	0.12	0.16	0.15	0.20	0.35
Pew Survey (2017)					
Overall	0.10	0.08	0.16	0.19	0.45
White	0.07	0.07	0.11	0.20	0.53
Black	0.30	0.08	0.28	0.08	0.22
Hispanic	0.08	0.09	0.25	0.18	0.37

Note:

Pew data from Aug.8-21, 2017 Survey of U.S. adults.

Table S6: Uniformed patrol officers by sex and race/ethnicity

	White	Black	Hispanic	Other	Total
Male	108	52	56	2	218
Female	12	12	6	1	31
Total	120	64	62	3	249

Note:

Sworn personnel in patrol division as of 31 December 2018.
249 officers is 72% of budgeted allowance (347 officers)

Table S7: Background Covariates by Treatment Assignment in T0 Survey

	Treat Mean	Treat SD	Ctrl Mean	Ctrl SD
Index of Primary Outcomes at T0	65.01	15.88	65.12	16.57
Age	43.95	14.01	44.08	14.21
Female	0.62	0.49	0.62	0.49
Household size	1.17	0.42	1.17	0.42
Annual income exceeds 50,000-59,999	0.43	0.50	0.42	0.49
Spanish speaker	0.01	0.12	0.01	0.12
Party Identification	2.25	2.25	2.37	2.37
Race/Ethnicity:				
White	0.57	0.50	0.57	0.49
Black	0.19	0.40	0.20	0.40
Hispanic	0.13	0.33	0.12	0.33
Other	0.11	0.32	0.11	0.31
Prior Police Contact:				
Face-to-face contact, last 12 Mos.	0.39	0.49	0.39	0.49
Any prior arrest	0.09	0.29	0.09	0.29
Any prior unfair treatment	0.14	0.35	0.15	0.35
Local Police District:				
District 1	0.07	0.25	0.06	0.24
District 2	0.21	0.41	0.20	0.40
District 3	0.05	0.22	0.05	0.21
District 4	0.04	0.20	0.06	0.24
District 5	0.02	0.13	0.02	0.14
District 6	0.03	0.17	0.02	0.15
District 7	0.21	0.41	0.23	0.42
District 8	0.07	0.25	0.06	0.24
District 9	0.22	0.41	0.21	0.41
District 10	0.08	0.28	0.08	0.28

Note:

Individual-level data from treatment (N = 1,007) and control (N = 1,006) groups.

Table S8: Background Covariates by Treatment Assignment in T1 Survey

	Treat Mean	Treat SD	Ctrl Mean	Ctrl SD
Index of Primary Outcomes at T0	64.80	16.23	65.44	16.26
Age	43.90	14.00	44.09	14.24
Female	0.62	0.49	0.62	0.49
Household size	1.17	0.41	1.16	0.41
Annual income exceeds 50,000-59,999	0.44	0.50	0.46	0.50
Spanish speaker	0.02	0.13	0.02	0.13
Party Identification	2.20	2.20	2.37	2.37
Race/Ethnicity:				
White	0.61	0.49	0.61	0.49
Black	0.18	0.38	0.18	0.38
Hispanic	0.11	0.31	0.11	0.31
Other	0.11	0.31	0.10	0.30
Prior Police Contact:				
Face-to-face contact, last 12 Mos.	0.38	0.49	0.39	0.49
Any prior arrest	0.08	0.27	0.08	0.27
Any prior unfair treatment	0.13	0.34	0.13	0.34
Local Police District:				
District 1	0.08	0.27	0.07	0.25
District 2	0.21	0.41	0.22	0.42
District 3	0.05	0.21	0.04	0.21
District 4	0.04	0.19	0.05	0.21
District 5	0.02	0.13	0.02	0.15
District 6	0.03	0.17	0.02	0.14
District 7	0.22	0.42	0.24	0.43
District 8	0.06	0.25	0.05	0.22
District 9	0.20	0.40	0.21	0.40
District 10	0.09	0.28	0.08	0.27

Note:

Individual-level data from treatment (N = 749) and control (N = 735) groups.

Table S9: Background Covariates by Treatment Assignment in T2 Survey

	Treat Mean	Treat SD	Ctrl Mean	Ctrl SD
Index of Primary Outcomes at T0	64.88	16.31	65.57	15.84
Age	44.00	13.97	44.39	14.24
Female	0.63	0.48	0.62	0.49
Household size	1.17	0.40	1.17	0.41
Annual income exceeds 50,000-59,999	0.45	0.50	0.48	0.50
Spanish speaker	0.01	0.12	0.01	0.08
Party Identification	2.15	2.15	2.41	2.41
Race/Ethnicity:				
White	0.64	0.48	0.65	0.48
Black	0.17	0.37	0.16	0.36
Hispanic	0.09	0.29	0.09	0.28
Other	0.10	0.30	0.11	0.31
Prior Police Contact:				
Face-to-face contact, last 12 Mos.	0.38	0.49	0.39	0.49
Any prior arrest	0.07	0.25	0.06	0.24
Any prior unfair treatment	0.12	0.32	0.11	0.32
Local Police District:				
District 1	0.09	0.28	0.07	0.25
District 2	0.22	0.41	0.23	0.42
District 3	0.04	0.19	0.05	0.22
District 4	0.03	0.18	0.05	0.22
District 5	0.01	0.10	0.01	0.11
District 6	0.03	0.16	0.02	0.13
District 7	0.24	0.43	0.24	0.43
District 8	0.05	0.23	0.05	0.22
District 9	0.21	0.41	0.20	0.40
District 10	0.09	0.28	0.09	0.28

Note:

Individual-level data from treatment (N = 546) and control (N = 523) groups.

Table S10: Randomization Inference (RI) for covariate balance

Survey Wave	Observed F-Statistic	0.025th Quantile	0.975th Quantile	RI P-value
T0 Survey	0.50	0.22	1.00	0.50
T1 Survey	0.54	0.31	1.17	0.71
T2 Survey	0.76	0.37	1.33	0.48

Note:

Quantiles of null distribution and RI P-values from 10,000 permutations of the experimental design.

Table S11: Randomization Inference (RI) for the sharp null hypothesis of no effect on response for any unit

Survey Wave	Observed Estimate	0.025th Quantile	0.975th Quantile	RI P-value
T1 Survey	0.01	-0.04	0.04	0.52
T2 Survey	0.02	-0.04	0.04	0.32

Note:

Quantiles of null distribution and RI P-values from 10,000 permutations of the experimental design.

Table S12: Randomization Inference (RI) for differential attrition

Survey Wave	Observed F-Statistic	0.025th Quantile	0.975th Quantile	RI P-value
T1 Survey	0.72	0.56	1.84	0.88
T2 Survey	0.93	0.54	1.82	0.65

Note:

Quantiles of null distribution and RI P-values from 10,000 permutations of the experimental design.

Table S13: Estimated treatment effects on primary outcome measures

	Index of Primary DVs	Legitimacy	Cooperation	Compliance	Performance
T1 Survey:					
ITT	2.95 (0.42)*	2.15 (0.54)*	2.26 (0.56)*	1.27 (0.64)*	3.20 (0.63)*
ATT	7.10 (0.99)*	5.17 (1.29)*	5.44 (1.34)*	3.06 (1.54)*	7.71 (1.49)*
T2 Survey:					
ITT	1.99 (0.52)*	1.50 (0.65)*	1.49 (0.63)*	0.95 (0.82)	2.08 (0.76)*
ATT	4.86 (1.26)*	3.67 (1.59)*	3.63 (1.55)*	2.26 (2.02)	5.12 (1.87)*

Note:

Covariate-adjusted point estimates with standard errors in parentheses.

* denotes statistical significance at pre-registered level

Table S14: Estimated treatment effects by Race/Ethnicity

	White	Black	Hispanic	Other
T1 Survey:				
ITT	2.17 (0.49)*	5.37 (1.28)*	4.64 (1.34)*	2.71 (1.35)*
ATT	5.95 (1.33)*	11.14 (2.54)*	7.12 (1.96)*	6.95 (3.72)
Pct. Sample	0.61	0.18	0.11	0.10
T2 Survey:				
ITT	1.38 (0.58)*	3.11 (1.60)	4.13 (1.82)*	3.21 (1.99)
ATT	3.84 (1.60)*	6.55 (3.34)	7.18 (3.02)*	7.30 (4.24)
Pct. Sample	0.65	0.16	0.09	0.10

Note:

Covariate-adjusted point estimates with standard errors in parentheses.

* denotes statistical significance at 0.05.

Table S15: Estimated treatment effects on secondary outcome measures

	Confidence in Police	Police Feeling Thermometer	Negative Beliefs about Police	Support Body Cameras	Increase Police by 10%	Trust in City Gov.
T1 Survey:						
ATT	3.80 (0.61)*	3.92 (0.85)*	-4.59 (0.68)*	-1.69 (0.71)*	4.06 (1.07)*	1.10 (0.75)
ATT	9.16 (1.41)*	9.47 (2.02)*	-11.08 (1.62)*	-4.06 (1.72)*	9.78 (2.57)*	2.64 (1.80)
T2 Survey:						
ATT	2.76 (0.73)*	4.04 (1.03)*	-2.98 (0.81)*	-0.80 (0.72)	2.07 (1.42)	NA
ATT	6.77 (1.76)*	9.97 (2.54)*	-7.28 (1.99)*	-1.95 (1.77)	4.97 (3.50)	NA

Note:

Covariate-adjusted point estimates with standard errors in parentheses.

* denotes statistical significance at pre-registered level.

Table S16: Estimated treatment effects by Quantile of Baseline Support

	Q1	Q2	Q3	Q4
T1 Survey:				
ITT	4.62 (1.59)*	2.71 (0.87)*	2.47 (0.82)*	1.91 (0.84)*
ATT	12.60 (4.21)*	7.23 (2.29)*	5.66 (1.87)*	3.94 (1.71)*
Pct. Sample	0.25	0.25	0.24	0.26
T2 Survey:				
ITT	4.70 (1.97)*	0.42 (1.03)	2.23 (1.02)*	0.94 (0.99)
ATT	12.61 (5.18)*	1.40 (2.75)	5.28 (2.47)*	1.84 (2.10)
Pct. Sample	0.24	0.25	0.26	0.25

Note:

Covariate-adjusted point estimates with standard errors in parentheses.

* denotes statistical significance at 0.05.

A Appendix A: Statistical Code

A.1 Functions used for outcome construction

```
# This function subsets a dataframe (df) to respondents (R) and the relevant
# columns (col_names), then scales the columns to have mean 0 (center = TRUE)
# and sd 1 (scale = TRUE) and sums across the rows to create a standardized
# index. It returns a dataframe with this new index added to the dataframe as a
# column with the desired label (new_name).
scale_sum <- function(df = NULL, R = NULL, col_names = NULL, new_name = NULL,
                      center = TRUE, scale = TRUE){
  # Response index
  index <- as.logical(R)
  index[is.na(index)] <- FALSE

  # subset to relevant columns in dataframe
  tmp <- df[index, col_names]

  # apply procedure, create new column vector w/ the appropriate name
  df[[paste(new_name)]] <- NA
  df[index, ][[paste(new_name)]] <-
    as.numeric(
      rowSums(
        scale(tmp, center = center, scale = scale)
      )
    )
  return(df)
}

# This function subsets a dataframe (df) to respondents (R) and the relevant
# columns (col_names), performs a principal components analysis, and extracts
# the first component from the PCA to use as an index. It returns a dataframe
# with this new index added to the dataframe as a column with the desired label
# (new_name).
# PCA is performed using the prcomp() function in base R, which used singular
# value decomposition of the data matrix, which is standardized to have
# mean 0 (center = TRUE) and sd 1 (scale = TRUE) by default, but it allows for
# custom transformations.
# It also allows for a user-specified rotation matrix of variable loadings (Q).
# The default constructs the rotation matrix using the subset of
# columns (col_names) in the dataframe (df).
scale_outcome <- function(df = NULL, R = NULL, Q = NULL, scale = TRUE,
                           center = TRUE, col_names = NULL, new_name = NULL){
  # Response index
```

```

index <- as.logical(R)
index[is.na(index)] <- FALSE

# Subset to desired columns.
X <- df[index, col_names]

# Unless a custom rotation matrix is provided, use prcomp function to
# perform PCA using singular value decomposition of the data matrix
if(is.null(Q))
  Q <- prcomp(X, scale. = scale, center = center)$rotation

# Multiply centered data by the rotation matrix to generate the rotated
# data matrix (P), then extract first column (y) or "principal component"
Z <- scale(as.matrix(X), scale = scale, center = center)
P <- Z %*% Q
y <- as.vector(P[, 1])

# The sign of the columns of the rotation matrix (Q) are arbitrary; this
# ensures more positive values indicate more positive attitudes
if(cor(y, Z[, 1], use = "complete.obs") < 0)
  y <- -y

# Add the index, with the desired label, and export the new dataframe, the
# rotation matrix, and the rotated data matrix
df[[paste(new_name)]] <- NA
df[index, ][[paste(new_name)]] <- y
return(list(df = df, Q = Q, P = P))
}

```

A.2 Randomization inference procedures

A.2.1 Covariate imbalance

```

# Subset to people who answered the baseline survey (and thus were assigned
# to treatment or control)
full_df <-
  cops_df %>%
  filter(R_t0 == 1) %>%
  mutate(dem_black = as.numeric(dem_race4 == "Black"),
         dem_white = as.numeric(dem_race4 == "White"),
         dem_hisp = as.numeric(dem_race4 == "Hispanic"),
         dem_other = as.numeric(dem_race4 == "Other"))

```

```

# Make design declaration
library(randomizr)
declaration <-
  with(full_df, {
    declare_ra(
      blocks = block_id,
      clusters = hh_id)
  })
}

# Obtain permutation matrix for 10000 random assignments based on the
# experimental design. Note: this takes several minutes to finish.
set.seed(1)
perm_mat <- obtain_permutation_matrix(declaration)

# Background covariates for predicting imbalance and differential attrition.
# Includes all variables used for blocking for (household-level) randomization.
X <- model.matrix(~ -1 + primary_dvs_index_t0 + dem_age + dem_female +
  hh_size + dem_inc_high + spanish + dem_pid7 +
  dem_white + dem_black + dem_hisp + dem_other +
  contact_arrest_t0 + contact_unfair_t0 + contact_f2f_t0 +
  spanish + factor(as.numeric(dist_num)),
  data = full_df)

# Binary response vectors for T1 survey and T2 survey
R_t1 <- full_df$R_t1
R_t2 <- full_df$R_t2

# Capture output
null_dist_t0 <- vector()
null_dist_t1 <- vector()
null_dist_t2 <- vector()

# Note: this takes several minutes to finish.
for(i in 1:ncol(perm_mat)){
  null_dist_t0[i] <- summary(lm(perm_mat[, i] ~ X))$f[1]
  null_dist_t1[i] <- summary(lm(perm_mat[, i][R_t1 == 1] ~ X[R_t1 == 1,]))$f[1]
  null_dist_t2[i] <- summary(lm(perm_mat[, i][R_t2 == 1] ~ X[R_t2 == 1,]))$f[1]
}

Z_Obs <- as.numeric(full_df$Z)
obs_f_t0 <- summary(lm(Z_Obs ~ X))$f[1]
obs_f_t1 <- summary(lm(Z_Obs[R_t1 == 1] ~ X[R_t1 == 1,]))$f[1]
obs_f_t2 <- summary(lm(Z_Obs[R_t2 == 1] ~ X[R_t2 == 1,]))$f[1]

# Calculate proportion of simulated statistics as extreme as the observed

```

```
# F-statistic. This is the randomization inference p-value.
ri_pvalue_t0 <- mean(null_dist_t0 >= obs_f_t0)
ri_pvalue_t1 <- mean(null_dist_t1 >= obs_f_t1)
ri_pvalue_t2 <- mean(null_dist_t2 >= obs_f_t2)
```

A.2.2 Sharp null on response behavior

```
# Test the sharp null (no covariates): no effect of Z on R for any unit.
# We use the ri2 package for convenience (https://github.com/acoppock/ri2)
library(ri2)

# RI for first post-treatment wave
ri_sharp_t1 <-
  conduct_ri(
    R_t1 ~ Z,
    declaration = declaration,
    sharp_hypothesis = 0,
    data = full_df,
    permutation_matrix = perm_mat,
    progress_bar = TRUE
  )

# Recode non-response from NA to 0 in order to avoid ri2 warning
full_df$R_t2[is.na(full_df$R_t2)] <- 0

# RI for second post-treatment wave
ri_sharp_t2 <-
  conduct_ri(
    R_t2 ~ Z,
    declaration = declaration,
    sharp_hypothesis = 0,
    data = full_df,
    permutation_matrix = perm_mat,
    progress_bar = TRUE
  )
```

A.2.3 Treatment-covariate interactions on response behavior

```
# Capture output
null_dist_t1 <- vector()
null_dist_t2 <- vector()
```

```

# Loop through permutation matrix and compute F-Statistic for each possible
# permutation. Note: this takes several minutes to finish.
for(i in 1:ncol(perm_mat)){
  fit_restricted_t1 <- lm(R_t1 ~ perm_mat[, i] + X)
  fit_unrestricted_t1 <- lm(R_t1 ~ perm_mat[, i]*X)
  null_dist_t1[i] <- anova(fit_restricted_t1, fit_unrestricted_t1)$F[2]

  fit_restricted_t2 <- lm(R_t2 ~ perm_mat[, i] + X)
  fit_unrestricted_t2 <- lm(R_t2 ~ perm_mat[, i]*X)
  null_dist_t2[i] <- anova(fit_restricted_t2, fit_unrestricted_t2)$F[2]
}

obs_f_t1 <- anova(lm(R_t1 ~ full_df$Z + X), lm(R_t1 ~ full_df$Z*X))$F[2]
obs_f_t2 <- anova(lm(R_t2 ~ full_df$Z + X), lm(R_t2 ~ full_df$Z*X))$F[2]

# Calculate proportion of simulated statistics as extreme as the observed
# F-statistic. This is the randomization inference p-value.
ri_pvalue_t1 <- mean(null_dist_t1 >= obs_f_t1)
ri_pvalue_t2 <- mean(null_dist_t2 >= obs_f_t2)

```

A.3 Effects on primary/secondary outcomes

A.3.1 Primary outcomes

```

# Package to implement estimators
library(estimatr)

primary_dvs <- c("coop_index", "comply_index", "legit_index",
                 "perform_index", "primary_dvs_index")

covariates <- c("dem_age", "dem_female", "dem_race4", "dem_pid7",
                 "dem_inc_high", "contact_arrest_t0", "contact_unfair_t0",
                 "contact_f2f_t0", "spanish")

# First Wave
gg_itt <- list()
gg_att <- list()

for(i in 1:length(primary_dvs)) {

  Y <- paste0(primary_dvs[i], "_t1")
  X <- c(paste0(primary_dvs[i], "_t0"), covariates)

```

```

gg_itt[[i]] <-
  lm_lin(as.formula(paste0(Y, " ~ Z")),
         covariates = as.formula(paste("~", paste(X, collapse = "+"))),
         clusters = hh_id, se_type = "stata", alpha = 0.08,
         data = cops_df) %>%
  tidy() %>%
  mutate_if(is.numeric, round, 3) %>%
  filter(term %in% c("(Intercept)", "Z")) %>%
  mutate(estimator = "OLS",
        term = ifelse(term == "(Intercept)", "Control", "Treatment"))

gg_att[[i]] <-
  iv_robust(as.formula(paste(Y, "~ D + ", paste(X, collapse = "+"),
                             "| Z + ", paste(X, collapse = "+"))),
             clusters = hh_id, se_type = "stata", alpha = 0.08,
             data = cops_df) %>%
  tidy() %>%
  mutate_if(is.numeric, round, 3) %>%
  filter(term %in% c("(Intercept)", "D")) %>%
  mutate(estimator = "IV",
        term = ifelse(term == "(Intercept)", "Control", "Treatment"))

}

est_primary_t1 <-
  bind_rows(
    (gg_itt %>% bind_rows()),
    (gg_att %>% bind_rows()) %>%
  filter(term != "Control") %>%
  mutate(outcome = factor(outcome, levels = rev(paste0(primary_dvs, "_t1"))),
         labels = rev(c("Cooperation", "Compliance",
                       "Legitimacy", "Performance",
                       "Index of Primary DVs"))),
  estimator = factor(estimator, levels = c("OLS", "IV")),
  Wave = "T1 Survey")

# Second Wave
gg_itt <- list()
gg_att <- list()

for(i in 1:length(primary_dvs)) {

  Y <- paste0(primary_dvs[i], "_t2")
  X <- c(paste0(primary_dvs[i], "_t0"), covariates)
}

```

```

gg_itt[[i]] <-
  lm_lin(as.formula(paste0(Y, " ~ Z")),
         covariates = as.formula(paste("~", paste(X, collapse = "+"))),
         clusters = hh_id, se_type = "stata", alpha = 0.08,
         data = cops_df) %>%
  tidy() %>%
  mutate_if(is.numeric, round, 3) %>%
  filter(term %in% c("(Intercept)", "Z")) %>%
  mutate(estimator = "OLS",
        term = ifelse(term == "(Intercept)", "Control", "Treatment"))

gg_att[[i]] <-
  iv_robust(as.formula(paste(Y, "~ D + ", paste(X, collapse = "+"),
                             "| Z + ", paste(X, collapse = "+"))),
             clusters = hh_id, se_type = "stata", alpha = 0.08,
             data = cops_df) %>%
  tidy() %>%
  mutate_if(is.numeric, round, 3) %>%
  filter(term %in% c("(Intercept)", "D")) %>%
  mutate(estimator = "IV",
        term = ifelse(term == "(Intercept)", "Control", "Treatment"))

}

est_primary_t2 <-
  bind_rows(
    gg_itt %>% bind_rows(),
    gg_att %>% bind_rows()) %>%
  filter(term != "Control") %>%
  mutate(outcome = factor(outcome, levels = rev(paste0(primary_dvs, "_t2"))),
         labels = rev(c("Cooperation", "Compliance",
                       "Legitimacy", "Performance",
                       "Index of Primary DVs"))),
  estimator = factor(estimator, levels = c("OLS", "IV")),
  Wave = "T2 Survey")

# Combine for presentation in Tables/Figures
est_primary <-
  bind_rows(est_primary_t1, est_primary_t2)

```

A.3.2 Secondary outcomes

```

# Package to implement estimators
library(estimatr)

primary_dvs <- c("coop_index", "comply_index", "legit_index",
                 "perform_index", "primary_dvs_index")

covariates <- c("dem_age", "dem_female", "dem_race4", "dem_pid7",
                 "dem_inc_high", "contact_arrest_t0", "contact_unfair_t0",
                 "contact_f2f_t0", "spanish")

# First Wave
gg_itt <- list()
gg_att <- list()

for(i in 1:length(primary_dvs)) {

  Y <- paste0(primary_dvs[i], "_t1")
  X <- c(paste0(primary_dvs[i], "_t0"), covariates)

  gg_itt[[i]] <-
    lm_lin(as.formula(paste0(Y, " ~ Z")),
           covariates = as.formula(paste("~", paste(X, collapse = "+"))),
           clusters = hh_id, se_type = "stata", alpha = 0.08,
           data = cops_df) %>%
    tidy() %>%
    mutate_if(is.numeric, round, 3) %>%
    filter(term %in% c("(Intercept)", "Z")) %>%
    mutate(estimator = "OLS",
          term = ifelse(term == "(Intercept)", "Control", "Treatment"))

  gg_att[[i]] <-
    iv_robust(as.formula(paste(Y, "~ D + ", paste(X, collapse = "+"),
                               "| Z + ", paste(X, collapse = "+"))),
               clusters = hh_id, se_type = "stata", alpha = 0.08,
               data = cops_df) %>%
    tidy() %>%
    mutate_if(is.numeric, round, 3) %>%
    filter(term %in% c("(Intercept)", "D")) %>%
    mutate(estimator = "IV",
          term = ifelse(term == "(Intercept)", "Control", "Treatment"))

}

est_primary_t1 <-
  bind_rows(

```

```

(gg_itt %>% bind_rows(),
 (gg_att %>% bind_rows())) %>%
filter(term != "Control") %>%
mutate(outcome = factor(outcome, levels = rev(paste0(primary_dvs, "_t1"))),
       labels = rev(c("Cooperation", "Compliance",
                     "Legitimacy", "Performance",
                     "Index of Primary DVs"))),
estimator = factor(estimator, levels = c("OLS", "IV")),
Wave = "T1 Survey")

# Second Wave
gg_itt <- list()
gg_att <- list()

for(i in 1:length(primary_dvs)) {

  Y <- paste0(primary_dvs[i], "_t2")
  X <- c(paste0(primary_dvs[i], "_t0"), covariates)

  gg_itt[[i]] <-
    lm_lin(as.formula(paste0(Y, " ~ Z")),
           covariates = as.formula(paste("~", paste(X, collapse = "+"))),
           clusters = hh_id, se_type = "stata", alpha = 0.08,
           data = cops_df) %>%
    tidy() %>%
    mutate_if(is.numeric, round, 3) %>%
    filter(term %in% c("(Intercept)", "Z")) %>%
    mutate(estimator = "OLS",
           term = ifelse(term == "(Intercept)", "Control", "Treatment"))

  gg_att[[i]] <-
    iv_robust(as.formula(paste(Y, "~ D + ", paste(X, collapse = "+"),
                                "| Z + ", paste(X, collapse = "+"))),
               clusters = hh_id, se_type = "stata", alpha = 0.08,
               data = cops_df) %>%
    tidy() %>%
    mutate_if(is.numeric, round, 3) %>%
    filter(term %in% c("(Intercept)", "D")) %>%
    mutate(estimator = "IV",
           term = ifelse(term == "(Intercept)", "Control", "Treatment"))

}

est_primary_t2 <-
bind_rows(

```

```

(gg_itt %>% bind_rows(),
 (gg_att %>% bind_rows())) %>%
filter(term != "Control") %>%
mutate(outcome = factor(outcome, levels = rev(paste0(primary_dvs, "_t2"))),
       labels = rev(c("Cooperation", "Compliance",
                     "Legitimacy", "Performance",
                     "Index of Primary DVs"))),
estimator = factor(estimator, levels = c("OLS", "IV")),
Wave = "T2 Survey")

# Combine for presentation in Tables/Figures
est_primary <-
bind_rows(est_primary_t1, est_primary_t2)

```

A.4 Heterogenous treatment effects

A.4.1 Race/Ethnicity

```

# Pre-registered sub-group analysis for differences by race/ethnicity
race_itt <- list()
race_att <- list()

race <- c("White", "Black", "Hispanic", "Other")

for(i in 1:length(race)){
  race_itt[[i]] <-
    lm_lin(primary_dvs_index_t1 ~ Z, covariates = ~ primary_dvs_index_t0 +
      dem_age + dem_female + dem_pid7 + dem_inc_high +
      contact_arrest_t0 + contact_unfair_t0 + contact_f2f_t0 +
      spanish, clusters = hh_id, se_type = "stata",
      data = cops_df, subset = dem_race4 == race[i]) %>%
    tidy() %>%
    mutate_if(is.numeric, round, 3) %>%
    filter(term %in% c("(Intercept)", "Z")) %>%
    mutate(estimator = "OLS",
           term = ifelse(term == "(Intercept)", "Control", "Treatment"),
           race = race[i])

  race_att[[i]] <-
    iv_robust(primary_dvs_index_t1 ~ D + primary_dvs_index_t0 + dem_age +
      dem_female + dem_pid7 + dem_inc_high + contact_arrest_t0 +
      contact_unfair_t0 + contact_f2f_t0 + spanish | Z +
      primary_dvs_index_t0 + dem_age + dem_female + dem_pid7 +

```

```

    dem_inc_high + contact_arrest_t0 + contact_unfair_t0 +
    contact_f2f_t0 + spanish, clusters = hh_id, se_type = "stata",
    data = cops_df, subset = dem_race4 == race[i]) %>%
tidy() %>%
mutate_if(is.numeric, round, 3) %>%
filter(term %in% c("(Intercept)", "D")) %>%
mutate(estimator = "IV",
      term = ifelse(term == "(Intercept)", "Control", "Treatment"),
      race = race[i])
}

est_race_t1 <-
bind_rows(
  (race_itt %>% bind_rows()),
  (race_att %>% bind_rows())) %>%
filter(term != "Control") %>%
mutate(outcome = "Index of Primary DVs",
      estimator = factor(estimator, levels = c("OLS", "IV")),
      Wave = "T1 Survey")

# Second Wave
race_itt <- list()
race_att <- list()

race <- c("White", "Black", "Hispanic", "Other")

for(i in 1:length(race)){
  race_itt[[i]] <-
    lm_robust(primary_dvs_index_t2 ~ Z + primary_dvs_index_t0 +
               dem_age + dem_female + dem_pid7 + dem_inc_high +
               contact_arrest_t0 + contact_unfair_t0 + contact_f2f_t0 +
               spanish, clusters = hh_id, se_type = "stata",
               data = cops_df, subset = dem_race4 == race[i]) %>%
tidy() %>%
mutate_if(is.numeric, round, 3) %>%
filter(term %in% c("(Intercept)", "Z")) %>%
mutate(estimator = "OLS",
      term = ifelse(term == "(Intercept)", "Control", "Treatment"),
      race = race[i])

  race_att[[i]] <-
    iv_robust(primary_dvs_index_t2 ~ D + primary_dvs_index_t0 + dem_age +
               dem_female + dem_pid7 + dem_inc_high + contact_arrest_t0 +
               contact_unfair_t0 + contact_f2f_t0 + spanish | Z +
               primary_dvs_index_t0 + dem_age + dem_female + dem_pid7 +

```

```

        dem_inc_high + contact_arrest_t0 + contact_unfair_t0 +
        contact_f2f_t0 + spanish, clusters = hh_id, se_type = "stata",
        data = cops_df, subset = dem_race4 == race[i]) %>%
tidy() %>%
mutate_if(is.numeric, round, 3) %>%
filter(term %in% c("(Intercept)", "D")) %>%
mutate(estimator = "IV",
       term = ifelse(term == "(Intercept)", "Control", "Treatment"),
       race = race[i])
}

est_race_t2 <-
bind_rows(
  (race_it >%> bind_rows()),
  (race_att >%> bind_rows())) %>%
filter(term != "Control") %>%
mutate(outcome = "Index of Primary DVs",
       estimator = factor(estimator, levels = c("OLS", "IV")),
       Wave = "T2 Survey")

# Combine estimates from each wave
est_race <-
bind_rows(est_race_t1, est_race_t2)

```

A.4.2 Quantiles of Baseline Support

```

# Estimate covariate-adjusted ITT/ATT by quantiles of baseline support
cops_df <-
  cops_df %>%
  mutate(quantcut = cut(primary_dvs_index_t0,
                        quantile(primary_dvs_index_t0, na.rm = T)))

quant_itt <- list()
quant_att <- list()

quant <- levels(cops_df$quantcut)

for(i in 1:length(quant)){
  quant_itt[[i]] <-
    lm_lin(primary_dvs_index_t1 ~ Z, covariates = ~ dem_race4 +
           dem_age + dem_female + dem_pid7 + dem_inc_high +

```

```

    contact_arrest_t0 + contact_unfair_t0 + contact_f2f_t0 +
    spanish, clusters = hh_id, se_type = "stata",
    data = cops_df, subset = quantcut == quant[i]) %>%
tidy() %>%
mutate_if(is.numeric, round, 3) %>%
filter(term %in% c("(Intercept)", "Z")) %>%
mutate(estimator = "OLS",
      term = ifelse(term == "(Intercept)", "Control", "Treatment"),
      quantcut = quant[i])

quant_att[[i]] <-
iv_robust(primary_dvs_index_t1 ~ D + dem_race4 + dem_age +
           dem_female + dem_pid7 + dem_inc_high + contact_arrest_t0 +
           contact_unfair_t0 + contact_f2f_t0 + spanish | Z +
           dem_race4 + dem_age + dem_female + dem_pid7 +
           dem_inc_high + contact_arrest_t0 + contact_unfair_t0 +
           contact_f2f_t0 + spanish, clusters = hh_id, se_type = "stata",
           data = cops_df, subset = quantcut == quant[i]) %>%
tidy() %>%
mutate_if(is.numeric, round, 3) %>%
filter(term %in% c("(Intercept)", "D")) %>%
mutate(estimator = "IV",
      term = ifelse(term == "(Intercept)", "Control", "Treatment"),
      quantcut = quant[i])
}

est_quantcut_t1 <-
bind_rows(
  (quant_itt %>% bind_rows()),
  (quant_att %>% bind_rows())) %>%
filter(term != "Control") %>%
mutate(outcome = "Index of Primary DVs",
       estimator = factor(estimator, levels = c("OLS", "IV")),
       Wave = "T1 Survey")

# Repeat for Second Wave Index
quant_itt <- list()
quant_att <- list()

for(i in 1:length(quant)){
  quant_itt[[i]] <-
    lm_lin(primary_dvs_index_t2 ~ Z, covariates = ~ dem_race4 +
           dem_age + dem_female + dem_pid7 + dem_inc_high +
           contact_arrest_t0 + contact_unfair_t0 + contact_f2f_t0 +
           spanish, clusters = hh_id, se_type = "stata",

```

```

    data = cops_df, subset = quantcut == quant[i]) %>%
tidy() %>%
mutate_if(is.numeric, round, 3) %>%
filter(term %in% c("(Intercept)", "Z")) %>%
mutate(estimator = "OLS",
      term = ifelse(term == "(Intercept)", "Control", "Treatment"),
      quantcut = quant[i])

quant_att[[i]] <-
iv_robust(primary_dvs_index_t2 ~ D + dem_race4 + dem_age +
           dem_female + dem_pid7 + dem_inc_high + contact_arrest_t0 +
           contact_unfair_t0 + contact_f2f_t0 + spanish | Z +
           dem_race4 + dem_age + dem_female + dem_pid7 +
           dem_inc_high + contact_arrest_t0 + contact_unfair_t0 +
           contact_f2f_t0 + spanish, clusters = hh_id, se_type = "stata",
           data = cops_df, subset = quantcut == quant[i]) %>%
tidy() %>%
mutate_if(is.numeric, round, 3) %>%
filter(term %in% c("(Intercept)", "D")) %>%
mutate(estimator = "IV",
      term = ifelse(term == "(Intercept)", "Control", "Treatment"),
      quantcut = quant[i])
}

est_quantcut_t2 <-
bind_rows(
  (quant_itt %>% bind_rows()),
  (quant_att %>% bind_rows())) %>%
filter(term != "Control") %>%
mutate(outcome = "Index of Primary DVs",
      estimator = factor(estimator, levels = c("OLS", "IV")),
      Wave = "T2 Survey")

# Combine estimates from each wave
est_quantcut <-
bind_rows(est_quantcut_t1, est_quantcut_t2)

```

B Appendix B: Pre-Analysis Plan and Supplements

Pre-Analysis Plan for September 2018 COPS Experiment



2018-09-15

Contents

Purpose	2
Background	2
Study outcomes	3
Primary outcomes	3
Cooperation	3
Compliance	3
Performance	3
Legitimacy	4
Index of all Primary Outcomes	4
Secondary outcomes	4
Index of Community Support	4
Police Feeling Thermometer	5
Police stereotypes	5
Benchmarking questions	5
Police policy questions	5
Trust in local government	6
Randomization	7
Estimation and Inference	7
Asymmetric hypothesis testing	8
Other	8
Missing values	8
Ceiling effects	8
Treatment effect heterogeneity	8
References	10



Purpose

Community-oriented policing strategies (COPS) have been heralded as an important public policy tool for building cooperative relationships between police officers and civilians to effectively solve community problems¹. COPS are believed to operate in large part by improving legitimacy: the belief that police have the authority to dictate proper behavior, and that the public should cooperate with police in their efforts to enforce the law². Civilians who view police as legitimate authorities tend to provide more positive evaluations of police performance, and are more willing to cooperate and comply with the law^{3,4}.

The case for COPS is largely built on case studies, cross-sectional surveys, and observational research designs⁴⁻⁶. At their core, COPS are designed to facilitate non-enforcement related contact between civilians and police officers. This seems especially promising given that contact is one of the most powerful tools for improving intergroup relations in general^{7,8}. However, the lack of experimental evidence on the effectiveness of COPS limits the ability of police departments in the United States to identify actionable strategies, and to convince skeptics about the value of COPS⁹.

This document describes a pre-analysis plan (PAP) for a randomized controlled trial designed to examine the effect of a COPS program developed in partnership with the [REDACTED]. The objective of this document is to foster research transparency by detailing the analyses that will be performed to examine the effectiveness of this program before any outcome data have been analyzed. We pre-commit to reporting all the analyses specified below. Any additional analyses that may be conducted – for example, to satisfy requests from reviewers as part of the academic peer-review process – will be clearly flagged as analyses that were not pre-registered.

Background

This experiment uses a mail to online panel design to measure outcomes¹⁰. We started by recruiting subjects in [REDACTED] to participate in an online panel survey called the “[REDACTED] Opinion Study 2018”. On August 15th 2018, approximately 50,000 [REDACTED] residents were sent an invitation to the mailing address listed in their voter registration record.

The mailers directed recipients to an online survey via a landing page at [www.\[REDACTED\]opinion.com](http://www.[REDACTED]opinion.com). Each respondent was provided with a unique login code in their recruitment letter, which they used to access the survey. The survey was available in English and Spanish. A dedicated phone number and email address were created to field respondent inquiries during the recruitment period, which ended on September 6th 2018.

At the time of this writing, 2,013 individuals nested in 1,852 households have completed the survey and provided their contact details to participate in followup surveys. On 9 September 2018 we randomly assigned 926 households (1,007 individuals) to receive treatment and 926 households (1,006 individuals) to receive control. Block random assignment took place at the household level using matched quartets (more details below).

Our partner organization – [REDACTED] – will canvass all 926 households in the treatment group to deliver a short (approximately 10 minute) conversation, and provide their contact information via department issued business cards. Canvassers will be assisted by Polis, a canvassing app that has been installed on the department issued cell phones of participating officers.

The [REDACTED] has selected 20 patrol officers from the B-Squad (shifts between 3.00pm to 11.00pm or 4.00pm to 12.00am) to participate in the study. Home visits will take place during the 6.00-8.00pm time frame during the weekdays, and prior to 8.00pm on the weekends. These time slots were selected to maximize the number of successful conversations, subject to [REDACTED] resource constraints.

The speed at which the canvass can be completed will largely depend on incoming calls for service, which must take priority over canvassing. The month of September was selected based on historical analyses of the [REDACTED] administrative data, which revealed it is almost always the least busy month as measured by calls for service. We anticipate canvassing to be completed within 10-20 days.

Study outcomes

In a survey experiment conducted on Amazon Mechanical Turk (MTurk) in March 2017, we measured civilian attitudes toward the police with a variety of widely-used survey measures^{4,11,12}. We then used factor analysis to select 19 questions, from a larger universe of 26, that tap four key dimensions: legitimacy, cooperation, compliance, and perceived effectiveness. For analysis purposes, we recode responses so that higher values indicate more positive attitudes toward the police. These four primary outcome measures are each based on a series of individual survey questions (detailed below). Secondary outcome measures are also described below.

Primary outcomes

Each of the following measures: **cooperation**, **compliance**, **performance**, **legitimacy** are constructed by combining multiple individual items to create a scale. Scale construction will be performed by standardizing the individual items (to mean 0 and standard deviation 1) associated with each domain and then summing them to create a single outcome measure for each of the four dimensions.

Cooperation

- How likely are you to provide information to the police?
- How likely are you to call the police to report a crime?
- How likely are you to report suspicious activity to the police?

For the field experiment, an additional question was added to the cooperation battery¹¹:

- How likely would you be to attend a community meeting to discuss problems in your neighborhood with the police?

All items are recorded on a 7-point scale from Extremely Unlikely to Extremely Likely with a neutral midpoint.

Compliance

- If the police tell you to do something, you should do it.
- There are times when it is ok to ignore what the police tell you to do (reverse coded)
- It is your duty to accept the decisions made by the police even when you disagree with them.
- It is your duty to do what the police tell you to do, even if you do not like the way they treat you.

All items are recorded on a 7-point scale from Strongly Disagree to Strongly Agree with a neutral midpoint.

Performance

- I have confidence that the police in [REDACTED] can do their job well.
- The police in [REDACTED] are effective at fighting crime.
- The police in [REDACTED] are effective at helping people who ask for help.
- When people in your neighborhood call the police for help, they respond quickly.

All items are recorded on a 7-point scale from Strongly Disagree to Strongly Agree with a neutral midpoint.

Legitimacy

- They care about the well-being of everyone they deal with
- They make fair and impartial decisions
- They act in ways consistent with your own ideas about what is right and wrong
- They have the same sense of right and wrong that you do
- They stand up for values that are important to you
- They behave according to the law when dealing with people
- They make decisions based on their biases or opinions (reverse coded)
- They are often dishonest (reverse coded)

All items are recorded on a 7-point scale from Strongly Disagree to Strongly Agree with a neutral midpoint.

Index of all Primary Outcomes

Although all the individual items described above are components of separate conceptual dimensions, they are all positively correlated. We will also examine the overall effect of the treatment on these attitudes toward the police by creating a single dependent variable constructed by extracting the first component from a principal components analysis (PCA) on all the primary outcome measures described above.

Secondary outcomes

None of the secondary outcomes listed here were part of the pilot study we conducted on MTurk. Unlike the primary outcome measures specified above, none of these measures are part of an established research literature.

All outcomes are coded so that higher values correspond to more positive evaluations of the police. We anticipate a positive effect on all secondary outcomes listed below, except support for body worn cameras. We have no expectations to pre-register about how treatment will affect support for body cameras.

Index of Community Support

These questions were developed from qualitative fieldwork and interviews with current and former officers in the [REDACTED] and are not part of any established research literature:

- The police make me feel safer in my neighborhood.
- The police are trying to make my community better.
- The police respect the people in my community.
- The police treat all people equally.
- The police are a part of my community.

All question prompts ask for specific evaluations about the local police department – [REDACTED]. Responses are recorded on a 7-point scale from “Strongly Disagree” to “Strongly Agree”. Responses to these questions were all highly correlated in the baseline survey ($\alpha = 0.90$). To increase precision they will all be combined into a single scale by standardizing the individual items (to mean 0 and standard deviation 1) and summing them.

Police Feeling Thermometer

The question prompts associated with all the primary and secondary outcome measures described above ask respondents to provide specific evaluations about their local police department, the [REDACTED]. As a measure of general affect toward “police”, we also included a 100 point “feeling thermometer” toward police embedded among various other groups. The feeling thermometer is a purely affective measure that appears in various public opinion surveys across time. We will analyze the feeling thermometer separately.

Police stereotypes

The treatment is a counter-stereotypical interaction with a police officer. We therefore measure stereotypes about police on the following dimensions using a 7 point scale where 4 is a neutral midpoint:

- Hardworking (1) to Lazy (7)
- Humble (1) to Arrogant (7)
- Intelligent (1) to Unintelligent (7)
- Trustworthy (1) to Untrustworthy (7)
- Compassionate (1) to Cold-hearted (7)

In particular, treated subjects should revise their beliefs that police are humble (as opposed to arrogant), compassionate (as opposed to cold), etc. These five stereotype dimensions will be combined into a single scale. Scale construction will be performed by standardizing the individual items (to mean 0 and standard deviation 1) associated with each domain and then summing them to create a single outcome measure.

As with the feeling thermometer, the stereotype questions were embedded into a battery with other groups. These other groups are: firefighters, politicians, doctors, school teachers, and religious leaders.

Benchmarking questions

We also included two questions about police that can be benchmarked to nationally representative time-series polling data:

- Do you agree or disagree with this statement: "I respect the police in my community." [7-point scale from "Strongly Disagree" to "Strongly Agree"]
- How much confidence do you have in each of the following institutions in your community? [The police; the city government; the courts; the public schools. (4) A lot; (3) Some; (2) Only a little; (1) None at all]

These items will be analyzed separately.

Police policy questions

Finally, we include two additional outcomes of specific policy support:

- As you may know, the city of [REDACTED] has recently required police to wear body cameras that record their activities while on duty. Do you support or oppose police body cameras?
- How would you feel about increasing the number of police on the street by 10 percent, even if it means fewer funds for other public services?

Both items are recorded on a 7-point scale from Strongly Oppose to Strongly Support, with a neutral midpoint. The question about body cameras was included out of general interest, and the final question was added as an additional measure of support for police¹³. These items will be analyzed separately. We speculate that

treatment may have a small positive effect on support for increasing the number of police on the street. We have no expectations to pre-register about support for police body cameras.

Trust in local government

We will also investigate whether treatment (an interaction with an agent of local government) increases trust in local government more broadly using the trust in government questions that have appeared in the American National Election Studies (ANES) survey since the 1960s. We recode values so that higher scores correspond to more positive evaluations of local government. We speculate that treatment may have a small positive effect on trust in local government.

- How often do you think you can trust the government [REDACTED] to do what is right? [Always (5); Most of the time (4); About half the time (3); Sometimes (2); Never (1)]
- Do you think that people in government [REDACTED] waste much of the money we pay in taxes? [Waste a lot (1); Waste some (2); Don't waste very much (3)]
- How many of the people running the government [REDACTED] are corrupt? [All (1); Most (2); About half (3); A few (4); None (5)]
- Would you say the government [REDACTED] is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people? [Run by a few big interests (0); For the benefit of all the people (1)]

We will create a single trust in local government measure by extracting the first component from a principal components analysis (PCA) on the four individual measures listed above.

Randomization

Randomization was conducted at the household level with the assistance of the `blockTools` and `randomizr` packages in R^{14,15}. Matched quartets of households were formed using the following blocking variables derived from a combination of administrative voterfile data and responses to the baseline survey:

- The proportion of residents above the median income category of \$50,000-59,999.
- The average age in the household
- The proportion of females in the household
- The proportion of residents in each of 4 major race/ethnicity categories (black, white, hispanic, or other)
- The household average on the first factor from a PCA on all the items underlying the 4 primary outcome measures: cooperation, legitimacy, compliance, and performance
- The proportion of residents who indicated any prior arrest by the [REDACTED]
- The proportion of residents who indicated having any face to face contact with the [REDACTED] in the last 12 months
- The proportion of residents who indicated any unfair treatment by the [REDACTED] in the past
- The proportion of spanish speaking residents in the household
- The household size
- The police district

1,852 households (clusters) of 2,013 individuals were randomly assigned to treatment or control across 463 blocks (quartets of households) using complete random assignment. The probabilities of assignment were constant across households. 926 households (1,007 individuals) were assigned treatment and 926 households (1,006 individuals) were assigned control.

Estimation and Inference

We anticipate fielding two waves of post-treatment surveys for a total of three waves of survey data. The primary goal of administering the second post-treatment survey is to examine the persistence of any attitude change induced by the police visit. As such, all subsequent surveys will include all outcome measures from the first wave.

Within three days of each visit, treated households will be invited to participate in the first followup survey, along with the control households in their corresponding quartet. Approximately 10 days later, they will be invited to participate in a second followup survey. We will file revised PAPs for the follow-up surveys if additional components are added that substantively alter the survey content.

Each home will receive one visit. Consistent with prior research on canvassing field experiments, we anticipate that most of the attempts at making contact will be unsuccessful (residents not home, unavailable, etc.). We therefore anticipate a high level of one sided non-compliance.

The [REDACTED] does not know the address of any household in the control condition, and we do not anticipate that any control households will be contacted as part of this program. We deemed a placebo canvass condition by the [REDACTED] (e.g. providing a non-COPS oriented treatment) as unrealistic, and believe the correct counterfactual is no contact.

Our main target parameters (for each outcome measure) are the intent-to-treat effect (ITT) and the complier average causal effect (CACE). We will use linear regression of the outcome on treatment assignment to estimate the ITTs, and instrumental variables regression to estimate the CACEs. Compliance will be measured

using the Polis app. We define “compliers” as the households that were 1) assigned to treatment; 2) visited by the [REDACTED]; and 3) direct human contact was made.

Regression estimators will be adjusted for pre-treatment values of the outcomes, as well as the pre-treatment covariates age, sex, race, and partisanship, and indicators for: income above the median category, any prior arrest by the [REDACTED], any face-to-face contact with [REDACTED] in the last 12 months, any reported unfair treatment by [REDACTED] in the past, and spanish speaking.

Standard errors will be clustered at the household level in all regression specifications, using a robust cluster variance estimator with a finite sample correction. This is often called the “Stata” version of the robust cluster variance estimator because it’s the version used by Stata software¹⁶.

Asymmetric hypothesis testing

For all ITT and CACE estimates, we will reject the null hypothesis if the estimate is in the bottom 1%, or the top 4%, of the sampling distribution. Specifically, if the point estimate is positive we will use a one-tailed p -value with a rejection threshold of 0.04 (0.08 two-tailed). If the estimate is negative we will use a one-tailed p -value with a rejection threshold of 0.01 (or 0.02 two-tailed). This preserves the property that the rejection region contains 5% of the sampling distribution. However, we focus on the bottom 1% and the top 4% instead of the typical bottom 2.5% and top 2.5%^{17,18}.

The only exception is the question about police body cameras. We do not have a prior expectation about the direction of any treatment effect on support for police body cameras, and we will use the conventional symmetric hypothesis test with $\alpha = 0.05$.

Other

Missing values

Missing values for demographic variables are rare, and are recoded to their means. This was only applied to a single unit in the first wave survey data, who did not provide an age and did not have their date of birth listed in the voter file.

For skipped questions (these are also rare) involving Likert-type responses we recode missing values as the neutral midpoint of “neither agree or disagree”. For feeling thermometer questions and questions about stereotypes, we apply the same rule and recoded skipped questions to their neutral midpoint. These items are both measured using visual analog scales that are presented to respondents with the neutral starting position (4 for the stereotype items and 50 for the feeling thermometer items).

Ceiling effects

Given the generally positive responses registered on our outcome measures in the baseline survey, we believe that “ceiling effects” are likely. For some respondents, for example, it’s simply impossible to record an increase in the perceived legitimacy or willingness to cooperate with police using the current survey instruments.

Treatment effect heterogeneity

We will examine treatment effect heterogeneity by 1) the quantiles of baseline support recorded in the pre-treatment survey and 2) the 4-category race variable (black, hispanic, white, other) with white as the reference group.

We may further investigate heterogenous treatment effects using Bayesian Additive Regression Trees (BART). However, given the anticipated constraints on statistical power, we will not present any additional heterogeneous effects analyses for specific sub-groups unless specifically required to do so by a reviewer, in which case we will note the deviation from the pre-analysis plan.

References

1. Century Policing. TF on 21st. Final report of the president's task force on 21st century policing. *Washington, DC: Office of Community Oriented Policing Services.* 2015.
2. Tyler TR. Enhancing police legitimacy. *The annals of the American academy of political and social science.* 2004;593(1):84-99.
3. Jackson J, Bradford B, Hough M, Myhill A, Quinton P, Tyler TR. Why do people comply with the law? Legitimacy and the influence of legal institutions. *British journal of criminology.* 2012.
4. Tyler TR, Jackson J. Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. *Psychology, Public Policy, and Law.* 2013;20(1):78.
5. Fagan J, Tyler TR. Legal socialization of children and adolescents. *Social justice research.* 2005;18(3):217-241.
6. Gill C, Weisburd D, Telep CW, Vitter Z, Bennett T. Community-oriented policing to reduce crime, disorder and fear and increase satisfaction and legitimacy among citizens: A systematic review. *Journal of Experimental Criminology.* 2014;10(4):399-428.
7. Pettigrew TF, Tropp LR. A meta-analytic test of intergroup contact theory. *Journal of personality and social psychology.* 2006;90(5):751.
8. Paluck EL, Green S, Green DP. The contact hypothesis revisited. 2017.
9. National Academies of Sciences Engineering, Medicine. *Proactive Policing: Effects on Crime and Communities.* National Academies Press; 2018.
10. Broockman D, Kalla J, Sekhon J. The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Political Analysis.* 2017;25(4):435-464.
11. Sunshine J, Tyler TR. The role of procedural justice and legitimacy in shaping public support for policing. *Law & society review.* 2003;37(3):513-548.
12. Tyler TR, Fagan J, Geller A. Street stops and police legitimacy: Teachable moments in young urban men's legal socialization. *Journal of Empirical Legal Studies.* 2014;11(4):751-785.
13. Mummolo J. Militarization fails to enhance police safety or reduce crime but may harm police reputation. *Proceedings of the National Academy of Sciences.* 2018.
14. Moore RT, Schnakenberg K. BlockTools: Blocking, assignment, and diagnosing interference in randomized experiments. *Version 05-6, August.* 2012.
15. Coppock A. *Randomizr: Easy-to-Use Tools for Common Forms of Random Assignment and Sampling.*; 2018. <https://CRAN.R-project.org/package=randomizr>.
16. Imbens GW, Kolesar M. Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics.* 2016;98(4):701-712.
17. Broockman D, Kalla J. Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science.* 2016;352(6282):220-224.
18. Olken BA. Promises and perils of pre-analysis plans. *Journal of Economic Perspectives.* 2015;29(3):61-80.

Supplement to ‘Pre-Analysis Plan for September 2018 COPS Experiment’



2018-10-02

Contents

Update on Progress	2
Revision to Reinterview Procedures	2
Revision to Secondary Outcomes	3

Update on Progress

This document is a supplement to the pre-analysis plan (PAP) filed on 15 September 2018. The study is still ongoing: neither treatment delivery or outcome measurement is complete. Although we anticipated treatment delivery to finish within 10-20 days, the project has been delayed due to several unexpected events. First, call volume has been unusually high and several Priority One calls for service have occurred in the city during the time periods that officers were scheduled to knock on doors. For example, a shooting on 22 September interrupted a large weekend canvass that was scheduled as several of the participating officers were first responders to this incident.

Second, severe thunderstorms and flooding prevented scheduled knocks on several days during the first two weeks of the study. It has rained on 12 of the 22 days since the study has been in the field, with at least one severe thunderstorm occurring each week.

Third, several officers who planned to participate in the study have been unable to participate for a variety of reasons. For example, two were out due to injuries sustained in the field (unrelated to this project). Two others have been diverted to special assignments since the start of the study and are unable to participate during their shifts.

Only 16 of the 20 officers have actually knocked on any doors at the time of this writing. One additional officer has been recruited and trained for the project, and we anticipate that 1-2 additional officers may be added in the coming weeks to provide additional support. At the time of this writing, only 489 of 926 houses assigned treatment have been visited, and outcome data have not been collected for the majority of respondents.

Analyses of the first wave responses will be conducted once the following conditions are met: 1) all 926 houses assigned to treatment have been visited; 2) all treated and control houses have been invited back to participate in the first post-treatment followup survey, and two weeks time has passed since all reinterview invitations have been delivered.

Revision to Reinterview Procedures

Three days after receiving treatment, treated households and the control households in their quartet are invited back to participate in the post-treatment followup survey. To increase response rates, we send a text message to those who provided their mobile phone numbers at the same time as the re-invitation email. We then send at least three reminder emails to respondents. The survey expires after two weeks have elapsed since the time of the initial reinterview solicitation.

As outlined in the PAP filed on 15 September 2018, we planned to invite respondents back for multiple post-treatment followup surveys. We initially planned to start recruitment for the second post-treatment followup survey approximately 10 days after respondents completed the first post-treatment followup survey. However, we have decided to delay the reinterview attempts until three weeks after responses to the first post-treatment followup survey have passed.

We had initially anticipated that we would have the budget to conduct more than two post-treatment followup surveys. However, we no longer believe a third followup survey is feasible. This is largely driven by the fact that we have decided to allocate more of our budget toward increasing the response rates to the first post-treatment survey wave.

We have found that although the majority of respondents finish the survey within the first few days of receiving their reinterview invitation, many respondents wait for more than a week before responding. To increase response rates to the first post-treatment survey, we decided to offer an additional \$5 incentive payment to those who have still not responded within a week of receiving their reinterview invitation.

Reinvitations to the second post-treatment survey will begin on 5 October 2018. This allows for three weeks to elapse between the first wave reinterview invitations that began on 14 September 2018. Reinterview

invitations for the second post-treatment wave will be sent three weeks after a respondent has finished the first post-treatment wave. Therefore the earliest possible response to the first post-treatment survey wave is 14 September 2018, and the earliest possible response to the second post-treatment survey wave is 5 October 2018. Respondents who do not complete the first post-treatment followup wave will not be invited to complete the second post-treatment followup wave.

Revision to Secondary Outcomes

The PAP uploaded to OSF on 15 September 2018 contained two minor errors. Among the secondary outcome measures listed in the PAP, we enumerated two benchmarking questions for analysis as outcome variables:

1. Do you agree or disagree with this statement: "I respect the police in my community." [7-point scale from "Strongly Disagree" to "Strongly Agree"]
2. How much confidence do you have in each of the following institutions in your community? [The police; the city government; the courts; the public schools. (4) A lot; (3) Some; (2) Only a little; (1) None at all]

The second question comes from Pew¹, and is embedded in a 4-item matrix that also asked respondents to rate the city government, the courts, and the public schools. This matrix of questions was not included in the first post-treatment survey, nor will it be included in any follow up surveys. These 4 questions were omitted in order to reduce the length of the questionnaire. Responses to this question in the pre-treatment survey will still be used for benchmarking our sample against nationally representative data.

We did retain the first single item question to use in post-treatment surveys, which is based on a question that Gallup has asked since 1965. However, the question wording and response categories we use differ from the Gallup question. The Gallup question wording is: "How much respect do you have for the police in your area – a great deal, some, or hardly any?", and the Gallup question has 3 response categories: "a great deal", "some", or "hardly any"².

The question we use in the pre-treatment and post-treatment surveys, by contrast, instead asks respondents to indicate their agreement or disagreement with the statement "I respect the police in my community", and has 7 response categories (Strongly Disagree to Strongly Agree) with a neutral midpoint. In our judgement, the revision is an improvement over the Gallup question, which only offers three response categories and does not include a neutral midpoint. However, we recognize the revision renders it unsuitable for direct comparison to the Gallup question, and it will not be used for benchmarking purposes.

Instead, we will incorporate our revised question into the 5-item "Community Support Index" listed in the PAP for a revised 6-item Community Support Index:

1. The police make me feel safer in my neighborhood.
2. The police are trying to make my community better.
3. The police respect the people in my community.
4. The police treat all people equally.
5. The police are a part of my community.
6. I respect the police in my community.

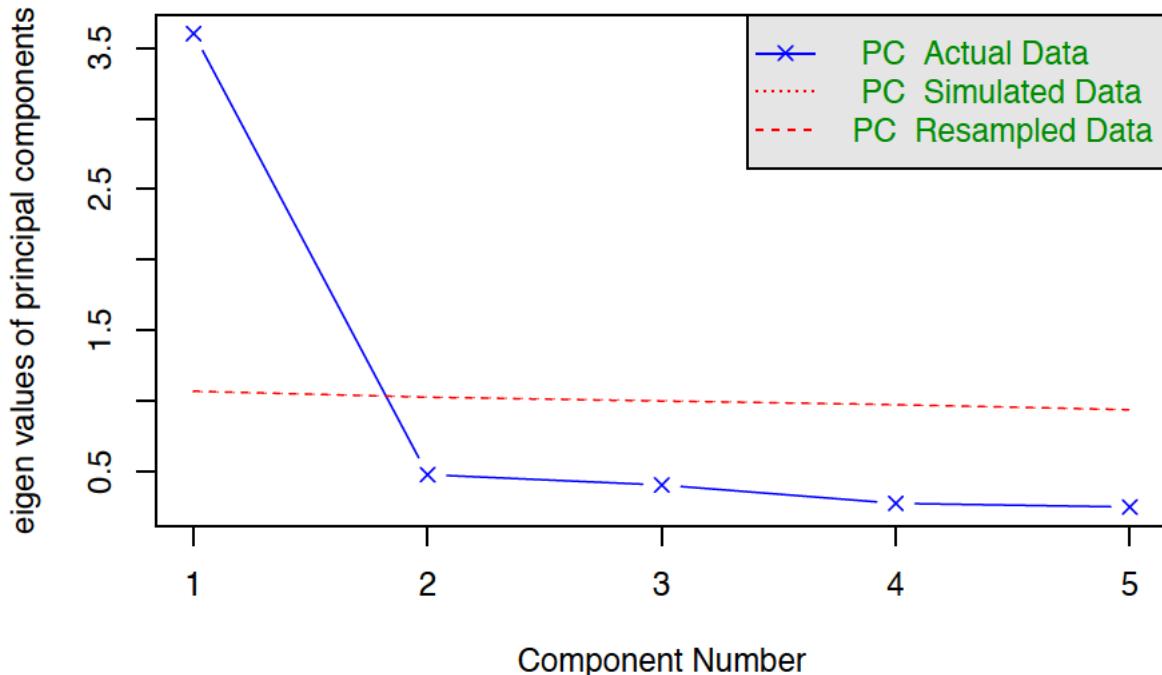
We believe this is justified for a few reasons. This revised Gallup question has the same number of response categories as the other 5-items in the index, and the question content – whether the respondent respects the police in their community – is very similar. Moreover, responses to this question were highly correlated with the other 5 questions in the baseline survey, and parallel analysis suggests 1 component is present in the data matrix of these 6 items:

¹see <http://www.pewsocialtrends.org/2016/09/29/the-racial-confidence-gap-in-police-performance/>

²See <http://www.gallup.com/file/poll/196631/161022RespectPolice.pdf>

```
nhos_t0_df %>%
  dplyr::select(-gallup, -safe, -equal, -better,
               -respect) %>%
  psych::fa.parallel(fa = "pc")
```

Parallel Analysis Scree Plots



```
## Parallel analysis suggests that the number of factors = NA and the number of components = 1
```

Second Supplement to ‘Pre-Analysis Plan for September 2018 COPS Experiment’



2018-10-28

Contents

Update on Progress	2
Timing of Reinterview Invitations	2
Revisions to Survey Items in T2 Survey	2

Update on Progress

This document is a second supplement to the pre-analysis plan (PAP) filed on 15 September 2018. The study is still ongoing. The last units were treated on 17 October 2018. Outcome measurement is still incomplete. At the time of this writing, all units in the study have been invited back for the first post-treatment followup survey (T1 Survey). The last units were invited on 20th October, 3 days after treatment. All T1 Survey invitations remain open for 2 weeks. Therefore, the T1 Survey will completely expire on 3 November 2018. At the time of this writing, 1,457 (out of 2,013) individuals have completed the T1 Survey, and 863 individuals have been invited to the T2 Survey, and 579 have completed it.

Timing of Reinterview Invitations

Three days after receiving treatment, treated households and the control households in their quartet were invited back to participate in the T1 Survey. In the “Supplement to ‘Pre-Analysis Plan for September 2018 COPS Experiment’” from October 2nd, we described a reinterview procedure for the T2 Survey where a unit became eligible for a re-interview invitation three weeks after they completed the T1 Survey. We subsequently realized that this was an error, as we had intended the reference date to be the time of receiving treatment, rather than the T1 Survey completion date.

Thus, here we are correcting this mistake, and clarifying that the eligibility criteria for the T2 Survey is *three weeks after the time of treatment*. The revised eligibility criteria for the T2 Survey are analogous to those used for the T1 Survey. Units become eligible for a T2 Survey invitation once the following criteria are satisfied:

- A response to the T1 Survey has been recorded, and an invitation to the T2 Survey has not yet been sent.
 - For units assigned treatment:
 - (2a) Three weeks (21 days) have passed since treatment (18 days since the T1 Survey invitation).
 - For units assigned control:
 - (2b) The unit belongs to a quartet where at least 1 unit assigned treatment satisfies condition (2a).

Our budget constraint does not allow for a third followup survey. As such, the T2 Survey will be the final phase of outcome measurement. The final invitations to the T2 Survey will be sent on November 7th (3 weeks after the last set of units were treated on 17th October). We anticipate that the study will be complete on November 21st, allowing the last batch of units invited to the T2 Survey a period of 2 weeks to respond.

Revisions to Survey Items in T2 Survey

The trust in government questions that were included in the T1 survey were not included in the T2 survey to make room for a 7-item Cognitive Reflection Test (CRT) battery. We plan to use CRT scores to investigate the association between cognitive reflection and stereotype endorsement with the expectation that participants with higher CRT scores are less likely to endorse stereotypes about police. For the purpose of exploratory analysis, we will also investigate whether cognitive reflection predicts treatment effect heterogeneity on police stereotypes (if applicable). We also plan to use CRT scores to investigate the association between cognitive reflection and the decision to enter into a raffle instead of choosing to receive a \$5 payment.

C Appendix C: Survey Instruments

C.1 Baseline (T0 Survey)

This study is conducted by researchers at [REDACTED] to learn about the views and opinions of the residents in [REDACTED]. Participation in this study will involve answering a series of survey questions that will only take about 10 minutes of your time. **After completing this short initial survey, you will be entered into a raffle to receive one of over a dozen \$75 gift cards that can be spent at your choice of 84 popular companies, including Amazon.com, or donated directly to one of 18 charities, such as the American Cancer Society.**

There are no known or anticipated risks to you for participating. We hope that our results will add to the knowledge about the views and opinions of [REDACTED]. **All of your responses will be held in confidence.** Only the researchers involved in this study and those responsible for research oversight will have access to the information you provide.

By completing the first survey, you will become a member of the 2018 [REDACTED] Opinion Study Panel, for which you may be invited to participate in additional follow-up surveys in the months ahead. **For each follow-up survey you take, you may choose to either receive a \$5 gift card or enroll in another raffle with even greater odds of winning a \$75 gift card.** We value your privacy, and are not interested in the responses of any particular person. Any contact information you provide will only be used to invite you to participate in our followup surveys and deliver gift cards. **This information will not be shared with any third-party, and will be deleted from our records once our study is completed.**

Participation in this study is completely voluntary. You are free to decline to participate, to end participation at any time for any reason, or to refuse to answer any individual question without penalty. Your decision whether or not to participate in this study will not affect your relationship with [REDACTED]

If you have any questions about this study, you may contact the principal investigator, [REDACTED]. If you would like to talk with someone other than the researchers to discuss problems or concerns, to discuss situations in the event that a member of the research team is not available, or to discuss your rights as a research participant, you may contact the [REDACTED] Human Subjects Committee [REDACTED]

[REDACTED] Would you like to participate in this study?

- I agree to participate
- I do not agree to participate

Thank you for being part of the [REDACTED] Opinion Survey! Before we begin, please

confirm the contact information we have on file is correct.

Name: \${e://Field/nameline}

Date of Birth: \${e://Field/Voter_BirthDate}

Address:

\${e://Field/address}
\${e://Field/city} \${e://Field/state} \${e://Field/mzip4}

- Everything looks good
- Something is incorrect

Please update your contact details using the form below. For your reference, the information we have on file is displayed on the left.

- Name: \${e://Field/nameline} _____
- Date of Birth: \${e://Field/Voter_BirthDate} _____
- Address: \${e://Field/address} _____
- City: \${e://Field/city} _____
- State: \${e://Field/state} _____
- ZIP: \${e://Field/mzip4} _____

First, we want to hear about how you feel about life in [REDACTED] in general. How many years have you lived in [REDACTED]?

All things considered, how much do you like living in [REDACTED] right now?

- A great deal
- A lot
- A moderate amount
- A little
- None at all

What's your favorite thing about living in [REDACTED]?

- My family and friends are in the area
- Job and career prospects
- Cost of living
- Something else _____

What do you think is the most important way that life in [REDACTED] could improve?

- Better economy and job prospects
- Quality of schools
- Traffic and access to transportation
- Access to quality health care
- Crime control
- Something else _____

In general, do you approve of the job [REDACTED] is doing in office?

- Strongly approve
- Somewhat approve
- Neither approve nor disapprove
- Somewhat disapprove
- Strongly disapprove
- No opinion/Never heard of her

How much confidence do you have in each of the following institutions in your community?

	A lot	Some	Only a little	None at all
The police department	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The courts	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
City government	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Public schools	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How people feel about their local police is of tremendous importance in the United States. We would like to know **how you feel about your local police department, the [REDACTED] Police Department**, and how any interactions you may have had with them have gone. Please answer these questions honestly and remember that your responses are confidential.

During the past 12 months, have you had any contact with the [REDACTED] police?

- Yes
- No

Were any of these contacts with a police officer in person? That is, did you interact with an officer face-to-face?

- Yes
- No

If you feel comfortable answering, can you tell us in a few words what the contact was about? If you'd rather not say, please just click next

Have you ever been arrested by the [REDACTED] Police?

- Yes
- No

Have you ever been treated unfairly by the [REDACTED] Police?

- Yes
- No

Do you agree or disagree with this statement: "**The police make me feel safer in my neighborhood.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

As you may know, [REDACTED] has recently required police to wear body cameras that record their activities while on duty. **Do you support or oppose police body cameras?**

- Strongly support
- Support
- Somewhat support
- Neither support nor oppose
- Somewhat oppose
- Oppose
- Strongly oppose

Do you agree or disagree with this statement: "**The police in [REDACTED] treat all people equally.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**The police are trying to make my community better.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**The police respect the people in my community.**"

- Strongly agree
 - Agree
 - Somewhat agree
 - Neither agree nor disagree
 - Somewhat disagree
 - Disagree
 - Strongly disagree
-

Do you agree or disagree with this statement: "**I respect the police in my community.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

How would you feel about increasing the number of police on the street by 10 percent, even if it means fewer funds for other public services?

- Strongly support
- Support
- Somewhat support
- Neither support nor oppose
- Somewhat oppose
- Oppose
- Strongly oppose

For the next set of questions, we will ask you to estimate how likely you are to engage in certain kinds of behaviors that involve contact with **your local police department**.

How likely would you be to attend a community meeting to discuss problems in your neighborhood with the police?

- Extremely likely
 - Moderately likely
 - Slightly likely
 - Neither likely nor unlikely
 - Slightly unlikely
 - Moderately unlikely
 - Extremely unlikely
-

How likely would you be to report suspicious activity to the police?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

How likely would you be to call the police to report a crime?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

If the police were looking for a suspect who was hiding, and you knew where that person was, how likely would you be to provide the police with information?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

We'd like to hear more about how you feel about the police in [REDACTED]. Do you agree or disagree with the statements below?

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
They care about the well-being of everyone they deal with	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They make fair and impartial decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They act in ways consistent with your own ideas about what is right and wrong	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They stand up for values that are important to you	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They behave according to the law when dealing with people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They make decisions based on their biases or opinions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

They are often dishonest	<input type="radio"/>						
They have the same sense of right and wrong that you do	<input type="radio"/>						

If the police tell you to do something, you should do it.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

There are times when it is alright to ignore what the police tell you to do.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

It is your duty to accept the decisions made by the police even when you disagree with them.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

It is your duty to do what the police tell you to do, even if you do not like the way they treat you.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

I have confidence that the police in [REDACTED] can do their job well.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

The police in [REDACTED] are effective at fighting crime.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

The police in [REDACTED] are effective at helping people who ask for help.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

When people in your neighborhood call the police for help, they respond quickly.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

The police are a part of my community.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Next we'd like to hear about how you feel about different levels of government. **These questions do not refer to any particular individual or political party.**

How often do you think you can trust the government to do what is right?

	Always	Most of the time	About half the time	Sometimes	Never
Federal government	<input type="radio"/>				
State [REDACTED]	<input type="radio"/>				
City [REDACTED]	<input type="radio"/>				

Do you think that people in government waste much of the money we pay in taxes?

	Waste a lot	Waste some	Don't waste very much
Federal government	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
State [REDACTED]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
City [REDACTED]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How many of the people running the government are corrupt?

	All	Most	About half	A few	None
Federal government	<input type="radio"/>				
State [REDACTED]	<input type="radio"/>				
City [REDACTED]	<input type="radio"/>				

Would you say the government is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people?

	Run by a few big interests	For the benefit of all the people
Federal government	<input type="radio"/>	<input type="radio"/>
State [REDACTED]	<input type="radio"/>	<input type="radio"/>
City [REDACTED]	<input type="radio"/>	<input type="radio"/>

Now we'll turn to a few questions about politics. Generally speaking, do you think of yourself as a...?

- Democrat
- Republican
- Independent

Would you call yourself a...

- Strong Republican
- Not very strong Republican

Would you call yourself a...

- Strong Democrat
- Not very strong Democrat

Do you think of yourself as closer to the...

- The Democratic Party
- The Republican Party

We'd like to get your feelings toward some people and groups who are in the news these days using something we call the **feeling thermometer**.

Here's how it works. You can choose any number between 0 and 100. The higher the number, the warmer or more favorable you feel toward that person or group; the lower the number, the colder or less favorable. You would rate the person or group at the 50 degree mark if you don't feel particularly warm or cold toward them. If you have never heard of the person or group, just check "Haven't heard of this person/group."

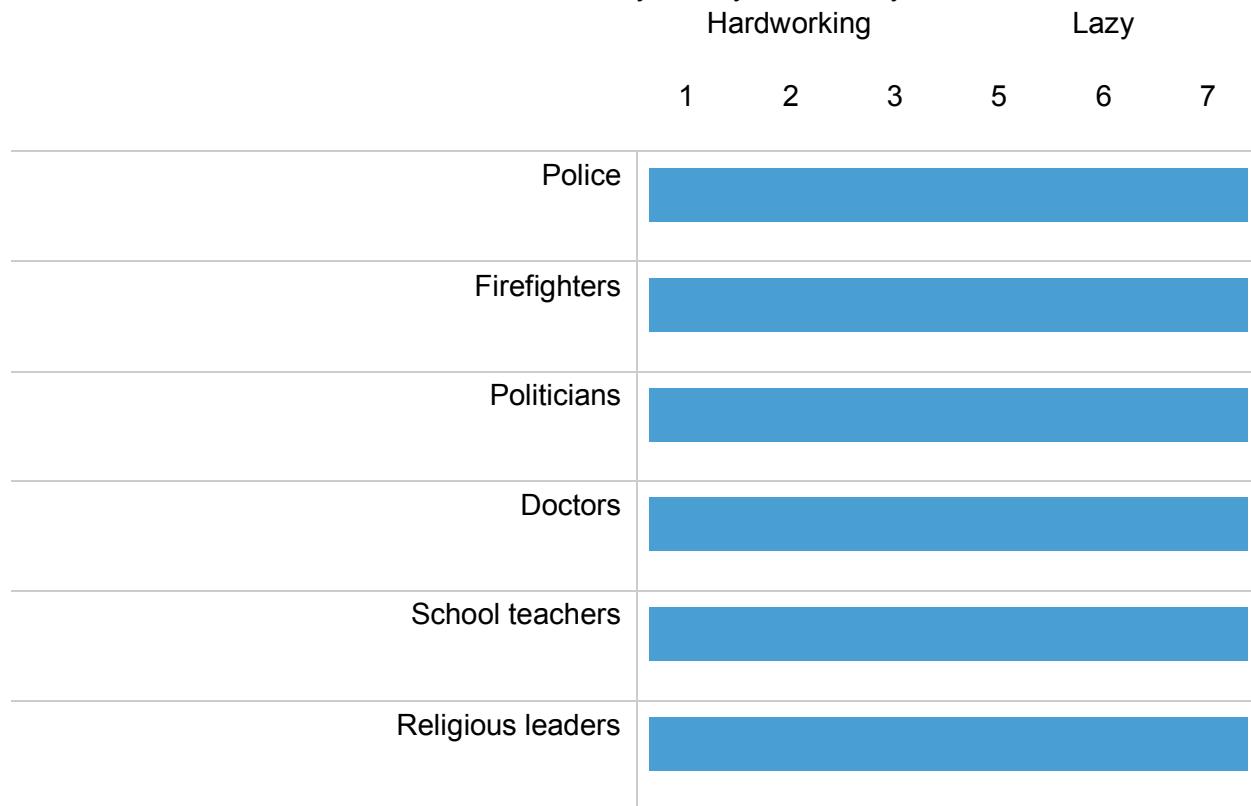
Haven't heard of this person/group

0 10 20 30 40 50 60 70 80 90 100

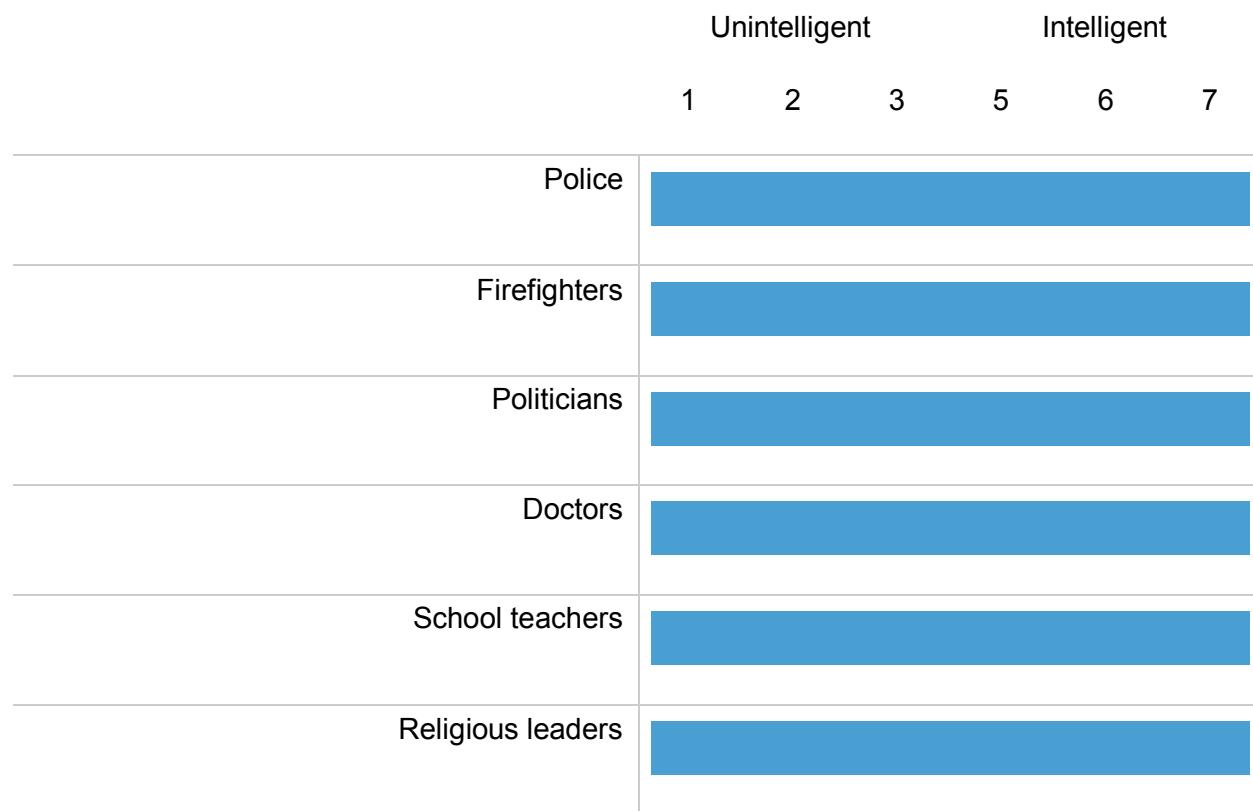
Donald Trump, President of the United States	
[REDACTED] US Senator [REDACTED]	
African-Americans	
Muslims in America	
Illegal immigrants in America	
Black Lives Matter	
Police officers	
National Rifle Association	
Firefighters	

In this section, you will see a series of seven-point scales on which the characteristics of people in a group can be rated. Please choose the number that comes closest to where you think people in the group stand.

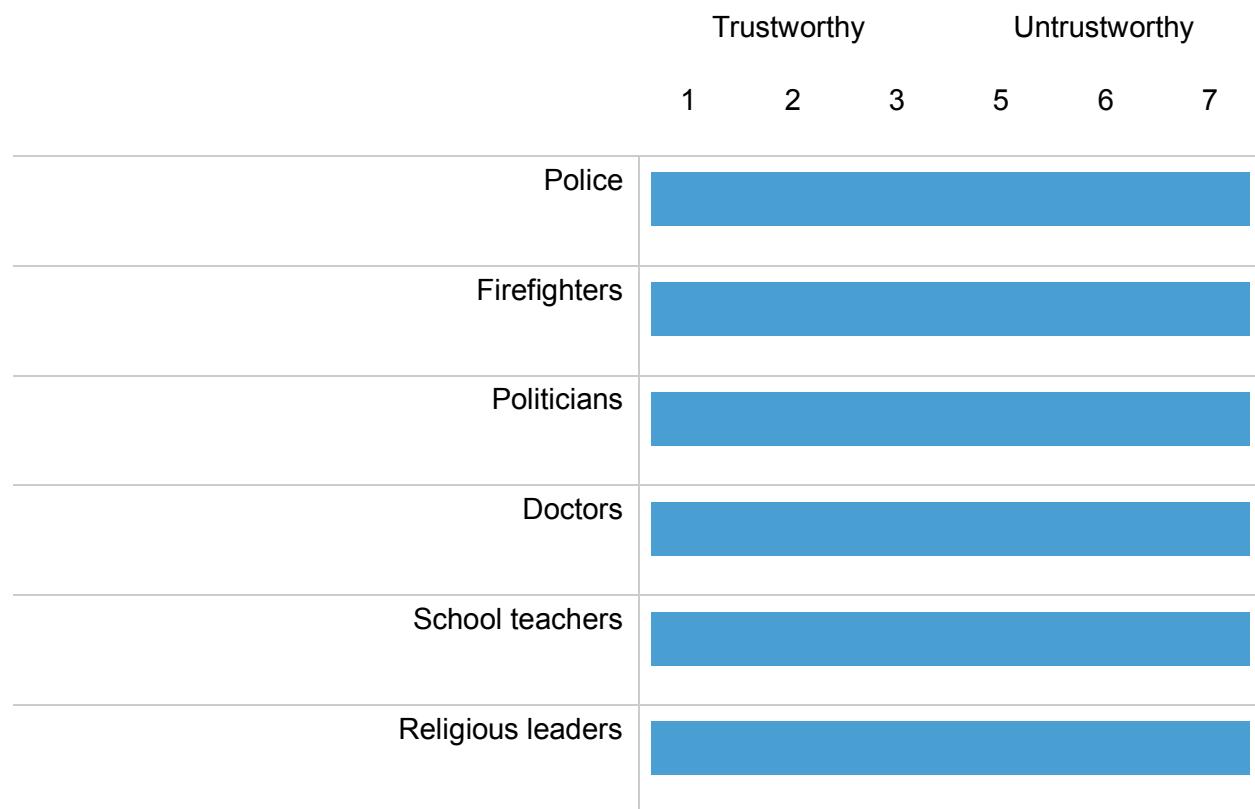
A score of 1 means you think almost all people in that group are "hardworking" and a score of 7 means you think all people in that group are "lazy". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



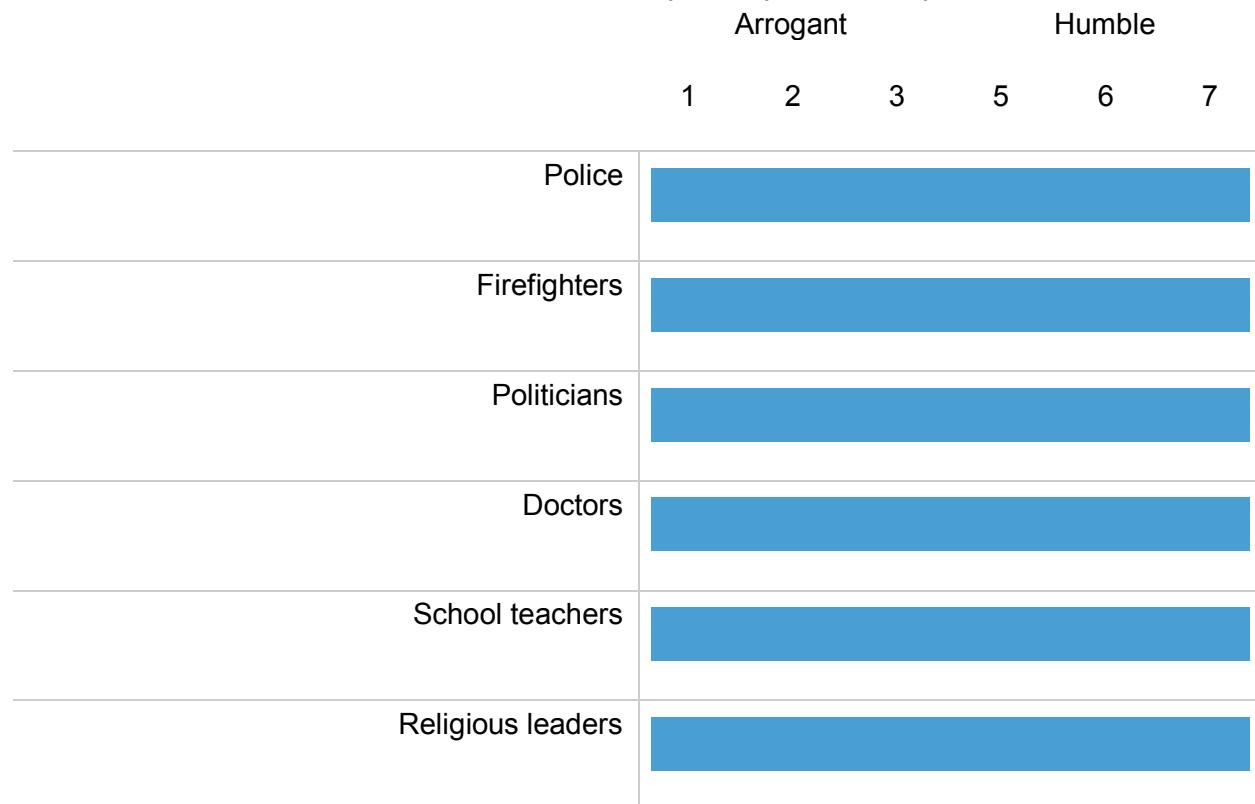
A score of 1 means you think almost all people in that group are "unintelligent" and a score of 7 means you think all people in that group are "intelligent". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



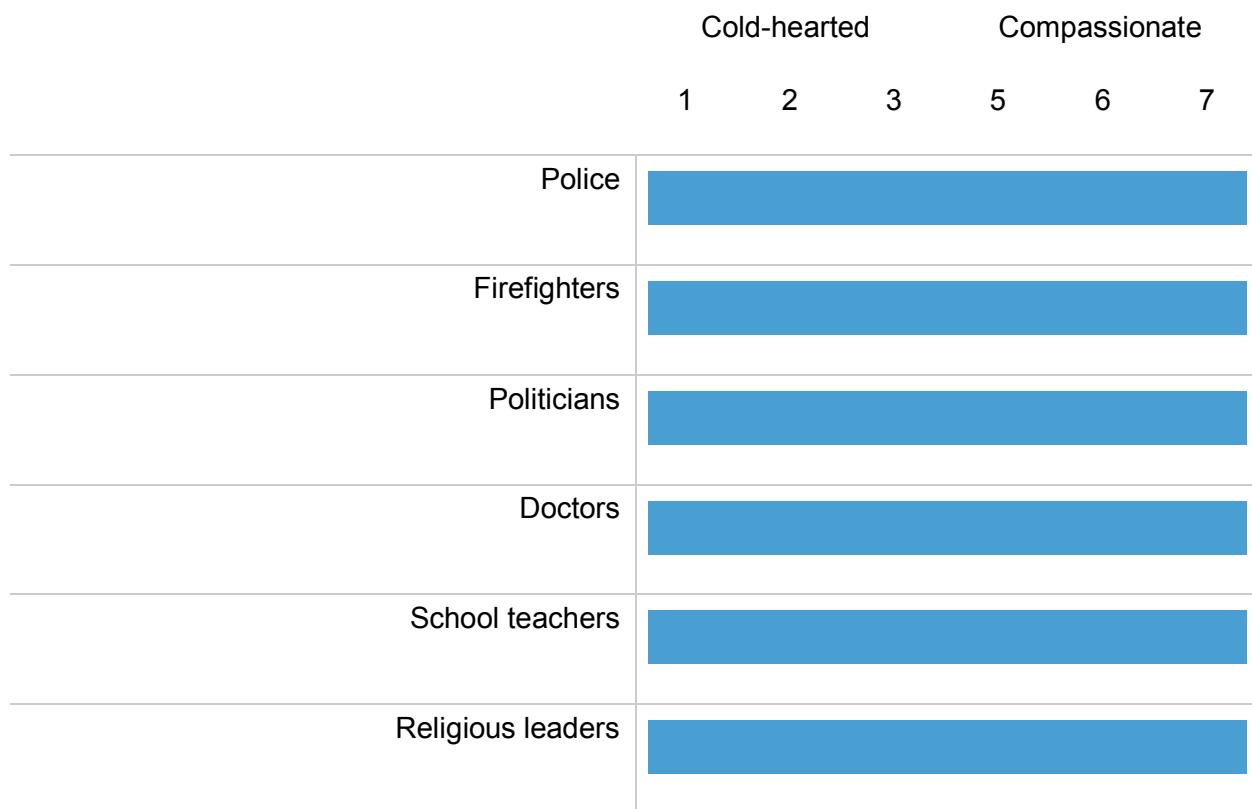
A score of 1 means you think almost all people in that group are "trustworthy" and a score of 7 means you think all people in that group are "untrustworthy". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



A score of 1 means you think almost all people in that group are "arrogant" and a score of 7 means you think all people in that group are "humble". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



A score of 1 means you think almost all people in that group are "cold-hearted" and a score of 7 means you think all people in that group are "compassionate". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



In this final section you will be asked a few background questions. Please answer honestly and remember that your responses are confidential.

Are you male or female?

Male

Female

What is the highest level of education you have completed?

- Did not graduate from high school
- High school graduate
- Some college, but no degree
- 2-year college degree
- 4-year college degree
- Postgraduate degree

Which statement best describes your current employment status?

- Working full time now
- Working part time now
- Temporarily laid off
- Unemployed
- Retired
- Permanently disabled
- Taking care of home or family
- Student
- Other _____

Which category represents your annual income?

- Less than \$5,000
- \$5,000 - \$7,499
- \$7,500 - \$9,999
- \$10,000-\$12,499
- \$12,500 - \$14,999
- \$15,000-\$19,999
- \$20,000-\$24,999
- \$25,000-\$29,999
- \$30,000-\$34,999
- \$35,00-\$39,999
- \$40,000-\$49,999
- \$50,000-\$59,999
- \$60,000-\$74,999
- \$75,000-\$99,999
- \$100,00-\$149,999
- \$150,000 or more

What racial or ethnic group best describes you?

- White
- Black or African American
- Hispanic or Latino
- Asian
- Native American
- Mixed
- Middle Eastern
- Other _____

Upon completing this survey, you may become eligible for a brief follow-up survey in a few weeks, for which you can either choose to receive \$5, or enroll in another raffle with an even greater chance of winning \$75.

We will also use this email address to contact you if you are one of the winners in the \$75 raffle.

Please enter your email address below so that we can send you an invitation for the follow-up survey. **This information is confidential and we will never give this information to anyone else.**

Please enter your mobile phone number below if you would like a call or text reminder when the next survey is available. This information is confidential and we will never give this information to anyone else.

It looks like you made a typo. We've detected that \${e://Field/enteredemail} is not a valid email address. Please try again.

We've detected that \${e://Field/enteredemail} is not a valid email address. Please enter a valid email address below if you want to participate in follow-up surveys. You may leave this blank if you do not want to participate in follow-up surveys.

Thank you for completing this survey! You have now been entered into a raffle to win one more than dozen \$75 prizes!

If you are selected as one of the winners, we will email you a confirmation link containing instructions within 24 hours of the raffle to \${e://Field/enteredemail} on September 6th. You may choose to spend the card at one of 84 popular companies, or donate to one of 18 charities.

You can see all the available options by visiting this link: <https://www.rewardsgenius.com/reward-link/>

C.2 First Followup (T1 Survey)

Thank you for participating in the 2018 [REDACTED] Opinion Survey! Click the ">>" button below to continue.

We want to make sure the data are accurate. This survey is for \${m://FirstName}
\${m://LastName}. Are you \${m://FirstName}?

- Yes
- No, I got the wrong survey

Please email [REDACTED] or call [REDACTED] for assistance.

First, we'd like to hear how you've been doing over the last few weeks. All things considered, over the last month, do you think life in [REDACTED] has been getting better, worse, or has stayed the same?

- Getting better
- Getting worse
- Stayed the same

Based on your experiences over the last few weeks, of all the issues listed below, which do you think is the most important for city leaders to consider?

- Garbage service
- Police presence
- Safer streets
- Public transport
- Red light cameras
- Something else _____

Over the last week, have you taken public transportation (e.g. a bus)?

- Yes
- No

Some people say public transit, like buses, should be expanded in [REDACTED]. Others say it is better to improve the roads we have. What do you think?

- More important to expand public transit
- More important to improve existing roads
- Neither

In general, do you approve of the job [REDACTED] is doing in office?

- Strongly approve
- Approve
- Somewhat approve
- Neither approve nor disapprove
- Somewhat disapprove
- Disapprove
- Strongly disapprove
- No opinion/never heard of her

The state government in [REDACTED] is considering changing the laws that prohibit the use of marijuana. Do you think the use of marijuana should be legal in [REDACTED]?

- No, marijuana should be illegal in all cases
- Yes, but only for medical reasons
- Yes, it should be legal for any reason, including recreational purposes

How people feel about their local police is of tremendous importance in the United States. We would like to know **how you feel about your local police department, the [REDACTED]**. Please answer these questions honestly and remember that your responses are confidential.

Do you agree or disagree with this statement: "**The police make me feel safer in my neighborhood.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**The police in [REDACTED] treat all people equally.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**The police are trying to make my community better.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**The police respect the people in my community.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**I respect the police in my community.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

How would you feel about increasing the number of police on the street by 10 percent, even if it means fewer funds for other public services?

- Strongly support
- Support
- Somewhat support
- Neither support nor oppose
- Somewhat oppose
- Oppose
- Strongly oppose

As you may know, the city of [REDACTED] has recently required police to wear body cameras that record their activities while on duty. **Do you support or oppose police body cameras?**

- Strongly support
- Support
- Somewhat support
- Neither support nor oppose
- Somewhat oppose
- Oppose
- Strongly oppose

For the next set of questions, we will ask you to estimate how likely you are to engage in certain kinds of behaviors that involve contact with **your local police department**.

How likely would you be to attend a community meeting to discuss problems in your neighborhood with the police?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

How likely would you be to report suspicious activity to the police?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

How likely would you be to call the police to report a crime?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

If the police were looking for a suspect who was hiding, and you knew where that person was, how likely would you be to provide the police with information?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

We'd like to hear more about how you feel about the police in [REDACTED] Do you agree or disagree with the statements below?

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
They care about the well-being of everyone they deal with	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They make fair and impartial decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They act in ways consistent with your own ideas about what is right and wrong	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They stand up for values that are important to you	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They behave according to the law when dealing with people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They make decisions based on their biases or opinions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

They are often dishonest	<input type="radio"/>						
They have the same sense of right and wrong that you do	<input type="radio"/>						

I have confidence that the police in [REDACTED] can do their job well.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

The police in [REDACTED] are effective at fighting crime.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

The police in [REDACTED] are effective at helping people who ask for help.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

When people in your neighborhood call the police for help, they respond quickly.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

The police are a part of my community.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

If the police tell you to do something, you should do it.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

There are times when it is alright to ignore what the police tell you to do.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

It is your duty to accept the decisions made by the police even when you disagree with them.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

It is your duty to do what the police tell you to do, even if you do not like the way they treat you.

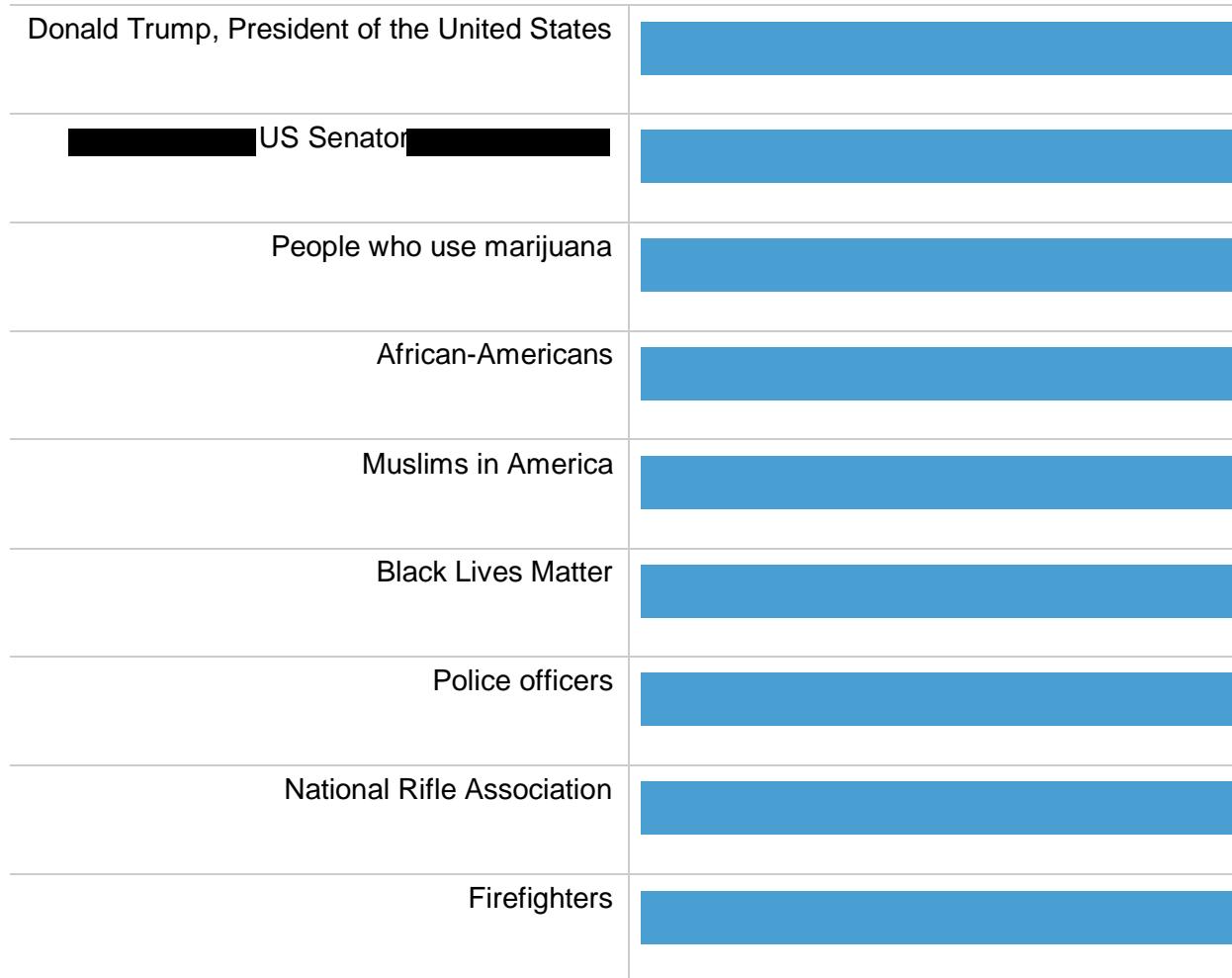
- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

We'd like to get your feelings toward some people and groups who are in the news these days using something we call the **feeling thermometer**.

Here's how it works. You can choose any number between 0 and 100. The higher the number, the warmer or more favorable you feel toward that person or group; the lower the number, the colder or less favorable. You would rate the person or group at the 50 degree mark if you don't feel particularly warm or cold toward them. If you have never heard of the person or group, just check "Haven't heard of this person/group."

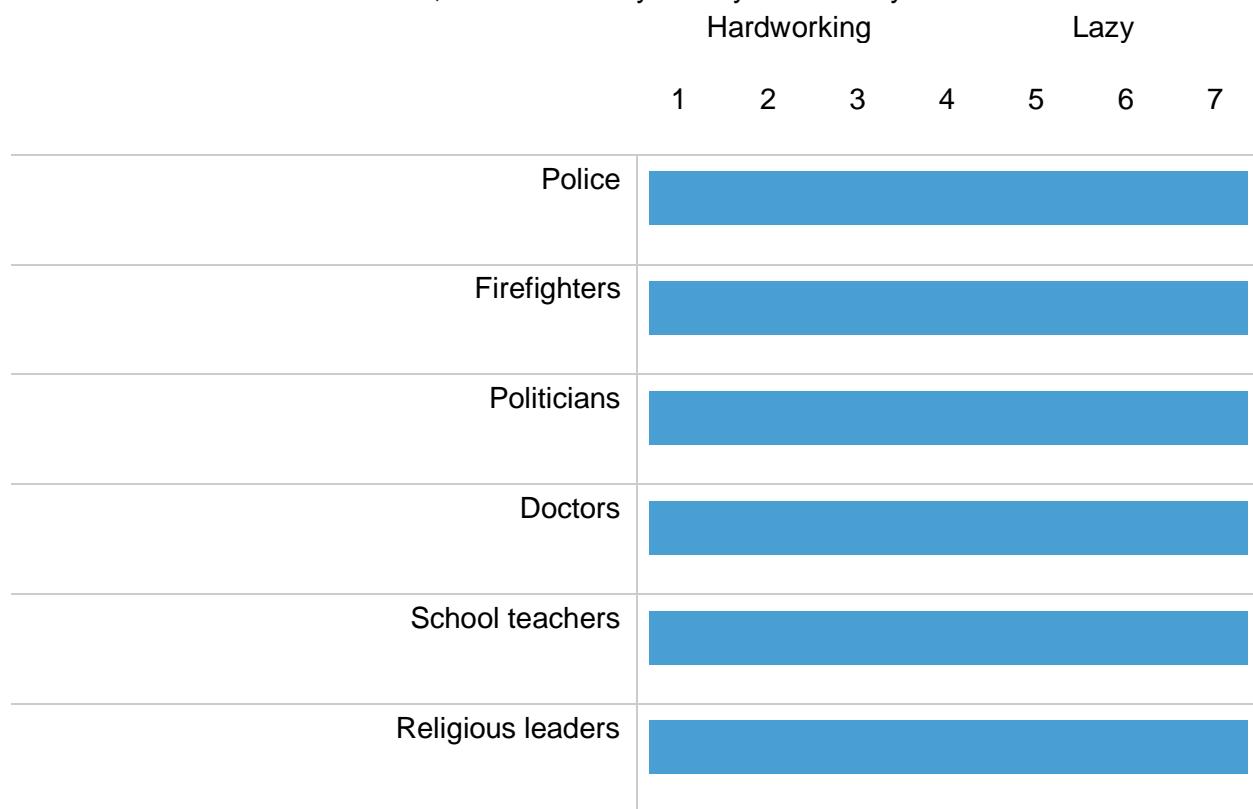
Haven't heard of this person/group

0 10 20 30 40 50 60 70 80 90 100

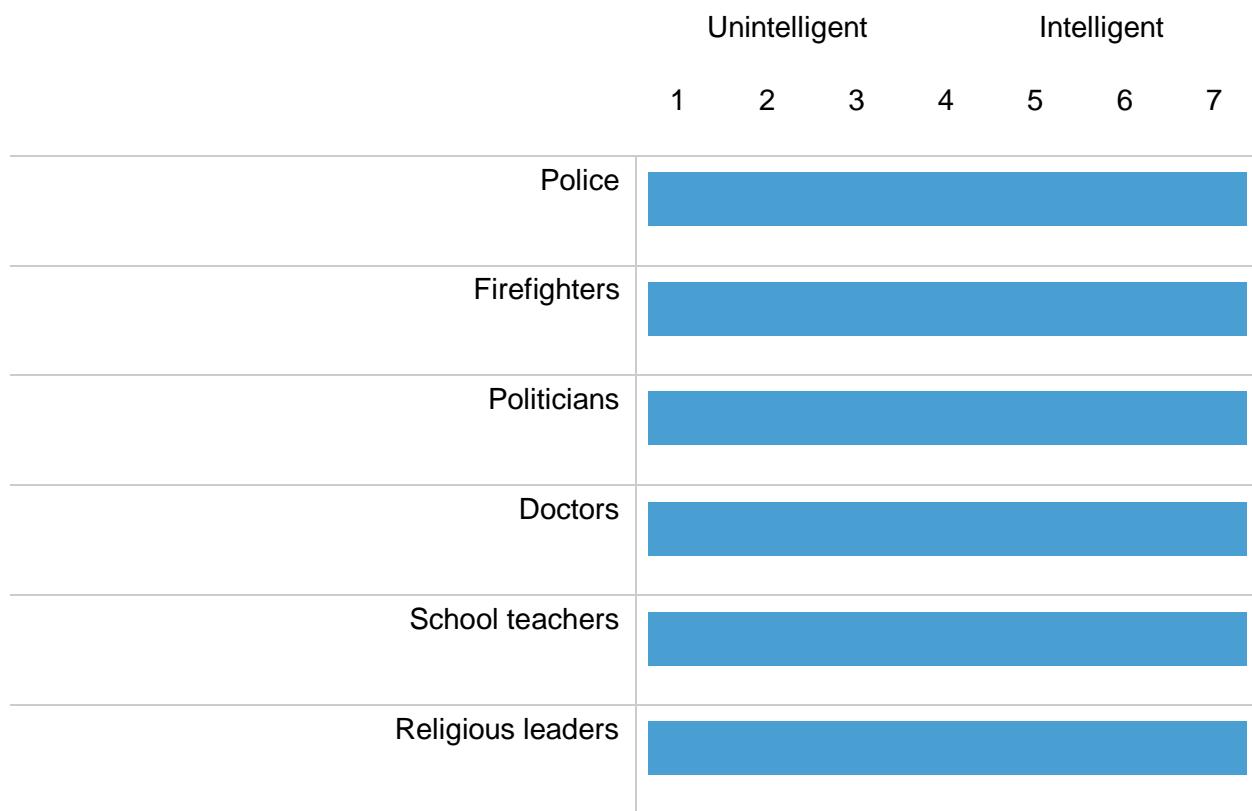


In this section, you will see a series of seven-point scales on which the characteristics of people in a group can be rated. Please choose the number that comes closest to where you think people in the group stand.

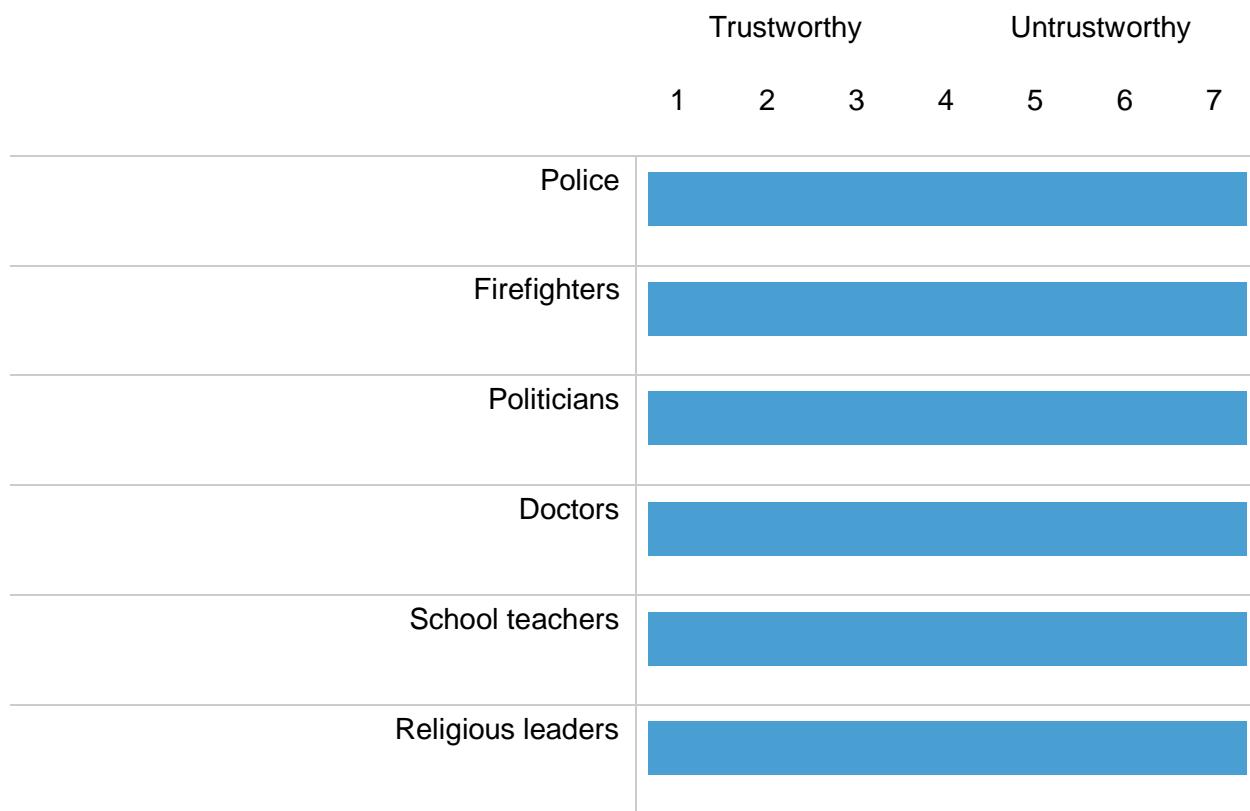
A score of 1 means you think almost all people in that group are "hardworking" and a score of 7 means you think all people in that group are "lazy". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



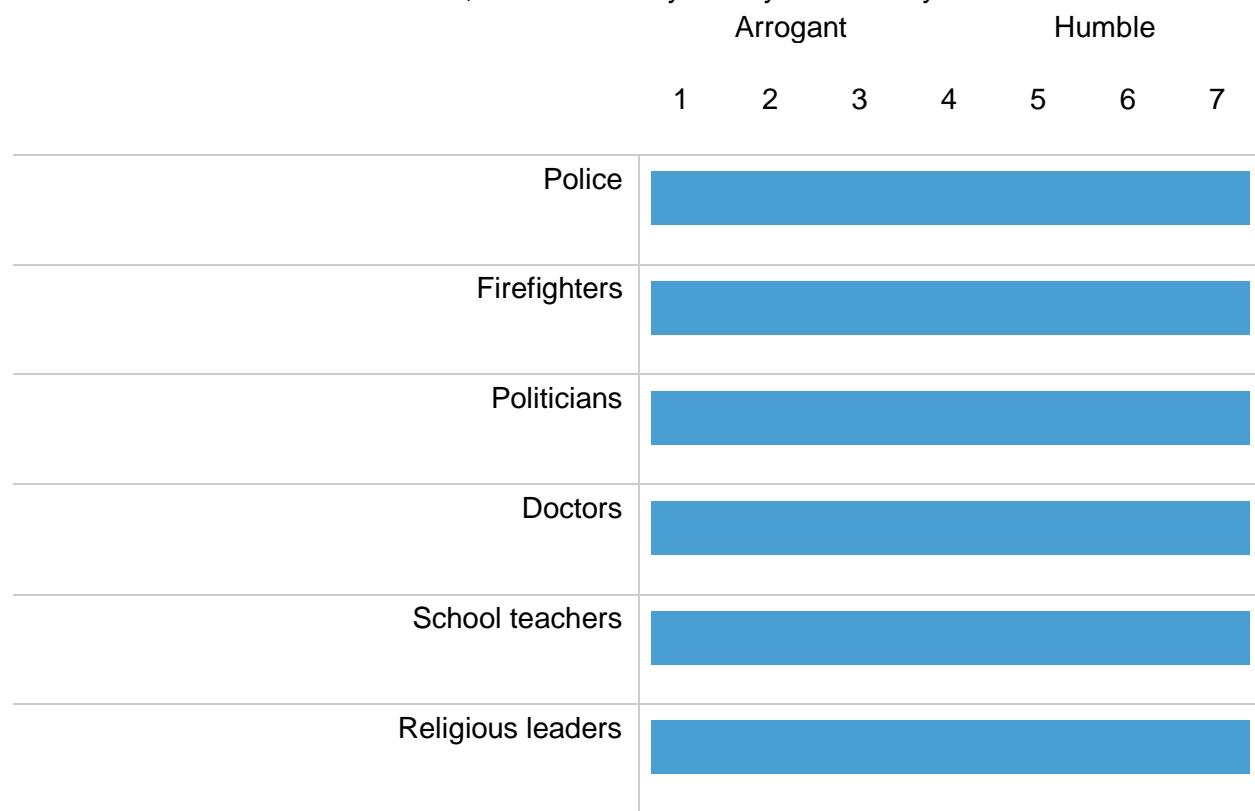
A score of 1 means you think almost all people in that group are "unintelligent" and a score of 7 means you think all people in that group are "intelligent". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



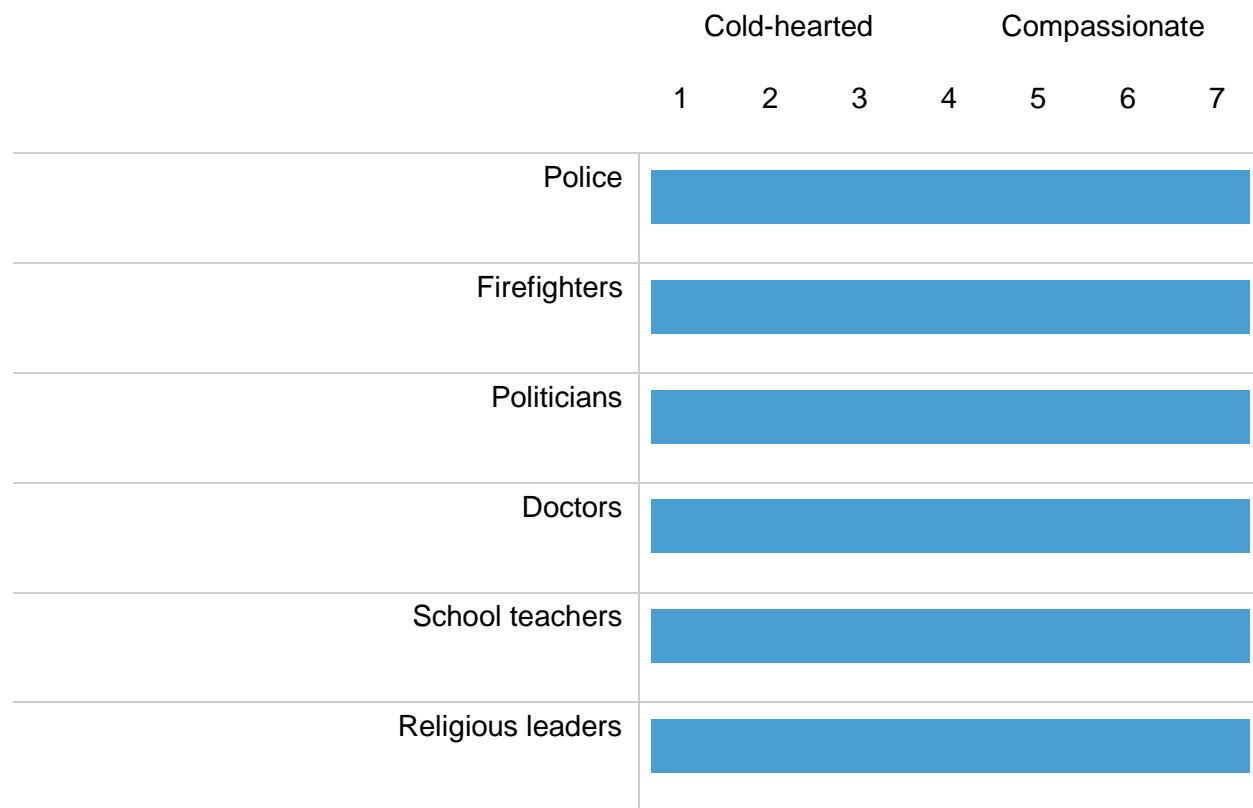
A score of 1 means you think almost all people in that group are "trustworthy" and a score of 7 means you think all people in that group are "untrustworthy". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



A score of 1 means you think almost all people in that group are "arrogant" and a score of 7 means you think all people in that group are "humble". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



A score of 1 means you think almost all people in that group are "cold-hearted" and a score of 7 means you think all people in that group are "compassionate". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



Next we'd like to hear about how you feel about different levels of government. **These questions do not refer to any particular individual or political party.**

How often do you think you can trust the government to do what is right?

	Always	Most of the time	About half the time	Sometimes	Never
Federal government	<input type="radio"/>				
State [REDACTED]	<input type="radio"/>				
City [REDACTED]	<input type="radio"/>				

Do you think that people in government waste much of the money we pay in taxes?

	Waste a lot	Waste some	Don't waste very much
Federal government	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
State [REDACTED]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
City [REDACTED]	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How many of the people running the government are corrupt?

	All	Most	About half	A few	None
Federal government	<input type="radio"/>				
State [REDACTED]	<input type="radio"/>				
City [REDACTED]	<input type="radio"/>				

Would you say the government is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people?

	Run by a few big interests	For the benefit of all the people
Federal government	<input type="radio"/>	<input type="radio"/>
State [REDACTED]	<input type="radio"/>	<input type="radio"/>
City [REDACTED]	<input type="radio"/>	<input type="radio"/>

Thank you for your participation! To thank you for your time, we'd like to offer you \$5.00, or you can enroll in another raffle with a chance to win one of over a dozen \$75 prizes. Which do you prefer?

Please note: all \$5.00 payments will be delivered to your email address by TangoCard. Please allow 1-2 business days for payments to be processed, and check your spam box for emails from noreply@tangocard.com.

The raffle will be conducted once all respondents have completed the survey. The winners will be notified via email on before November 5th.

- A \$5.00 payment
- Another raffle

C.3 Second Followup (T2 Survey)



Thank you for participating in the 2018 [REDACTED] Opinion Survey! Click the ">>" button below to continue.

We want to make sure the data are accurate. This survey is for \${m://FirstName}
\${m://LastName}. Are you \${m://FirstName}?

- Yes
- No, I got the wrong survey

Please email [REDACTED] or call [REDACTED] for assistance.

First, we'd like to hear how you've been doing over the last few weeks. All things considered, over the last month, do you think life in [REDACTED] has been getting better, worse, or has stayed the same?

- Getting better
- Getting worse
- Stayed the same

Based on your experiences over the last few weeks, of all the issues listed below, which do you think is the most important for city leaders to consider?

- Garbage service
- Police presence
- Safer streets
- Public transport
- Red light cameras
- Something else _____

What do you think is the most important way that life in [REDACTED] could improve?

- Better economy and job prospects
- Quality of schools
- Traffic and access to transportation
- Access to quality health care
- Crime control
- Something else _____

In general, do you approve of the job [REDACTED] is doing in office?

- Strongly approve
- Approve
- Somewhat approve
- Neither approve nor disapprove
- Somewhat disapprove
- Disapprove
- Strongly disapprove
- No opinion/never heard of her

How people feel about their local police is of tremendous importance in the United States. We would like to know **how you feel about your local police department, the [REDACTED]** [REDACTED]. Please answer these questions honestly and remember that your responses are confidential.

Do you agree or disagree with this statement: "**The police make me feel safer in my neighborhood.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**The police in [REDACTED] treat all people equally.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**The police are trying to make my community better.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**The police respect the people in my community.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

Do you agree or disagree with this statement: "**I respect the police in my community.**"

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

How would you feel about increasing the number of police on the street by 10 percent, even if it means fewer funds for other public services?

- Strongly support
- Support
- Somewhat support
- Neither support nor oppose
- Somewhat oppose
- Oppose
- Strongly oppose

As you may know, the city of [REDACTED] has recently required police to wear body cameras that record their activities while on duty. **Do you support or oppose police body cameras?**

- Strongly support
- Support
- Somewhat support
- Neither support nor oppose
- Somewhat oppose
- Oppose
- Strongly oppose

For the next set of questions, we will ask you to estimate how likely you are to engage in certain kinds of behaviors that involve contact with **your local police department**.

How likely would you be to attend a community meeting to discuss problems in your neighborhood with the police?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

How likely would you be to report suspicious activity to the police?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

How likely would you be to call the police to report a crime?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

If the police were looking for a suspect who was hiding, and you knew where that person was, how likely would you be to provide the police with information?

- Extremely likely
- Moderately likely
- Slightly likely
- Neither likely nor unlikely
- Slightly unlikely
- Moderately unlikely
- Extremely unlikely

We'd like to hear more about how you feel about the police in [REDACTED]. Do you agree or disagree with the statements below?

	Strongly agree	Agree	Somewhat agree	Neither agree nor disagree	Somewhat disagree	Disagree	Strongly disagree
They care about the well-being of everyone they deal with	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They make fair and impartial decisions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They act in ways consistent with your own ideas about what is right and wrong	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They stand up for values that are important to you	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They behave according to the law when dealing with people	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
They make decisions based on their biases or opinions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

They are often dishonest	<input type="radio"/>						
They have the same sense of right and wrong that you do	<input type="radio"/>						

I have confidence that the police in [REDACTED] can do their job well.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

The police in [REDACTED] are effective at fighting crime.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

The police in [REDACTED] are effective at helping people who ask for help.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

When people in your neighborhood call the police for help, they respond quickly.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

The police are a part of my community.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

If the police tell you to do something, you should do it.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

There are times when it is alright to ignore what the police tell you to do.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

It is your duty to accept the decisions made by the police even when you disagree with them.

- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

It is your duty to do what the police tell you to do, even if you do not like the way they treat you.

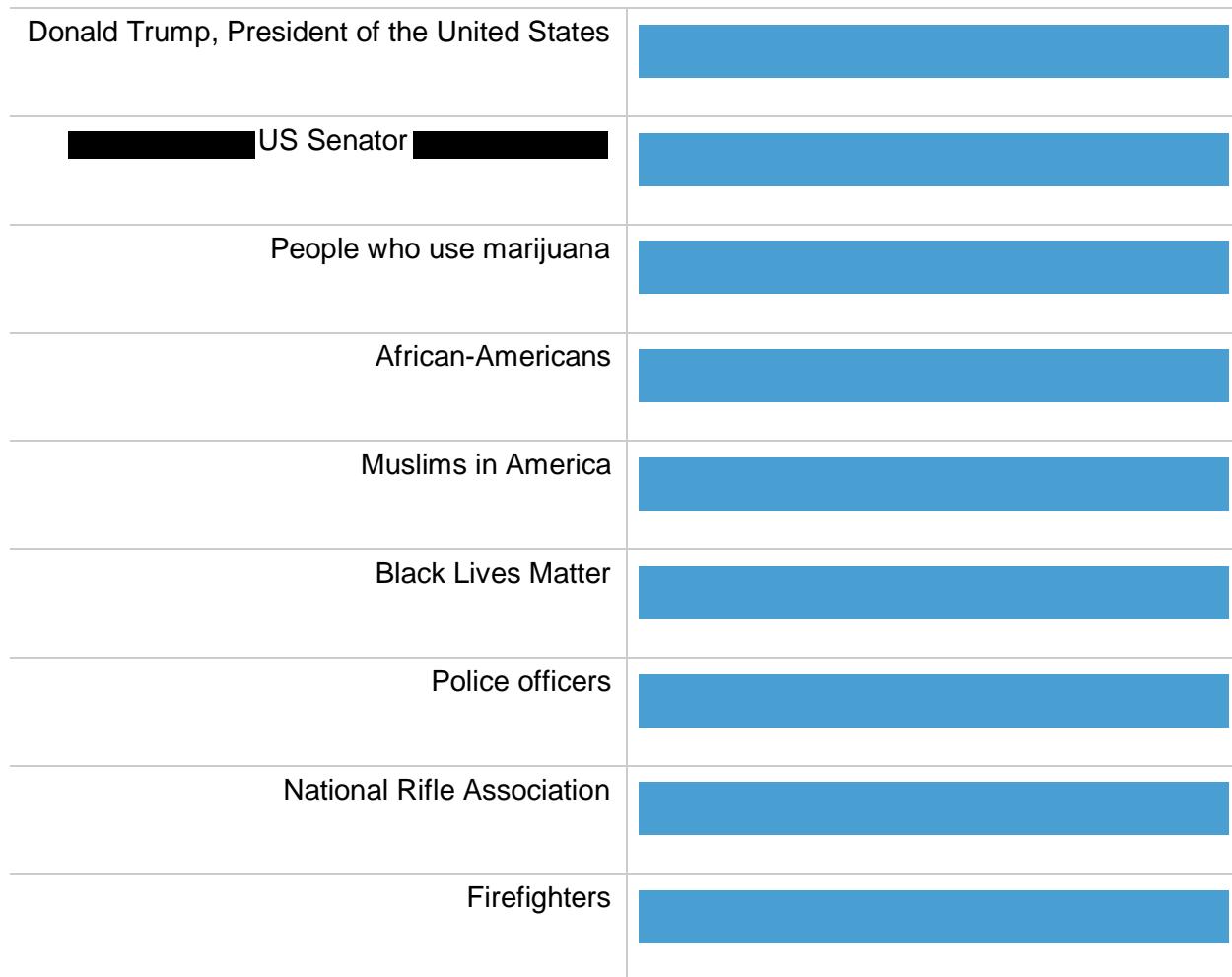
- Strongly agree
- Agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Disagree
- Strongly disagree

We'd like to get your feelings toward some people and groups who are in the news these days using something we call the **feeling thermometer**.

Here's how it works. You can choose any number between 0 and 100. The higher the number, the warmer or more favorable you feel toward that person or group; the lower the number, the colder or less favorable. You would rate the person or group at the 50 degree mark if you don't feel particularly warm or cold toward them. If you have never heard of the person or group, just check "Haven't heard of this person/group."

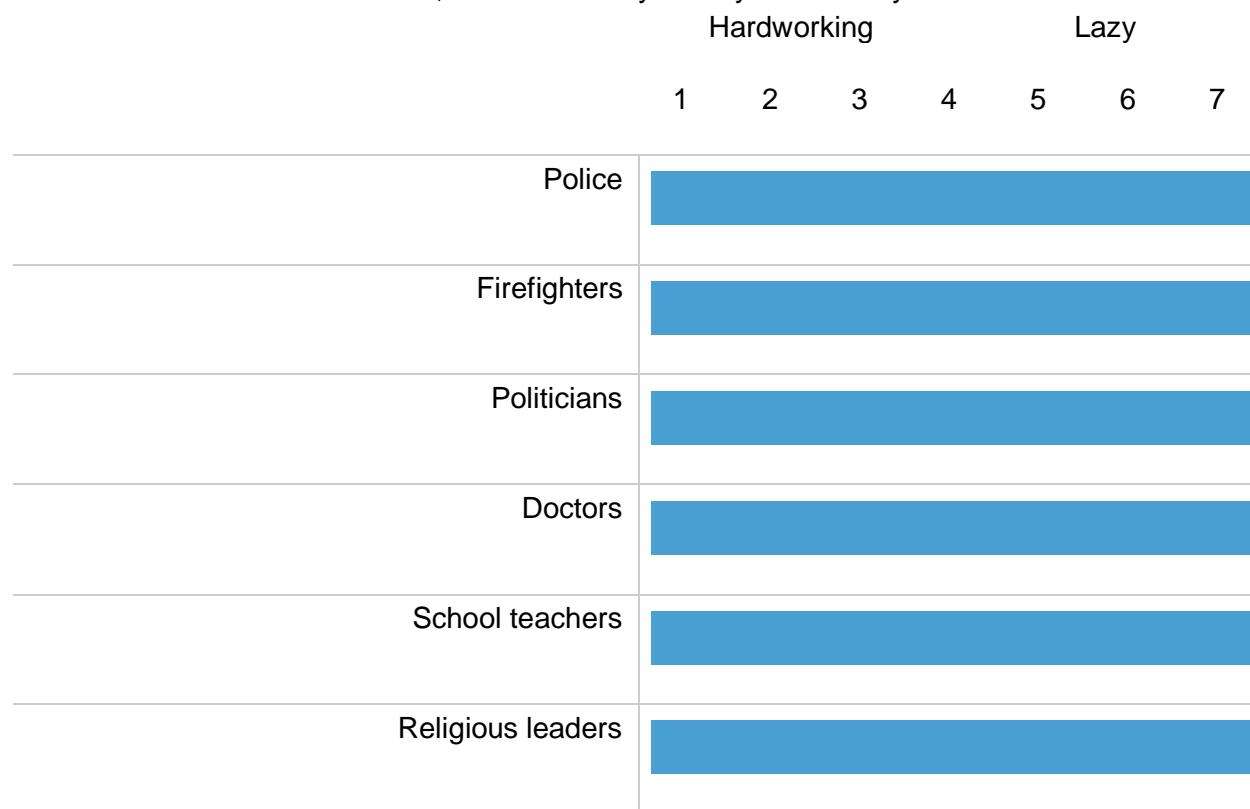
Haven't heard of this person/group

0 10 20 30 40 50 60 70 80 90 100

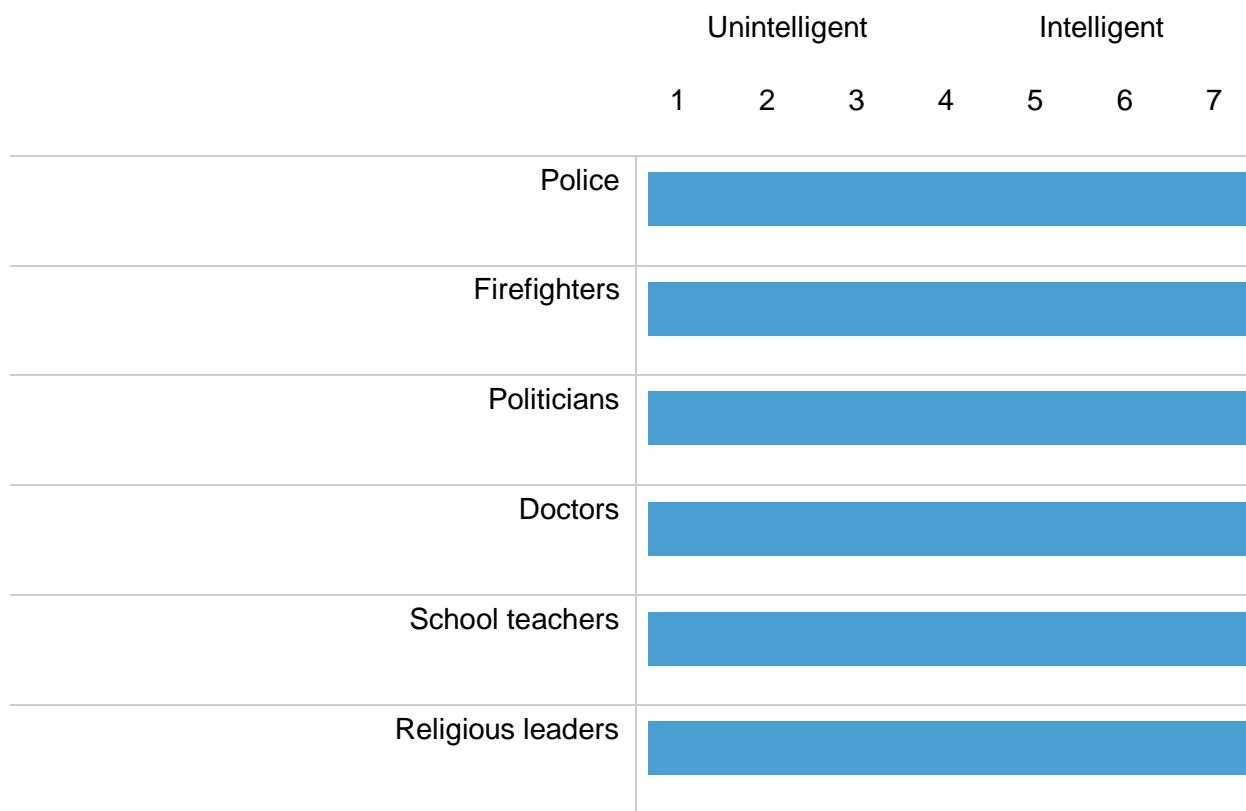


In this section, you will see a series of seven-point scales on which the characteristics of people in a group can be rated. Please choose the number that comes closest to where you think people in the group stand.

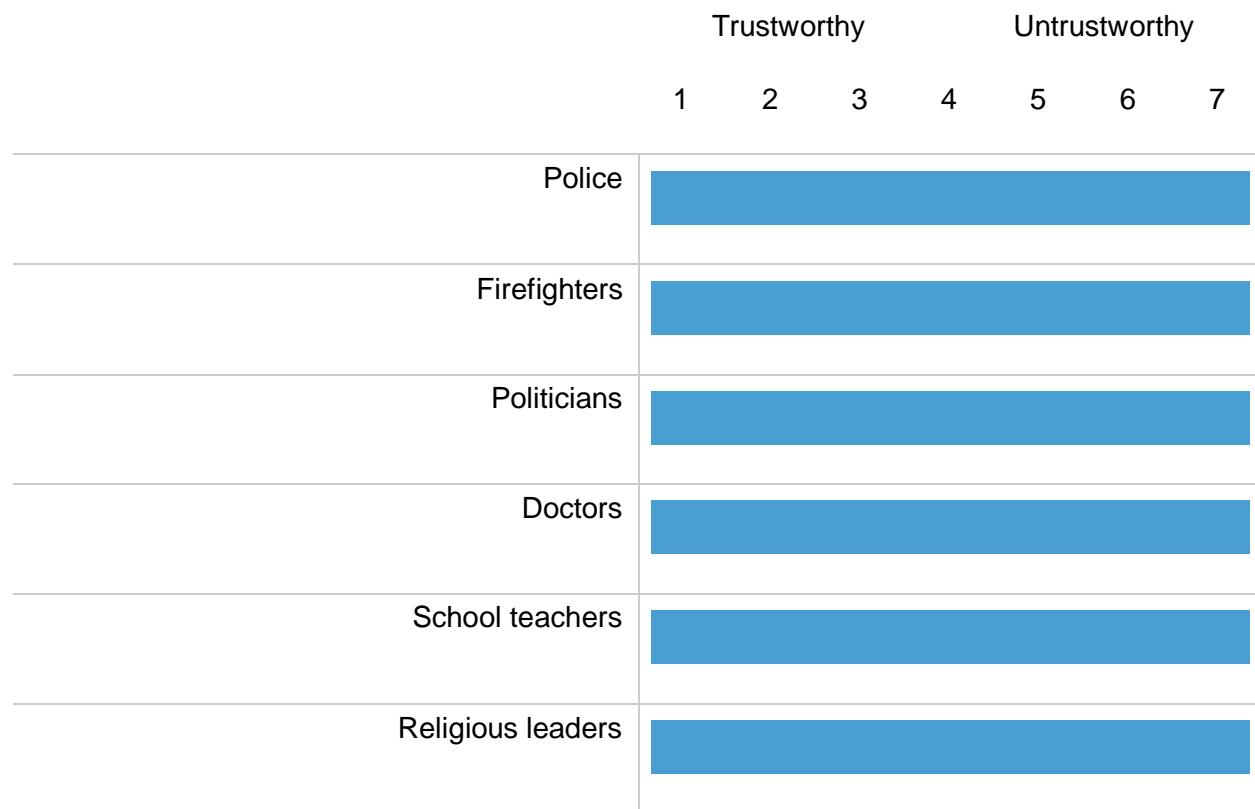
A score of 1 means you think almost all people in that group are "hardworking" and a score of 7 means you think all people in that group are "lazy". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



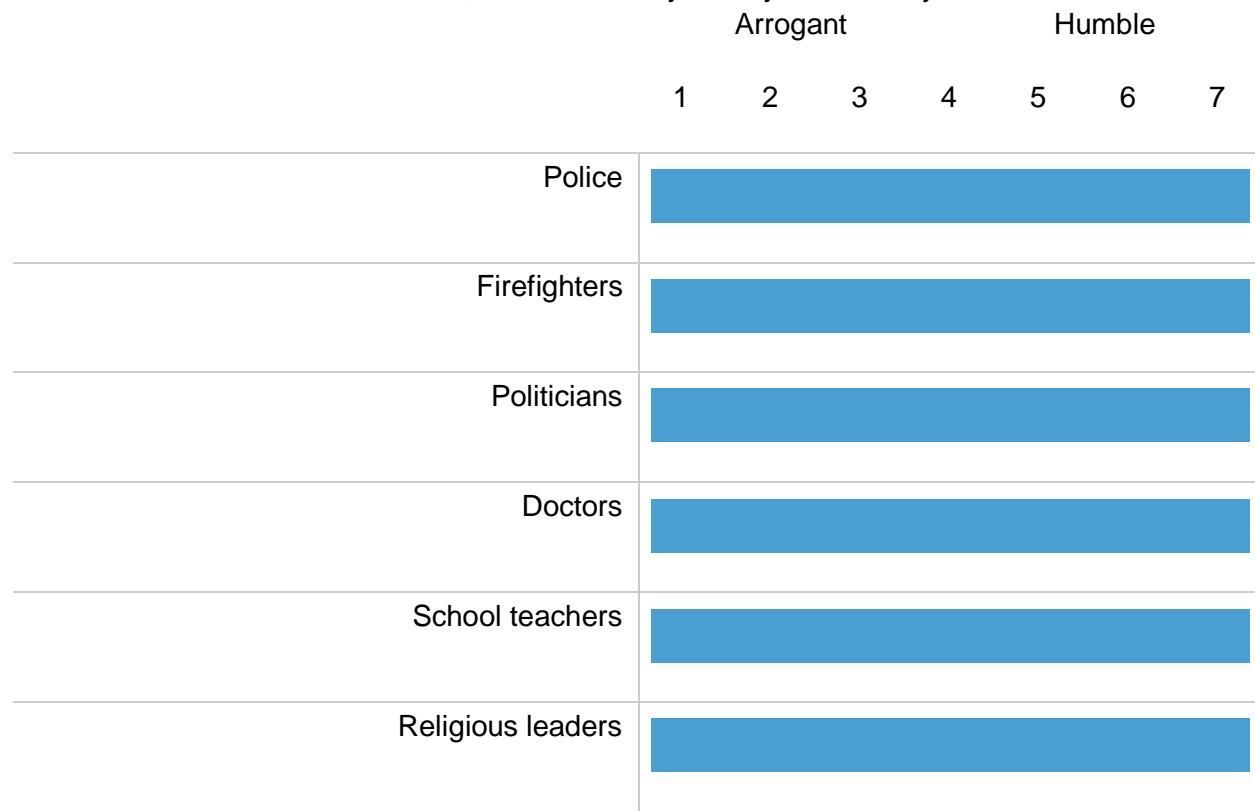
A score of 1 means you think almost all people in that group are "unintelligent" and a score of 7 means you think all people in that group are "intelligent". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



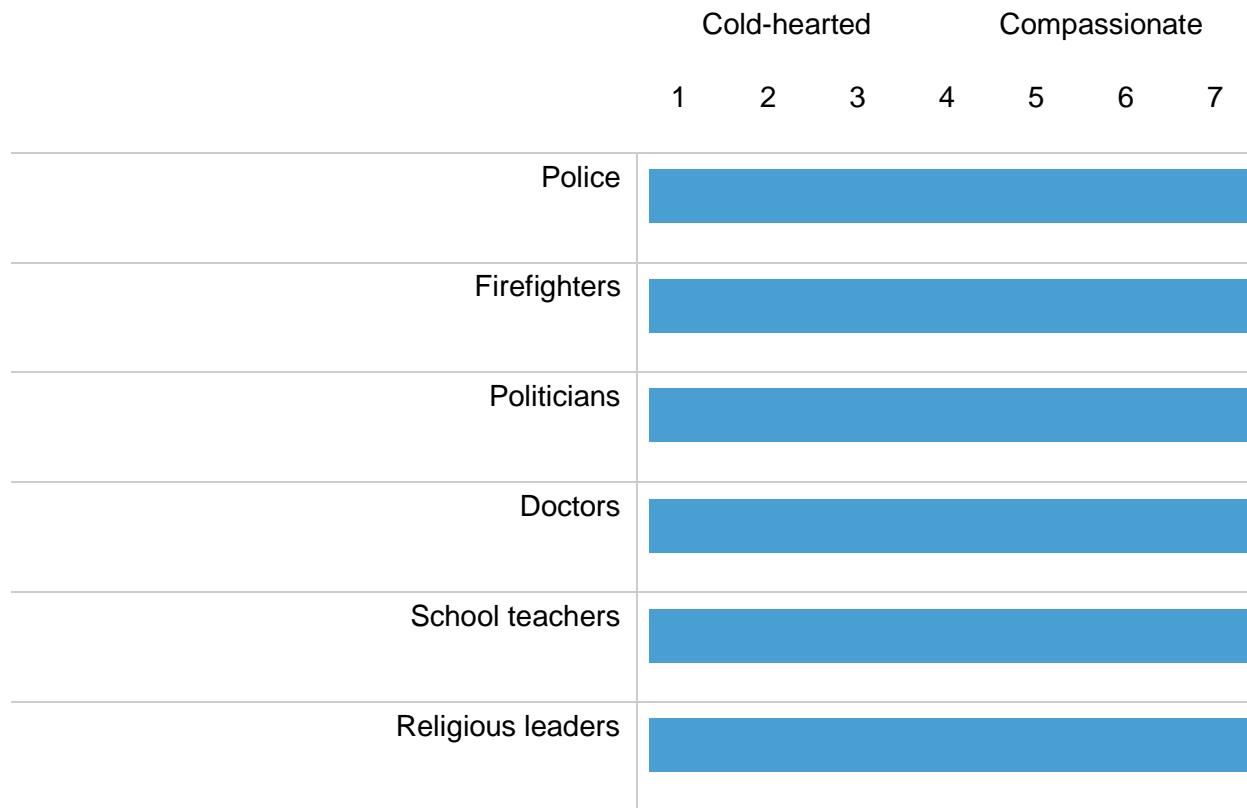
A score of 1 means you think almost all people in that group are "trustworthy" and a score of 7 means you think all people in that group are "untrustworthy". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



A score of 1 means you think almost all people in that group are "arrogant" and a score of 7 means you think all people in that group are "humble". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



A score of 1 means you think almost all people in that group are "cold-hearted" and a score of 7 means you think all people in that group are "compassionate". A score of 4 means you think the group is not towards one end or the other, and of course you may choose any number in between.



In the following section you will be asked a new set of questions. Please do your best to answer as accurately as possible, and please do not use the internet to search for answers.

The ages of Mark and Adam add up to 28 years total. Mark is 20 years older than Adam. How many years old is Adam?

If it takes 10 seconds for 10 printers to print out 10 pages of paper, how many seconds will it take 50 printers to print out 50 pages of paper?

On a loaf of bread, there is a patch of mold. Every day, the patch doubles in size. If it takes 40 days for the patch to cover the entire loaf of bread, how many days would it take for the patch to cover half of the loaf of bread?

If you're running a race and you pass the person in second place, what place are you in?
(Please enter a number; for example, 5 means "5th place")

A farmer had 15 sheep and all but 8 died. How many are left?

Emily's father has three daughters. The first two are named April and May. What is the third daughter's name?

How many cubic feet of dirt are there in a hole that is 3' deep x 3' wide x 3' long?

Thank you for your participation! To thank you for your time, we'd like to offer you \$5.00, or you can enroll in another raffle with a chance to win one of over a dozen \$75 prizes. Which do you prefer?

Please note: all \$5.00 payments will be delivered to your email address by TangoCard. Please allow 1-2 business days for payments to be processed, and check your spam box for emails from noreply@tangocard.com.

The raffle will be conducted once all respondents have completed the survey. The winners will be notified via email on or before 26th of November.

- A \$5.00 payment
- Another raffle

References and Notes

1. O of Community Oriented Policing Services, *Community Policing Defined* (Office of Community Oriented Policing Services, 2012).
2. T. R. Tyler, Psychological perspectives on legitimacy and legitimization. *Annu. Rev. Psychol.* **57**, 375–400 (2006).
3. T. R. Tyler, J. Jackson, Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. *Psychol. Public Policy Law.* **20**, 78 (2014).
4. P. J. Carr, L. Napolitano, J. Keating, We never call the cops and here is why: A qualitative examination of legal cynicism in three philadelphia neighborhoods. *Criminology.* **45**, 445–480 (2007).
5. M. Desmond, A. V. Papachristos, D. S. Kirk, Police violence and citizen crime reporting in the black community. *Am. Sociol. Rev.* **81**, 857–876 (2016).
6. J. Sunshine, T. R. Tyler, The role of procedural justice and legitimacy in shaping public support for policing. *Law Soc. Rev.* **37**, 513–548 (2003).
7. J. A. Shjarback, J. Nix, S. E. Wolfe, The ecological structuring of police officers' perceptions of citizen cooperation. *Crime Delinq.* 0011128717743779 (2017).
8. M. S. Phelps, D. Pager, Inequality and punishment: A turning point for mass incarceration? *Ann. Am. Acad. Political Soc. Sci.* **663**, 185–203 (2016).
9. H. Goldstein, Toward community-oriented policing: Potential, basic requirements, and threshold questions. *Crime Delinq.* **33**, 6–30 (1987).
10. C. M. Lum, C. S. Koper, *An evidence-assessment of the recommendations of the president's task force on 21st century policing* (International Association of Chiefs of Police, 2016).
11. O. W. Wilson, *Police administration* (McGraw-Hill, 1950).
12. J. Hawdon, J. Ryan, Police-resident interactions and satisfaction with police: An empirical test of community policing assertions. *Crim. Justice Policy Rev.* **14**, 55–74 (2003).
13. P. K. Manning, Community policing. *Am. J. Police*, 205–227 (1984).
14. C. Gill, D. Weisburd, C. W. Telep, Z. Vitter, T. Bennett, Community-oriented policing to reduce crime, disorder and fear and increase satisfaction and legitimacy among citizens: A systematic review. *J. Exp. Criminol.* **10**, 399–428 (2014).
15. R. K. Brunson, “Police don’t like black people”: African-american young men’s accumulated police experience. *Criminol. Public Policy.* **6**, 71–101 (2007).
16. J. Soss, Lessons of welfare: Policy design, political learning, and political action. *Am. Political Sci. Rev.* **93**, 363–380 (1999).
17. V. M. Weaver, A. E. Lerman, Political consequences of the carceral state. *Am. Political Sci. Rev.* **104**, 817–833 (2010).
18. A. White, Misdemeanor disenfranchisement? The demobilizing effects of brief jail spells on potential voters. *Am. Political Sci. Rev.* 1–14.
19. A. E. Lerman, V. Weaver, Staying out of sight? Concentrated policing and local political action. *Ann. Am. Acad. Political Soc. Sci.* **651**, 202–219 (2014).

20. J. Fagan, T. R. Tyler, Legal socialization of children and adolescents. *Soc. Justice.* **18**, 217–241 (2005).
21. J. H. Skolnick, *Justice without trial: Law enforcement in democratic society* (Wiley, 1975).
22. J. Goldstein, Police discretion not to invoke the criminal process: Low-visibility decisions in the administration of justice. *Yale Law J.* **69**, 543–594 (1960).
23. G. W. Allport, *The Nature of Prejudice*. Basic books, 1954.
24. T. F. Pettigrew, L. R. Tropp, A meta-analytic test of intergroup contact theory. *J. Pers. Soc. Psychol.* **90**, 751 (2006).
25. E. L. Paluck, S. A. Green, D. P. Green, The contact hypothesis re-evaluated. *Behav. Public Policy*, 1–30 (2018).
26. D. Broockman, J. Kalla, Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*. **352**, 220–224 (2016).
27. A. Scacco, S.S. Warren. Can social contact reduce prejudice and discrimination? *Am. Political Sci. Rev.* **112**, 654–677 (2018).
28. E. L. Paluck, D. P. Green, Prejudice reduction: What works? A review and assessment of research and practice. *Annu. Rev. Psychol.* **60**, 339–367 (2009).
29. National Academies of Sciences Engineering, Medicine, *Proactive policing: Effects on crime and communities* (The National Academies Press, Washington, DC, 2018).
30. M. A. Wycoff, W. G. Skogan, Pate Antony M., L. W. Sherman, A. Sampson, *Citizen contact patrol: The houston field test, executive summary*. (Washington, DC: Police Foundation, 1985).
31. R. Blair, S. Karim, B. Morse, Establishing the rule of law in weak and war-torn states. *Am. Political Sci. Rev.* <https://doi.org/10.1017/S0003055419000121>, 1-17 (2019).
32. Weisburd, D., Morris, N. A., & Ready, J. (2008). Risk-Focused Policing at Places: An Experimental Evaluation. *Justice Quarterly*, 25(1), 163–200.
33. D. Broockman, J. Kalla, J. Sekhon, The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. *Political Anal.* **25**, 435–464 (2017).
34. T. R. Tyler, J. Fagan, A. Geller, Street stops and police legitimacy: Teachable moments in young urban men’s legal socialization. *J. Empir. Leg. Stud.* **11**, 751–785 (2014).
35. M. Levi, A. Sacks, T. Tyler, Conceptualizing legitimacy, measuring legitimating beliefs. *Am. Behav. Sci.* **53**, 354–375 (2009).
36. Jackson, J. and Gau, J.M., Carving up concepts? Differentiating between trust and legitimacy in public attitudes towards legal authority. In *Interdisciplinary perspectives on trust* (pp. 49–69). Springer, 2016.
37. J. Mummolo, Militarization fails to enhance police safety or reduce crime but may harm police reputation. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9181–9186 (2018).
38. M. Levi, L. Stoker, Political trust and trustworthiness. *Annu. Rev. Political Sci.* **3**, 475–507 (2000).
39. J. M. Gau, R. K. Brunson, Procedural justice and order maintenance policing. *Justice Q.* **27**, 255–279 (2010).
40. T. R. Tyler, J. Fagan, Legitimacy and cooperation: Why do people help the police fight crime in their communities? *Ohio State J. Crim. Law*, **6**, 231–275 (2008).

41. L. Mazerolle, S. Bennett, J. Davis, E. Sargeant, M. Manning, Procedural justice and police legitimacy: a systematic review of the research evidence, *J. Exp. Criminol.* **9**, 245–274 (2013).
42. Reaves, B. A. *Census of State and Local Law Enforcement Agencies, 2008*. (Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice, 2011).
43. R. T. Moore, K. Schnakenberg, BlockTools. Version 0.5-6, August (2012).
44. A. Coppock, Randomizr: Easy-to-use tools for common forms of random assignment and sampling (2018; <https://CRAN.R-project.org/package=randomizr>).
45. K. Imai, K. Khanna, Improving ecological inference by predicting individual ethnicity from voter registration records. *Political Anal.* **24**, 263–272 (2016).
46. A. S. Gerber, D. P. Green, Field Experiments (WW Norton, 2012).
47. W. Lin, Standard operating procedures for Don Green’s lab at Columbia. Version 1.05 (2016) (available at https://github.com/acoppock/Green-Lab-SOP/blob/master/Green_Lab_SOP.pdf).
48. J. D. Angrist, J.-S. Pischke, Mostly harmless econometrics (Princeton university press, 2008).
49. G. W. Imbens, D. B. Rubin, Causal inference in statistics, social, and biomedical sciences (Cambridge University Press, 2015).
50. B. A. Olken, Promises and perils of pre-analysis plans. *J. Econ. Perspect.* **29**, 61–80 (2015).
51. D. P. Green, M. C. McGrath, P. M. Aronow, Field experiments and the study of voter turnout. *J Elections Public Opin Parties.* **23**, 27–48 (2013).
52. J. L. Kalla, D. E. Broockman, The minimal persuasive effects of campaign contact in general elections. *Am. Political Sci. Rev.* **112**, 148–166 (2018).