How Does Community Policing Affect Police Attitudes?

An Experimental Test and a Theory of Bureaucrat-Citizen Contact

American Political Science Review (Accepted)

Dotan Haim* Matthew Nanes† Nico Ravanilla‡

August, 2024

Abstract

In contrast to the expansive work on how community policing affects *citizens'* attitudes towards police, existing research says little about how community policing affects *officers'* attitudes towards citizens. We examine officer-facing outcomes using an experiment in the Philippines, in which a random subset of a province's 705 officers were assigned to intensive community policing activities for seven months. Treatment officers saw improved understanding of citizen concerns, but did not develop greater empathy or trust towards civilians, nor an increased sense of accountability for citizen-facing misconduct. We build from the experiment to develop an inductive theory of bureaucrat-citizen contact, relying on qualitative observations and exploratory analyses of heterogeneous effects. We propose that contact with citizens is only likely to improve attitudes among frontline bureaucrats who are not ex-ante embedded in their communities. Moreover, contact may have negative effects when it reveals threats to bureaucrats' personal safety.

Acknowledgements and Funding

This project would not have been possible without the participation of several key individuals, who we thank for their integral contributions to this research. Ms. Grace Labalan of IPA led the research team in Sorsogon and provided expert project management and oversight. General Ronaldo Cabral of the Philippine National Police provided vision, guidance, and dedicated service to his community by partnering with us in this research. Lt. Col. Malu Calubaquib of the Philippine National Police was instrumental in coordinating between the PNP and our research team, and in helping design public messaging around our intervention. We are grateful for helpful comments on this manuscript provided by Matthew Baum, Tanisha Fazal, Nahomi Ichino, Kelly Matush, Aidan Milliff, Michael Weintraub and audiences at Dartmouth College, University of California San Diego, Emory University, Florida State University, the 2021 International Studies Association, 2022 American Political Science Association, and 2022 Empirical Studies of Conflict meetings. We also thank Graeme Blair, Cesi Cruz, Craig McIntosh, Jeremy Weinstein, and members of the EGAP Metaketa project for their feedback on the research design. This research was generously funded by Evidence in Governance and Politics (EGAP), and by the UC San Diego Policy Design and Evaluation Lab (PDEL).

*Assistant Professor, Dept. of Political Science, Florida State University; dhaim@fsu.edu

†Associate Professor, Dept. of Political Science, Saint Louis University; matthew.nanes@slu.edu

‡Associate Professor, School of Global Policy & Strategy, UC San Diego; nravanilla@ucsd.edu

Tensions between police officers and citizens have reached historic highs in many places around the world. In the US, protests spread rapidly in response to high-profile deaths of Black citizens at the hands of the police. In the Philippines, a violent drug war left tens of thousands of people dead and eroded citizen trust. As a remedy to poor police-citizen relations, community advocates, police leadership, and politicians frequently call for "community policing." In this model, officers engage with citizens in informal settings to learn about community needs and build personal links that encourage trust (Skogan and Hartnett 1997). Existing studies of community policing focus almost exclusively on citizen-facing outcomes, finding mixed effects on citizen attitudes, crime reporting, and crime rates (Gill et al. 2014; Blair et al. 2021).

However, trust is a two-way street. Relatively little attention is paid to how *police officers* are affected by increased contact with community members, an omission that prevents a full understanding of how police-community relations can be sustainably improved. The guiding philosophy behind community policing requires that street-level police officers strengthen their understanding of local public safety concerns, reduce suspicions of citizens, and view citizen misconduct as a serious issue for which they will be held accountable (Mastrofski, Worden, and Snipes 1995; Cordner 1995). These attitudinal changes are a critical step towards breaking the "us versus them" culture among officers that sets the stage for shirking, abuse, and discrimination (Sparrow, Moore, and Kennedy 1990; Ingram, Terril, and Pauline 2018).

More broadly, despite the abundance of work on street-level bureaucrats' role as de facto policymakers (Lipsky 1980; Pepinsky, Pierskalla, and Sacks 2017), we have limited evidence on how different governance strategies affect state agents' perceptions of citizens. Our work contributes to scholarship on frontline service delivery by individual bureaucrats at the local level, especially in contexts with weaker government capacity for oversight and a higher incidence of patronage (Brierley et al. 2023). Existing work in this area focuses on how bureaucrats' implementation decisions are shaped by institutional structure and

oversight (Gulzar and Pasquale 2017; Mummolo 2018; Slough 2022), as well as a range of individual characteristics like bureaucrats' personality traits (**?**), local embeddedness (Tsai 2007; Bhavnani and Lee 2018; Haim, Nanes, and Davidson 2021), and shared identity with citizens on the dimensions like race, religion, and gender (Moser 1968; Meier and Nicholson-Crotty 2006; White, Nathan, and Faller 2015; Pfaff et al. 2021; Ba et al. 2021; Blair et al. 2022). The most common approach is to treat bureaucrats' characteristics as fixed traits that exogenously shape policy implementation. Instead, we test how one crucial characteristic of frontline bureaucrats *forms*.[1] In other words, we posit that bureaucrats' understanding of community concerns and their attitudes towards the community – key mechanisms through which "representative bureaucracy" and embeddedness are often argued to affect outcomes – are endogenous to the governance strategies and social contact with citizens that bureaucrats experience while on the job.

In this article, we develop and test our theory in two stages. In the first stage, we hypothesize that community policing will have positive effects on five officer-facing outcomes: 1) sharing citizens' public safety concerns, 2) trust in citizens, 3) empathy for citizens' concerns, 4) organizational accountability for citizen-facing misconduct, and 5) taking corruption seriously. We test these pre-registered hypotheses with a randomized community policing experiment in Sorsogon Province of the Philippines, a semi-rural province with a population of just under one million people. Of the approximately 705 eligible officers in Sorsogon Province, a random subset of 245 officers were assigned to participate in the community policing program, while officers in the control group undertook only their regularly assigned duties. This experimental design helps overcome inferential challenges to estimating the

[1]Research on the origins of bureaucratic characteristics tends to focus on the process by which bureaucrats are selected (Kaufman 1956; Kuipers 2022; Toral 2023). A closer analogy to our approach is Mo and Conn (2018), who find that Teach for America participants developed more favorable attitudes towards minority communities.

effects of community policing in an observational setting, where officers' and departments' selection into community policing likely correlates with pre-existing officer attitudes. Zooming out, this is one of the first field experiments in political science aimed at uncovering the determinants of street-level bureaucrats' attitudes.[2]

We find that participation in community policing led to a 24% increase in overlapping safety concerns identified by both officers and citizens, but did not result in improvements in officers' empathy, trust, accountability, or views of corruption. Our model is sufficiently powered to rule out substantively large effects on officer attitudes: on our composite index of officer attitudes, the largest effect that falls within the 90% confidence interval is a 4% improvement in attitudes. These findings suggest promising evidence that officers can learn valuable information through community policing, but they also introduce an important puzzle regarding why this information did not translate to improved attitudes.

The second stage of our project addresses this puzzle. We develop an inductive theory of why, and under what conditions, officer-citizen interactions may fail to positively affect street-level officers' attitudes. Conversations with our field staff who attended thousands of community policing meetings, along with our own interviews with officers, led us to develop a better sense of how officer attitudes were shaped by the intervention, which we corroborated with exploratory analyses of heterogeneous effects. We theorize that officers' participation in community policing is a form of interpersonal contact with an outgroup. Consistent with the "contact hypothesis" literature, increased interactions with outgroups can generate different results depending on the context of those interactions (Allport 1954; Pettigrew 1998). Specifically, we posit that community policing can shape officers' attitudes in different ways depending on 1) their pre-existing "embeddedness" in the community and 2) whether their experiences raise the salience of threats to officer safety.

---

[2]Blair et al. (2021) offers a notable exception. Before presenting our results, we elaborate on how our article builds on this work.

First, regarding embeddedness, we propose that increased contact with citizens is only likely to improve attitudes among officials who are not already highly familiar with the communities they are tasked with serving. Increased contact with community members will only change officers' attitudes when it provides information that contradicts misconceptions, reduces previously-held anxieties about interacting with citizens, and overcomes an inability to understand local citizens' perspectives. These conditions are less likely to be met for embedded officers, who already experience extensive contact with citizens during their daily lives. Our data show that participation in community policing did not change the attitudes of officers who lived in the same province where they served, but substantially improved attitudes among officers from outside the province.

Second, we contend that the safety context where officers are charged with implementing community policing can determine the direction of attitudinal shifts. Interpersonal contact with citizens in "safe" communities can help dispel officers' negative views, but these same interactions can have counterproductive effects in communities where they raise the salience of threats to officer safety. Consistent with this view, treatment officers who participated in community policing in places where insurgents were active saw reduced trust in citizens, while officers who participated in areas without insurgents experienced a marginally significant increase in trust. An important implication is that rather than *overcoming* the effects of neighborhood safety context as is commonly assumed, community policing can instead *heighten* the effects of neighborhood context on officer attitudes.

This project contributes to a more complete understanding of the way that contact between citizens and government agents affects government service provision. Our findings help explain why evidence of community policing's effectiveness is inconsistent at best, especially in the Global South (Weisburd and Eck 2004; Blair et al. 2021). By identifying contexts where community policing might backfire, we do not suggest that the solution is for the police to become less engaged, but rather that different police reforms may be more appropriate in these contexts, perhaps in combination with community policing. This re-

search also contributes to literatures on other types of frontline government agents, including bureaucrats and military personnel. For example, if engaging in "hearts and minds" style counterinsurgency in dangerous areas undercuts trust among the soldiers and development personnel charged with extending services, this may increase their propensity to use force against civilians or distribute services unfairly (Lyall, Blair, and Imai 2013). A key takeaway is that in order to understand the citizen-facing effects of strategies to improve governance, it is crucial to understand the effects on the state agents charged with implementing these strategies.

Theory Part 1 (Deductive)

Community policing generates frequent, informal interactions between police officers and citizens during town-hall meetings, citizen-facing foot patrols, and problem-solving sessions with community leaders, with the intent of generating mutual trust and understanding between both parties (Skogan and Hartnett 1997; Skogan 2004). In this section, we develop our theory for how participating in community policing should affect officer attitudes. We begin by defining five attitudinal outcomes, focusing on the attitudes that research shows are most likely to translate to officer behaviors that shape public safety.

Why Officer Attitudes Matter

Our work builds on a substantial body of research that examines the causes and effects of public servants' attitudes. Work in public administration focuses on two categories of attitudes that correlate with the quality of services bureaucrats provide: bureaucrats' assessment of clients and their adherence to bureaucratic mission (Keiser 2010; **?**; Maynard-Moody and Musheno 2003). In political science, the most common approach is to treat these attitudes as *mechanisms* and proxy for their existence using data on bureaucrats' demographic characteristics, though there exist some notable exceptions that measure bureaucrats' attitudes

directly (Sabatier, Loomis, and McCarthy 1995; Kuipers 2022; Toral 2023). Applied to the study of policing, research tends to focus on three aspects of client assessment that we label 1) *sharing citizen concerns*, 2) *trust* in citizens, and 3) *empathy* for citizens' concerns. When it comes to adherence to bureaucratic mission, policing scholars are especially attuned to officers' views on 4) organizational *accountability* for misconduct, and 5) the seriousness of misconduct, especially *corruption*. Variation in these attitudes is important because patrol officers have substantial discretion over when to enforce laws and what sanctions to apply, similar to the discretion afforded other street-level bureaucrats (Lipsky 1980). In particular, officers' attitudes correspond to policing practices such as the quality of communication, the amount of effort devoted to investigating citizen tips, the prioritization of some citizens over others, and the propensity to harass or harm citizens.

Understanding and *sharing citizen concerns* is often considered the first step towards officers being able to do their job effectively. When officers are in touch with the issues that are important to citizens, they can better allocate their time to activities that make a substantial dent in improving order and earning citizens' trust (Skogan and Hartnett 1997). Moreover, a better understanding of citizens' concerns is considered to be the key mechanism that leads to improvements in other officer attitudes (Paoline, Myers, and Worden 2000). When officers better understand citizens' perspective on public safety issues, they are less likely to resort to harmful stereotypes of civilians as being unpredictable and suspicious.

Next, research devotes substantial attention to officers' *trust* in citizens, a concept that is particularly prevalent in work on "police culture." We conceptualize trust as a range of officer beliefs that constitute an "us versus them" mentality, including perceptions of citizen intentions and the threat posed by citizens (Sparrow, Moore, and Kennedy 1990). For example, Paoline (2003) defines a key aspect of police culture as the "prescription to be suspicious and maintain the edge over citizens" due to the "terror of a hostile and unpredictable citizenry." Trust in citizens varies greatly across officers, and this variation has strong links to officer behavior (Paoline 2003). For example, studies find that suspicion towards citizens

is correlated with officer abuse and discrimination (Ingram, Terril, and Pauline 2018), as well as lower productivity (Moon and Zager 2007).

A third set of client-facing attitudes relates to officers' *empathy* for citizens' concerns, which we define as the belief that citizens' concerns and complaints are warranted. While this captures only a narrow subset of the overarching concept of empathy, it comports with the mechanisms theorized to link officer empathy with improved policing outcomes. Inzunza (2015) outlines two fundamental components of empathy that are salient in the policing literature. The first is whether officers place themselves in citizens' shoes when hearing their reports and complaints, which includes both affective (feeling what others feel) and cognitive (perspective-taking) dimensions of empathy (Zaki 2014; Todd and Galinsky 2014). This leads to a second, behavioral manifestation of empathy, whereby officers outwardly convey that they are receptive to citizen-provided information (Barrett-Lennard 1981). When officers convey that reports and complaints are justified and worth taking seriously, citizens – especially those from marginalized groups and those who experience sexual violence – are more likely to report positive contact with officers and follow through on initial reports (Birzer 2008; Maddox, Lee, and Barker 2011; Turgoose et al. 2017).

Finally, scholarship suggests that officer behavior is shaped by two aspects of their adherence to bureaucratic mission. The first regards officers' perception of organizational *accountability*. When officers believe that police leadership can monitor citizen-facing misconduct, takes citizen complaints seriously, and would seriously reprimand misbehavior, officers may begin treating citizens as partners rather than adversaries (Cordner 1995; Crank 2014). Accountability links to a second important concept, which is whether officers internalize that harming citizens constitutes serious misconduct. This includes officers' views on the seriousness of activities like *corruption*. The perception of organizational accountability can make adversarial comments towards citizens a taboo topic, leading to a change in norms about the appropriateness of misconduct (Haarr 2001). The literature suggests that improvements in officer behavior are more sustainable when officers begin treating issues surrounding mis-

conduct in accordance with a "logic of appropriateness," independently of whether they are worried about being punished for bad behavior (Sparrow, Moore, and Kennedy 1990; March 1994).

How Community Policing Influences Officer Attitudes

In the absence of regular, informal contact with citizens, a reliance on reactive policing gives officers a skewed view of the community by disproportionately bringing them into contact with citizens when a crime has occurred. Officer attitudes tend to depend heavily on rough indicators of "neighborhood context," such as the demographic makeup, homocide rate, or reputation for citizen resistance to the police (Terrill and Reisig 2003). Prior research suggests that officers commonly underestimate the proportion of the community that supports their efforts to improve public safety (Skolnick 2011). These misperceptions can lead to alienation from the community and a feeling that citizens do not share the officer's values (Twersky-Glasner 2005). In short, it is easy for officers to resort to harmful stereotypes of the typical citizen's experiences, intentions and behaviors.

A core tenet of community policing is that positive, informal interactions with citizens can overcome these misperceptions (Paoline, Myers, and Worden 2000). When it comes to client assessments, community policing should first and foremost provide officers with information that leads them to *share citizens' concerns*. Officers' priorities may shift as they understand that issues appearing relatively minor to them are actually significant problems for citizens. This should translate to improvements in officer *trust* by giving them a more accurate view of the prevalence of threats to officer safety and the proportion of citizens who support police efforts. These interactions should also allow officers to understand citizen

motivations when they choose to report issues to the police, generating greater *empathy* and leading officers to take citizen reports more seriously (Schuck and Rosenbaum 2005).[3]

A second key goal of community policing is to signal increased organizational *accountability* and change officers' beliefs about the seriousness of misconduct like *corruption*. By implementing community policing, police leadership signals that they care about the quality of citizen engagement (Cordner 1995). Community policing also typically involves strengthening procedures by which officers can be held accountable for misconduct, including efforts to emphasize the importance of reporting fellow officers. Moreover, when officers are placed in more frequent contact with law-abiding citizens, the hope is that, over time, they will internalize the seriousness of misconduct and the importance of serving people rather than on 'doing policing' (Bracey 1992).

Drawing from the literatures discussed above, we propose five hypotheses for how community policing should shape officer attitudes towards citizens:[4]

> **H1 (Shared Concerns):** Officers who participate in community policing will become more likely to share an understanding of citizens' public safety concerns.

> **H2 (Trust):** Officers who participate in community policing will become more likely to express trust in community members.

---

[3]The influential "Final Report of the President's Task Force on 21st Century Policing" (2015) advocates for community policing because it spurs officers to engage in perspective-taking and treat citizen reports with greater seriousness.

[4]The pre-analysis plans specifying these hypotheses are available at `https://osf.io/tfpx7/` and `https://osf.io/zxefj/`. The *Trust* hypothesis is not included in our pre-analysis plan. We discuss this oversight, as well as evidence that we ex-ante intended to include this outcome, in a broader introductory memo to the PAP found in the SI.

H3 (Empathy): Officers who participate in community policing will become more likely to express empathy for the seriousness of citizen reports.

H4 (Accountability): Officers who participate in community policing will feel more accountable for citizen-facing misconduct.

H5 (Corruption): Officers who participate in community policing will be more likely to view corruption as serious misconduct.

A small set of existing studies aims to empirically test a similar set of hypotheses regarding the effects of community policing on police attitudes. A first group of observational studies provides useful evidence that is nevertheless susceptible to various threats to causal inference. Paoline, Myers, and Worden (2000) compare attitudes between specialized community policing officers and regular beat officers in two US police departments. Hayeslip and Cordner (1987) and Rosenbaum, Yeh, and Wilkinson (1994) both use longitudinal research designs, tracking officer attitudes before and after they implemented community policing. All three studies find that community policing corresponded with improved officer attitudes towards citizens. However, these results may be biased if the same factors that drove officer and departmental decisions to adopt community policing are correlated with officer attitudes, or if other changes that occurred concurrently with community policing programs shaped officer attitudes.

Another relevant study is Blair et al. (2021), who implemented a set of community policing experiments in the Global South and report null results of community policing on officer attitudes, including empathy, corruption, and accountability.[5] We see our work as building on (rather than contradicting) the results of Blair et al. (2021), which foregrounds

[5]Our outcome measures capturing H3-H5 are consistent with the measures used in Blair et al. (2021).

the effects on crime and citizen-facing outcomes. At a basic level, we contribute to the advancement of research on this topic by elucidating the theory behind the predicted effects on officers and incorporating two additional outcomes (*Shared Concerns* and *Trust*) that are crucial to our interpretation of the results. More importantly, our inductive theory and the evidence we present in the latter half of the paper can help scholars and policymakers interpret *why* and *under what conditions* social contact with citizen may fail to translate into improved attitudes, both among police officers and bureaucrats more generally.

Context and Research Design

We evaluate the impact of community policing on police officers' attitudes towards citizens in Sorsogon Province of the Philippines. Sorsogon is one of 81 provinces in the country and it encompasses 541 "barangays," an administrative akin to a neighborhood or village. The population of about one million people lives in a few semi-urban centers, along with a range of coastal, lowland, and upland rural barangays. Sorsogon's poverty rate of 21.6% is just above the national average of 18%, and the province displays very little ethnic or religious diversity; 95% of the population is Roman Catholic.

Public safety provision is handled by two institutions. The Philippine National Police (PNP) is the national-level policing institution, though leadership at the provincial level is granted significant autonomy over policing practices and resource utilization. The PNP is relatively well-equipped compared to other lower middle-income countries, but they lack the manpower to fully patrol Sorsogon's broad and diverse geography. Before we implemented the community policing initiative described below, more than 45% of surveyed Sorsogon residents said they saw a PNP officer once per month or less. To fill these gaps, barangays employ a chief "tanod" and 8-20 regular tanods, semi-professional civilian safety officers responsible for preventing minor crimes, resolving disputes, and helping citizens communicate

more serious issues to the PNP. Tanods are jointly appointed by the elected barangay captain and the barangay "kagawad" (councilor) in charge of peace and order.

Residents in Sorsogon face a variety of public safety issues. The most common are petty theft, neighborhood/family disputes, vehicle accidents, public intoxication, and juvenile delinquency, with the latter two issues often translating to harassment, trespassing, and property damage. More severe crimes like assault, armed robbery, rape, and murder occur at rates similar to other semi-rural provinces in the country. Finally, Sorsogon is a hotspot for the New People's Army (NPA), a nationwide communist insurgency that has been active in the province since the late 1960s. According to military intelligence reports, the NPA was active in every municipality in Sorsogon in 2013, but withdrew from about one-third of these municipalities by 2015. As we discuss below, practical and ethical considerations led us to narrow our sample to 298 barangays (out of 541) with limited NPA presence. NPA members still travel through these areas and attempt to identify potential sympathizers, but violence is rare. In these areas, the PNP, rather than the military, is the primary government security provider that attempts to deter NPA activity.

The "One Sorsogon" Community Policing Intervention

Sorsogon's community policing intervention began as a reform initiative led by the Provincial Police Chief. He wanted to implement a community-centered approach to contrast with the violent "War on Drugs" happening elsewhere in the country. Even though virtually no drug-war related violence occurred in Sorsogon, the Police Chief expressed concern that the PNP's local reputation was harmed by the national-level environment. Prior to the intervention, he implemented a number of small-scale programs aimed at bolstering the service-oriented side of policing, such as a program to engage with at-risk youth in schools. However, practically all community engagement activities prior to the intervention were under the domain of a designated "Police Community Relations" (PCR) Director at each municipal station, along with a handful of PCR officers at the Provincial Police Office, none of whom were eligible

for our intervention. Regular beat officers in Sorsogon almost never engaged in the usual components of community policing, such as town halls or meetings with community leaders.

The new initiative, called "One Sorsogon," spanned seven months and centered on directing rank-and-file officers to work closely with barangay leaders and ordinary citizens to identify the issues that most threatened public safety in their communities, and devise locally appropriate solutions together. We randomized the selection of PNP officers who participated in the community policing activities, which is the design feature we use to draw inferences in the present study.[6] Control officers did not participate in any of the activities described below, and experienced policing-as-usual for the duration of the study.

The program was a bundled treatment that kicked off with a two-day training on best practices for community policing, led by a professional consultant.[7] At the training, treatment officers were paired with the barangay leaders in their assigned barangay (who were also in attendance), practiced using crime reports and tanod logbooks to identify the most pressing local public safety issues, learned techniques for empathetic communication, and focused on recognizing and reporting officer misconduct, including corruption.

Treatment officers then attended six monthly meetings with leaders of their assigned barangay, lasting just under two hours on average, and held at the municipal police station. The first three meetings focused on identifying the most pressing local public safety problems, devising a plan to address those issues, and proposing a budget of up to 5,000 pesos (about USD $100) to help with implementation, while the next three meetings focused on implementing this plan and monitoring its progress. This aspect of community policing is

---

[6]We discuss intervention details, prior policing practices, and other aspects of the randomized trial in greater detail in the SI.

[7]The community policing programs assessed in the literature tend to involve trainings with similar length and content (e.g. Paoline, Myers, and Worden 2000).

known as "Problem-Oriented Policing" (POP) (Goldstein 1990; Weisburd et al. 2010). In our case, POP teams were composed of two PNP officers, the barangay captain, the barangay councilor in charge of peace and order, the chief tanod, and 2-4 regular tanods. The POP aspect of the intervention was designed to enhance interactions with barangay leaders, first and foremost. We conceive of these leaders as being comparable to an important class of "community leaders" focused on by prior research on community policing, who typically have a more complete understanding of crime in their community and can better assist the police in implementing reforms. From the perspective of the PNP, barangay leaders were discussed as "citizen representatives," and they play a similar role to informal village leaders and community watch members in other parts of the Global South.

Between meetings, the officers assigned to One Sorsogon were directed to engage with a broader range of citizens in the communities where they were assigned. Officers regularly traveled to their assigned barangay to engage in citizen-facing foot patrols, gather citizen perspectives, and disseminate information about police activities. The implementation strategies of the POP teams also tended to directly involve officers, for example by building community gardens and passing out sports equipment with at-risk youth. Treatment officers met monthly with their station's police chief and PCR director to discuss the progress of the intervention, and were encouraged to surface reports of police misconduct during these meetings. Finally, the intervention concluded with a community town-hall in each barangay, during which PNP officers received input from the broader community and communicated their plans going forward. On average, each culminating town hall was attended by 80 citizens and lasted more than three hours.

Our study is best interpreted as an evaluation of community policing as it is actually implemented in many real world contexts, especially those outside of the US, Australia, and Europe, where most studies of community policing take place. Similar to many countries with moderate or weak police capacity, the Sorsogon PNP faces pressures associated with difficult terrain, weak infrastructure, under-staffing, decentralization of resources, and leadership

rotations that disrupt policy continuity. Under these circumstances, community policing programs (including those supported by international donors like the UN Department of Peacekeeping) tend to be localized, limited to several months of intensive implementation, and involve officers who are also assigned to other duties. We note that One Sorsogon was a homegrown initiative, and we did not alter the duration of the program originally specified by the Provincial Chief. In this context, the intervention was relatively intensive, and many PNP officers expressed that One Sorsogon represented a wholesale change in the scope of their community-facing duties. That said, we acknowledge that community policing may produce different effects if paired with an overhaul in resources, staffing procedures, or long-term commitments from national leadership.

Ethics

By the time we approached the Sorsogon PNP about partnering on an evaluation of One Sorsogon, we had spent approximately one year conducting observational research on policing in the province. We previously conducted a large survey on policing and developed an extensive network of local academics, citizens, community leaders, police officers, and individuals with connections to the NPA, all of which helped us to better understand the risks of the study and the ways we could mitigate those risks. To briefly summarize three key points, we learned from our contacts in the area that the Sorsogon Police Chief was a well-known reformer who opposed the violent tactics of President Duterte's Drug War, which primarily took place in major urban areas. We confirmed the low rate of Drug War related violence in Sorsogon using surveys, conversations with local human rights activists, and data from ACLED's special project on the Philippine Drug War. We also relied heavily on our local network in each municipality to identify and exclude areas where NPA presence was strong enough to make the intervention unsafe. Finally, our research staff attended every POP meeting and town hall, and conducted extensive audits in each barangay to detect potential

abuse. We expand on these points, as well as other aspects of research ethics and human subjects protections, in the SI.

Experimental Design

The Provincial PNP Chief agreed to randomize several aspects of the One Sorsogon intervention, including the barangays where it was implemented: Of the 298 study barangays, 99 were randomly selected to receive the PNP's community policing program. In the present study, we draw empirical leverage from the randomized assignment of *officers* to these barangays.[8] From the list of 705 eligible PNP officers serving in Sorsogon, we randomly selected 198 officers to participate in the community policing intervention, two for each participating barangay. Blocking took place at two levels. First, we blocked on police station (from among the 17 stations in the province) to assign treatment officers to a POP team in a barangay within their station's jurisdiction. We also blocked on rank, assigning one senior and one junior officer to each team.

A number of officers initially assigned to treatment later had to withdraw from their POP team, mostly due to being reassigned to national-level training programs or a different station. When an officer was reassigned, they were replaced by the next officer on our list of randomly selected officers (in the same station and of the same rank). In total, 47 randomly selected replacement officers were assigned to POP teams during the course of the intervention. Most replacements occurred during the first week of the intervention due to a region-wide training and replacement cycle. As a result, the average replacement officer was

[8]The officer randomization only occurred within the treatment barangays in the larger study, meaning we cannot leverage the barangay-level randomization for analyzing effects on officers. Eligible officers included those with the rank Police Officer (PO) or Senior Police Officer (SPO). Non-uniformed personnel and upper-level police leadership were excluded.

assigned to attend 5.5 of the 7 community policing meetings. Our main results focus on the Intent to Treat effect (ITT), meaning that 245 officers (198 original plus 47 replacements) are considered treated. Treatment officers displayed a high degree of balance with control officers on their demographic characteristics, as well as their baseline responses to the survey outcomes (SI Table A.3).

Outcomes

We surveyed officers before and after the intervention and constructed indices of *sharing citizen concerns* (H1), *trust* in citizens (H2), *empathy* for citizen concerns (H3), perceptions of organizational *accountability* for citizen misconduct (H4), and perceptions of *corruption* as serious misconduct (H5).[9]

Sharing citizens' concerns (H1) was measured by asking officers what they thought were the three most important public safety concerns (from a list of 13 items) in the municipality where they were assigned, and then matching these responses with the four most common responses to the same question among community leaders who lived in that municipality.[10] Table 1 shows that, at baseline, citizens were far more likely than the police to identify minor crimes like public intoxication, theft, and illegal gambling. The police, on the other hand, were more likely to identify crimes like rape, illegal drug use, and murder. We expect that the *Shared Concerns* measure is correlated with broader factual knowledge

[9]The text of each survey item and details about index construction are located in the SI. All outcome measures (including the umbrella *Officer Attitude Index* described later) were pre-registered except the *Trust* index, which was an unintentional omission we detail in our SI memo introducing the PAP.

[10]In the SI, we show that the results look similar when using several alternative versions of the *Shared Concerns* measure.

about the community, though limited survey space precluded asking additional questions measuring factual knowledge. For example, officers expressed to our research team that participating in the intervention gave them a better sense of which barangay leaders were the most effective local partners for motivating community action.

Table 1: Most important public safety issues according to officers and citizens

Issue	**PNP**	**Citizens**	**Difference**
Public Intoxication	15.1%	37.2%	-22.1%
Theft	22.1%	36.5%	-14.4%
Illegal Gambling	30.1%	39.5%	-9.4%
Police Abuse	2.6%	4.3%	-1.7%
Sexual Harassment	9.1%	10.7%	-1.6%
Robbery	4.7%	5.8%	-1.1%
Vehicle Theft	2.8%	3.7%	-0.9%
Vehicle Accidents	55.5%	55.2%	0.3%
Illegal Guns	6.1%	1.8%	4.3%
Domestic Abuse	12.7%	4.6%	8.1%
Murder	36.6%	26.5%	10.1%
Illegal Drug Use	61.1%	48.7%	12.4%
Rape	41.5%	16.1%	25.4%
Observations	*773*	*2,983*	

Citizens and officers were each presented with the above list of public safety issues and asked to choose which three were the most important concerns in their municipality. The listed percentages show how often each issue was included in a respondent's top three.

To capture the attitudinal outcomes in H2-H5, we used indices of several questions on the officer survey. Answers were captured using a slide bar that translated to a 100-point scale. The *Trust* index is composed of three questions: 1) citizens care about officer well-being, 2) citizen-provided information is likely to be accurate, and 3) how safe for officers is the municipality where you are assigned. We included the question on safety in the trust index based on literature that links officers' perceptions of safety to their suspicions of the citizenry. Acknowledging that other factors may shape officer perceptions of safety, we separately report results for the individual components of the *Trust* index. The *Empathy* index is composed of two questions: 1) when people complain about the police, it is because

they have a good reason, and 2) most things people report to the police are worth taking seriously. With the recognition that empathy is a broader concept than what we measured here, these questions are intended to capture one theoretically important aspect of how empathy manifests in policing, described in our theory section.

The *Accountability* and *Corruption* indices are both composed of questions about two hypothetical scenarios involving police misconduct. In one scenario, an officer accepts a bribe in exchange or not issuing a citation; in the other, an officer routinely accepts unsolicited gifts while on patrol. After respondents are presented with these scenarios, they are asked whether 1) they would report this behavior themselves, 2) they think their peers would report this behavior, and 3) the extent to which police leadership would punish this behavior. Along with a separate question (unrelated to the hypothetical scenarios) on 4) whether police leadership takes citizen complaints about officers seriously, these questions compose the *Accountability* index. Officers were also asked whether 1) they consider the behavior to be serious misconduct and 2) most other police officers consider it serious misconduct. These questions compose the *Corruption* index. Our focus on corrupt misconduct aligns with concepts raised during the community policing training and monthly check-ins with station leadership. It also accords with the PNP's nationwide focus on tackling corruption in the lead up to the intervention period.[11]

Finally, to capture officers' overall attitudes towards citizens, we construct a composite *Officer Attitude Index*, which is an "index of indices" incorporating all the attitudinal outcomes described above. To create each index, we standardized and summed all items,

---

[11]See: https://www.aljazeera.com/news/2017/1/30/philippine-police-suspend-drug-war-to-tackle-corruption. Our study does not include questions regarding a separate pre-registered hypothetical scenario about abusive behavior, which we discuss in the SI memo introducing the PAP.

then divided by the number of constituent terms.[12] One major concern was that officers' answers about citizens could be biased upwards through social desirability bias. To mitigate this concern, surveys were self-enumerated on an iPad in a private location, which has been shown to reduce sensitivity bias (Nanes and Haim 2020).

Estimation

We analyze the average treatment effects using the following linear regression model:

$$Y_{i,endline} = \alpha + \beta_1 TREAT_i + \beta_2 Y_{i,baseline} + \beta_3 Blocks_i + \beta_4 MissBaseline_i + \epsilon_i \quad (1)$$

We regressed each of our attitudinal outcome measures at endline ($Y_{i,endline}$) on an indicator for whether the officer was assigned to One Sorsogon ($TREAT_i$), controlled for the same attitudinal measure at baseline ($Y_{i,baseline}$), and included fixed effects for the blocking variables ($Blocks_i$). If there is missing data in the baseline outcome used as a control, we add an indicator variable for missingness ($MissBaseline_i$), and replace missing values with zeros (Gerber and Green 2012). Standard errors are clustered at the station-level using White clustered-robust standard errors. As a robustness check, we run the regression using only the endline cross-section, which does not rely on the imputation of baseline values:

$$Y_{i,endline} = \alpha + \beta_1 TREAT_i + \beta_2 Blocks_i + \epsilon_i \quad (2)$$

The structure of equation 1 is preregistered; equation 2 is not. To account for different treatment probability across blocks, which may introduce bias if treatment effects are not assumed to be the same across subjects (Gerber and Green 2012), we include robustness checks using weighted regression in the SI. The number of clusters in our study (17) is

[12] In the SI, we show that the results are robust to using a pre-registered alternative version of the indices that imputes missing responses to individual items.

relatively small, though simulations suggest that White cluster-robust standard errors tend to perform well in experimental studies with a similar number of clusters.[13]

Results

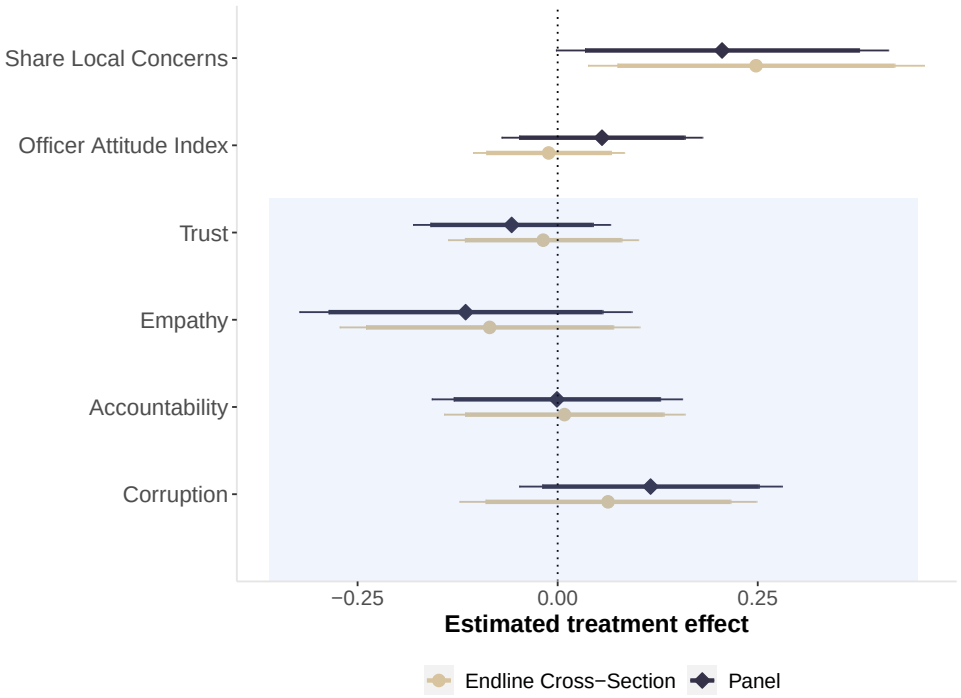
Figure 1 shows the results for our main hypotheses. Consistent with H1, participating in community policing caused officers to become significantly more likely to share the public safety concerns of civilians in their jurisdiction. The treatment effect is large in magnitude, representing a 24% increase in the number of public safety concerns officers matched with citizens (above an average of 1.07 matched issues in the control group). On the other hand, we find no evidence that participation in community policing impacted officers' attitudes towards civilians in terms of trust (H2), empathy (H3), accountability (H4), or corruption (H5). These null results are striking in light of the significant effects on *Shared Concerns*, a key mechanism the literature suggests should be driving attitudinal outcomes.

Before turning to our inductive theory, we address possible design-based explanations for the null results, including 1) power, 2) compliance, 3) ceiling effects, 4) attrition, and 5) spillovers. First, our model is sufficiently powered to rule out substantively large effects on officer attitudes. To assess this, we use the method in Rainey (2014) to calculate the bounds of the substantive effects on our combined *Officer Attitude Index* that fall within the 90% confidence interval. The largest substantive effect consistent with the cross-sectional model would be a negligible 2.5-point (or 4%) average increase, or a 2.7-point average decrease, on the 100-point outcome scale.

---

[13] https://declaredesign.org/blog/posts/how-misleading-are-clustered-ses-in-designs-wit
html. We acknowledge that this is dependent on a number of assumptions, including the intra-cluster correlation. In the SI, we show that the results are robust to using block-bootstrapped standard errors.

Figure 1: Effects of Community Policing on Main Outcomes



Estimates from the model controlling for baseline outcomes (equation 1) are in black; estimates from the cross-sectional model (equation 2) are in tan. Confidence intervals are displayed at the 95% level (thin bars) and 90% level (thick bars). Blue shading indicates results for the sub-indices that compose the Officer Attitude Index. Full regression results are in SI Table A.5.

It is unlikely that the results are driven by low compliance. Treatment officers attended 84% of the POP meetings and over 90% of the town halls to which they were assigned, and our field staff reported that most officers visited their assigned barangay at least once between each meeting. Ceiling effects are another unlikely explanation for the majority of the null results. The average baseline answer to the raw survey questions (on the 100-point scale) was 73 for all items in the overall *Attitude Index*, including 74 for the *Trust* questions, 72 for the *Empathy* questions, and 55 for the *Accountability* questions, leaving plenty of room

for movement in attitudes. The one index where we have substantial concerns about ceiling effects is *Corruption*, for which the average baseline value was 84.[14]

Fourth, regarding attrition: Of the 705 eligible officers we surveyed at baseline, we were unable to survey 190 at endline, due mostly to transfers, trainings, and retirements. Between baseline and endline, 183 new officers who would otherwise have been eligible to participate on a community policing team were hired or transferred to Sorsogon. In the SI, we show that there was no differential attrition between the treatment and control groups. The demographic characteristics and survey responses of "new" officers were highly balanced with the officers who dropped out of the study. Furthermore, the results of the model controlling for baseline outcomes and the endline cross-section are nearly identical, increasing our confidence that results are not driven by attrition.

A final concern is spillovers, which might bias towards null effects if treatment officers shared their experiences and opinions with control officers, or if control officers experienced an increased commitment to citizen engagement from their superiors. The first concern is mitigated by the fact that it would be surprising to find a large effect on the *Shared Concerns* outcome if treatment officers shared information with control officers. In addition, both types of spillovers would bias municipality-level heterogeneous effects towards null, which is contradicted by the patterns displayed in the following section. To further explore this issue, we conducted several analyses modeling spillovers, which we present in the SI. We approximate control officers' "exposure" to treatment using data on 1) family ties between officers (using the method pioneered by Cruz, Labonne, and Querubin 2017) and 2) officers' beat patrol partner assignments. At the individual level, we find no evidence that control officers' attitudes differentially changed based on their family ties or beat partnerships with treatment officers. We also explored heterogeneous effects based on the size and family

[14]See the SI for a more detailed breakdown of compliance and ceiling effects, including a Figure that displays the baseline distribution of all indices.

network density of police-stations, leveraging the assumption that officers are more likely to interact with each other in smaller, denser stations. Treatment effects did not differ significantly within the range of these moderators where there was sufficient data to support a substantive interpretation of the marginal effects. Together, these analyses suggest that spillovers were weak, at best.

Theory, Part 2 (Inductive) and Exploratory Analyses

Our main results suggest that community policing may be a good first step towards officers internalizing community concerns, but they also cast doubt on commonly-held assumptions about how community policing affects police officer attitudes. In the remainder of the paper, we endeavor to develop a theory that explains why, and under what conditions, community policing may change officers' attitudes towards citizens. To generate this theory, we rely on our own qualitative observations, as well as the observations of our field staff who attended each POP and town hall meeting. We asked our field staff to take detailed notes on each meeting and report back to us, with a particular focus on identifying mechanisms that might be driving the program's (in)effectiveness. After the intervention ended, we conducted informal interviews with police officers about their experiences during One Sorsogon. Finally, we conducted several exploratory analyses using our survey data to corroborate the insights garnered from these qualitative observations.[15]

The conventional wisdom motivating our pre-registered theory assumes that "officers' perceptions of danger are disproportionate to the objective risk because of the unpredictability and potential severity of the threats," and as a result, "greater knowledge of the people and places with which officers have contact could reduce... the perceived risk" (Paoline,

[15]The goal of this section is not hypotheses testing, and the heterogeneity analyses are not pre-registered. Additional details on our qualitative methodology are in the SI.

Myers, and Worden 2000, p. 581). We came to believe that this assumption lacks sufficient nuance when it comes to a) the nature of officers' prior beliefs and experiences, b) the content of information generated by community policing, and c) the narrow focus on information, as opposed to a wider range of mechanisms that moderate the effects of social contact.

In particular, we identified two sources of variation that challenge the universal applicability of this assumption. First, officers' pre-existing *embeddedness* in the community shaped the potential for community policing to effect changes in their beliefs. Officers from outside Sorsogon Province were more likely to hold negative stereotypes about local citizens that were attenuated by community policing, but additional contact with citizens was insufficient to change the attitudes of officers local to Sorsogon. Second, the local *security context* shaped whether officers' experiences during community policing heightened or lessened perceptions of threats to their personal safety. Officers' suspicions about citizens were reduced when they conducted community policing in municipalities where NPA presence was lower, but increased when NPA presence was higher.

Our qualitative observations led us to view the process by which community policing shapes officers' attitudes towards citizens through the lens of *intergroup contact theory* (Allport 1954; Pettigrew 1998). The "us versus them" outlook that many officers hold towards citizens is similar to the types of prejudices that individuals hold towards social outgroups. Scholars have previously adapted scope conditions of the "contact hypothesis" to better understand citizen views towards the police (Karim 2020; Rosenbaum et al. 2005), but not vice versa. In contrast to much of the conventional wisdom on community policing, the contact theory literature suggests that merely conveying knowledge about an outgroup has a weak effect on attitudes (Pettigrew and Tropp 2008; Hopkins, Sides, and Citrin 2019). The effects of social contact are mediated by 1) the *content* of information about the outgroup, 2) whether the nature of interactions reduces *anxiety* about outgroup interactions, 3) whether contact induces an increase in *perspective-taking*, and 4) whether contact creates new *shared*

identities with outgroup members (Allport 1954; Pettigrew and Tropp 2008; Tajfel 1978; Gaertner et al. 1999; Hewstone, Rubin, and Willis 2002).

These mechanisms help make sense of our observations about how officer embeddedness and the local safety context shaped the effect of community policing on officers. The effects of social contact should be muted among officers with existing local ties because they likely already have better information, less anxiety, existing shared perspectives, and shared identities with local citizens. Moreover, when citizens raise issues related to the NPA, it can heighten officers' perception of threats to their safety, thereby generating "negative contact" that increases anxiety, worsens stereotypes of citizens, and hardens officer identities. In the remainder of this section, we lay out the evidence we gathered on these conditions and refine our expectations about how increased contact with citizens can affect the attitudes of frontline state agents.

Embeddedness

A first main takeaway from our field staff's observations was that officers from within Sorsogon responded differently to community policing than officers from outside the province. Early on, officers local to Sorsogon appeared more comfortable during the POP meetings and were quicker to understand citizens' perspectives when issues were raised. Yet some of these local officers expressed to us that they viewed community policing in similar terms to what Crank (2014, p.311) characterizes as "departmental bullshit," stating that they already had enough prior experience in the community to do their job well. On the other hand, our field staff indicated that as the program progressed, non-local officers seemed increasingly at ease, displayed greater effort, and improved their communication style during the POP

meetings, despite starting at a lower baseline on these metrics.[16] Our conversations with non-local officers indicated that community policing helped dispel some of their suspicions of citizens that stemmed from Sorsogon's reputation for being a hotbed of citizen resistance (manifesting in insurgency and drug trafficking).

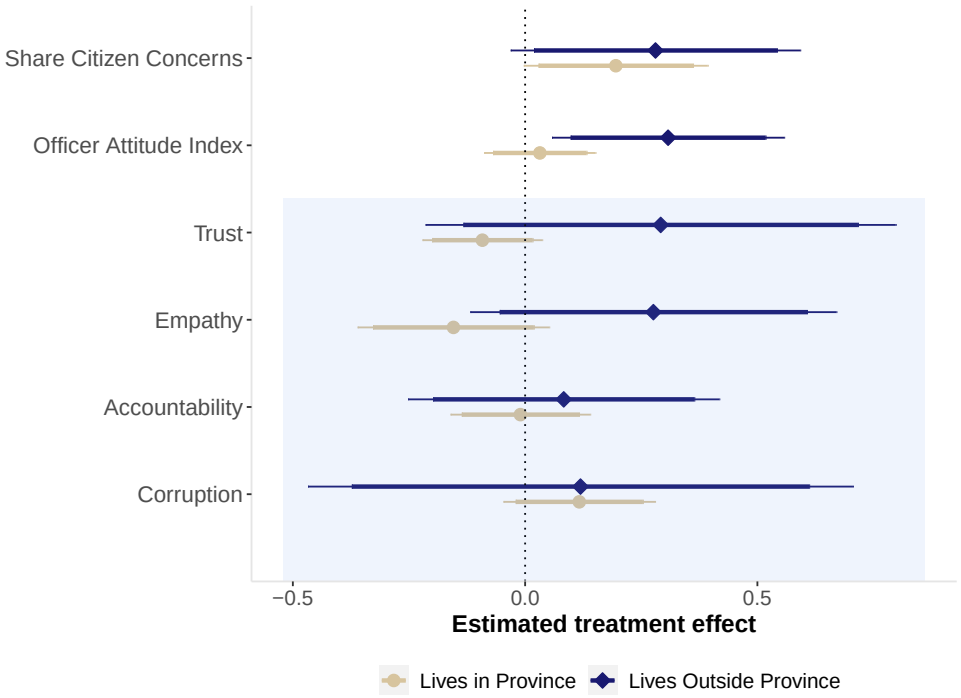
To further explore these notions, we looked to the survey data, which confirmed that officers from outside Sorsogon had worse attitudes towards local citizens at the start of the intervention. Among non-local officers, there was 8% less overlap in *Shared Concerns* with issues deemed important by citizens. Average responses to the survey questions on *Trust* and *Empathy* were also eight points lower among officers from outside Sorsogon at baseline, though average answers of 75/100 among officers from Sorsogon left room for upwards movement for those officers as well. We then explored whether the patterns noted by our field staff were reflected in heterogeneous effects. Results for models interacting the treatment variable with an indicator of whether officers lived in Sorsogon at the start of the intervention are shown in Figure 2.

Corroborating our field team's observations, the intervention had large positive effects on the overall *Officer Attitude Index* among officers from outside Sorsogon, but no effect among officers who lived inside the province. By the end of the program, treatment officers from outside Sorsogon had virtually caught up to officers from inside the province in terms of their attitudes towards citizens. Consistent with our inductive theory, outlined below, the heterogeneous effects are driven primarily by officers' *Trust* and *Empathy*, the two outcomes most clearly linked to social contact with citizens. Our theory does not speak to whether top-down *Accountability* or views on the appropriateness of *Corruption* should be shaped by officer embeddedness, and we do not find significant heterogeneous effects for these outcomes.

---

[16]We observed this change in the period between the training and the culminating town halls, suggesting that officer outcomes were shaped by experiences during the POP meetings and intervening barangay visits.

Figure 2: Heterogeneous Effects of Community Policing, by Officer Embeddedness



All estimates are from the specification controlling for baseline outcomes (equation 1). Marginal effects displayed in blue are for officers from outside Sorsogon; estimates displayed in tan are for officers local to Sorsogon. Confidence intervals are displayed at the 95% level (thin bars) and 90% level (thick bars). Full regression results are in SI Table A.6.

We probe the robustness of these patterns in the SI. Results hold when controlling for officer characteristics that may correlate with embeddedness, including rank, education, religion, age, gender, time in service, commute time, and public service motivation. They also hold when interacting control variables with the treatment indicator. The inclusion of station fixed effects mitigates concerns about the PNP disproportionately assigning non-embedded officers to municipalities where the program was more likely to succeed (for example, stations with higher leadership buy-in), because the model compares effects among embedded and non-embedded officers at the same station. Still, station assignments of officers from outside and inside Sorsogon are balanced on crime rate, NPA presence, and a range of socio-

demographics, and the results hold when replacing the station effects with municipality-level controls. Finally, results are similar when using an alternative measure of embeddedness – whether officers live in the *municipality* where they are assigned.

Drawing from these observations, we posit that programs intended to enhance interactions with citizens will be more likely to improve attitudes among officers who are less embedded ex-ante, assuming that they hold more negative suspicions of citizens than local officers.[17] Our theory relies on two propositions that challenge common assumptions in the community policing literature. First, existing work largely assumes that, under a system of reactive policing, informal and non-threatening interactions with citizens are rare for all officers. We contend that embedded officers are already more likely to have these types of interactions with local citizens during their daily lives, outside of work. Research suggests that prior contact can increase the likelihood that people will seek out future contact (Pettigrew 1998), meaning that embedded officers may also already be more likely to engage with citizens in the course of their regular policing duties. Simply put, community policing is less likely to generate new experiences with citizens for embedded officers, preventing the social contact mechanisms from taking hold.

Second, the canonical community policing literature places outsized emphasis on *information* about citizens as being the key mechanism that can unlock improvements in officer trust and empathy. The results displayed in Figure 2 suggest that *Shared Citizen Concerns* increased for both embedded and non-embedded officers, meaning it is unlikely that non-embedded officers saw improved attitudes solely because they benefited more in terms of information gained through contact. This is not to say that the informational mechanism is altogether unimportant. Rather, our claim is that information may need to be activated to-

[17]This effect depends on officers' prior perceptions. We would expect the opposite effect when unembedded officers hold disproportionately positive stereotypes.

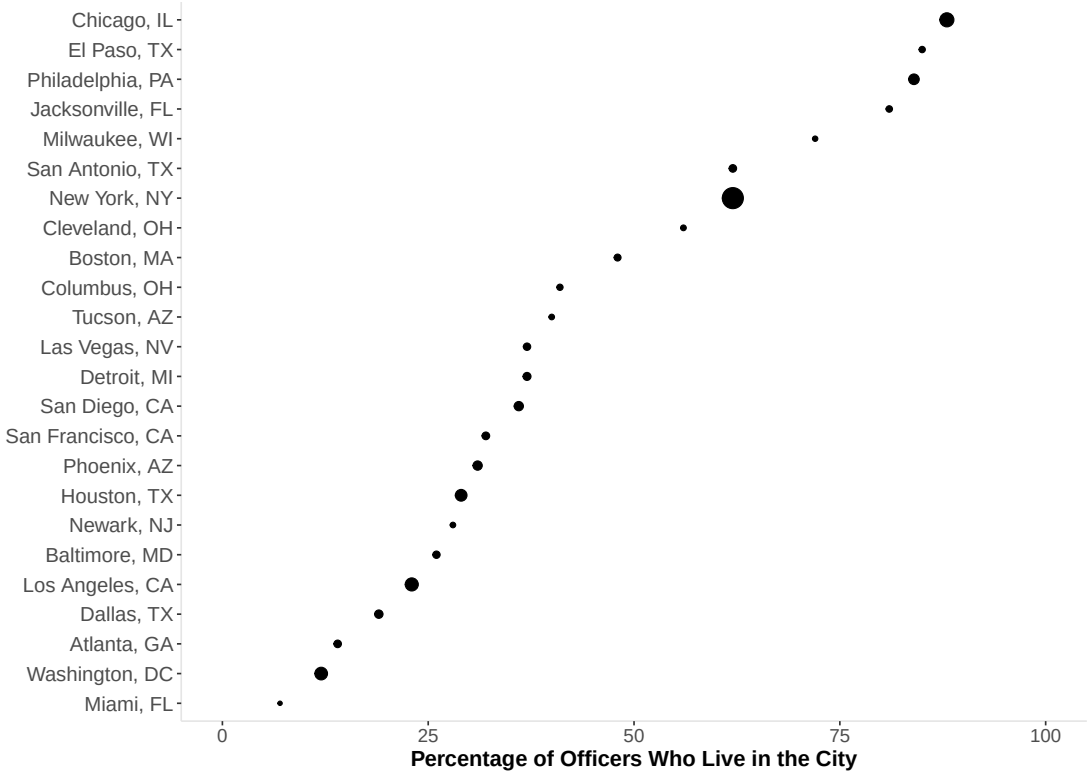
gether with other mechanisms in order for contact with citizens to result in improved officer attitudes, and that these mechanisms may be blunted for embedded officers.

Consistent with our field staff's observations about local officers' comfort and communication style during the POP meetings, we theorize that embedded officers are already less likely to be *anxious* about interacting with citizens in the absence of community policing, as well as better able to take the *perspective* of local citizens. This translates to reduced potential to see movement in their attitudes. Anxiety-reduction is a crucial moderator shaping outgroup attitudes (Pettigrew 1998; Pettigrew and Tropp 2008), and information has a bigger effect on outgroup attitudes when paired with improved perspective-taking (Pettigrew and Tropp 2008; Kalla and Broockman 2023; Adida et al. 2024). It may also be the case that embedded officers already share aspects of their identity with local citizens, in this case, as a "Sorsogueño." When individuals share identity features, their attitudes are less malleable because they tend to see new information as "situational," rather than contributing to their views of an outgroup (Paolini et al. 2014; Lyall 2010; Lyall, Blair, and Imai 2013). In contrast, unembedded officers are more likely to develop a new common group identity shared with locals, a key mechanism through which social contact can trigger changes to outgroup attitudes (Gaertner et al. 1999; Hewstone, Rubin, and Willis 2002).

Our theory is important in light of the tremendous variation in police embeddedness around the world. This variation is driven by a debate about the tradeoff between officer effectiveness and partiality, highlighted by the literature on bureaucratic embeddedness (Pepinsky, Pierskalla, and Sacks 2017). Embedded bureaucrats often exhibit greater local knowledge, higher effort and provide better services (Tsai 2007; Ricks 2016; Bhavnani and Lee 2018), but they are also prone to favoritism and further alienating marginalized populations (Fjeldstad 2005; Haim, Nanes, and Davidson 2021). Reflecting this tradeoff, police departments in the US differ in whether they require officers to live within the city boundaries where they serve (Kennedy et al. 2017), and officer embeddedness varies from 7% of officers living locally in Miami to 88% of officers living locally in Chicago (see Figure

3). This discussion is also salient in countries with weaker police capacity and a high risk of corruption. In Mexico, for example, policymakers debate whether to assign local officers who are better at identifying key actors in the drug trade, or non-local "Federales" who are less likely to collude with traffickers.

Figure 3: Distribution of Police Embeddedness in US Cities



Data from the U.S. Equal Employment Opportunity Commission (EEOC) and the Census Bureau. Displayed are the 25 largest US police forces - points are sized by the number of officers employed by the force. For an expanded description of these data, see https://fivethirtyeight.com/features/most-police-dont-live-in-the-cities-they-serve/

In light of this debate, we offer a few conjectures on where community policing might be most beneficial. Our results imply that community policing provides experiences similar to those embedded officers already hold, but this does not mean that community policing should be eschewed in favor of recruiting officers who are already embedded in their community. Optimistically, community engagement could help close the gap in trust and empathy between embedded and non-embedded officers, perhaps allowing communities to benefit from

the upsides of police embeddedness without being subject to as many of the downsides. One idea would be to focus community policing efforts in places where police embeddedness is low, or to disproportionately assign unembedded officers to these programs, while focusing on other reforms to improve attitudes among officers with extensive local experience. That said, officer attitudes are not the only relevant policing outcomes; embedded officers may still be more effective at using the platform of community policing to improve citizen-facing outcomes. If this is the case, police departments may face a tradeoff, whereby they can either prioritize improving the attitudes of officers or citizens, but not both.

Safety Context and Threats to Personal Safety

A second major takeaway from our qualitative observations concerns how the local safety context shaped the content of information to which officers were exposed while participating in community policing. Learning more about the issues that citizens dealt with on a day-to-day basis shaped officers' understanding of citizen-facing public safety issues, but it also affected assessments of their own personal safety. Specifically, community policing heightened the salience of threats to officer safety in areas that were proximate to the NPA insurgency.

Although none of the POP teams officially designated the NPA threat as their focus issue, we found that barangay leaders in NPA-affected municipalities used the POP meetings to raise instances where NPA personnel were sighted or suspected of recruiting at-risk youth.[18] Our field staff reported that when these issues were raised, officers tended to act uncomfortable and express outward frustration that citizens were not more forthcoming with real-time information about NPA activities. This reaction was likely tied to officers' fears of

[18]We think the heterogeneous effects of NPA presence were driven almost entirely by the POP aspect of the program. Citizens virtually never raised the NPA issue during town halls, and it was not part of the initial training.

NPA ambushes, which was the most common threat to officer safety mentioned in our interviews. Many officers told stories about someone they knew on the force who had been killed, injured, or kidnapped by the NPA. When ambushes occur, officers often blame the citizen population for alerting the NPA about police movements or, at the very least, failing to provide the police with information that would allow them to avoid unexpected encounters.

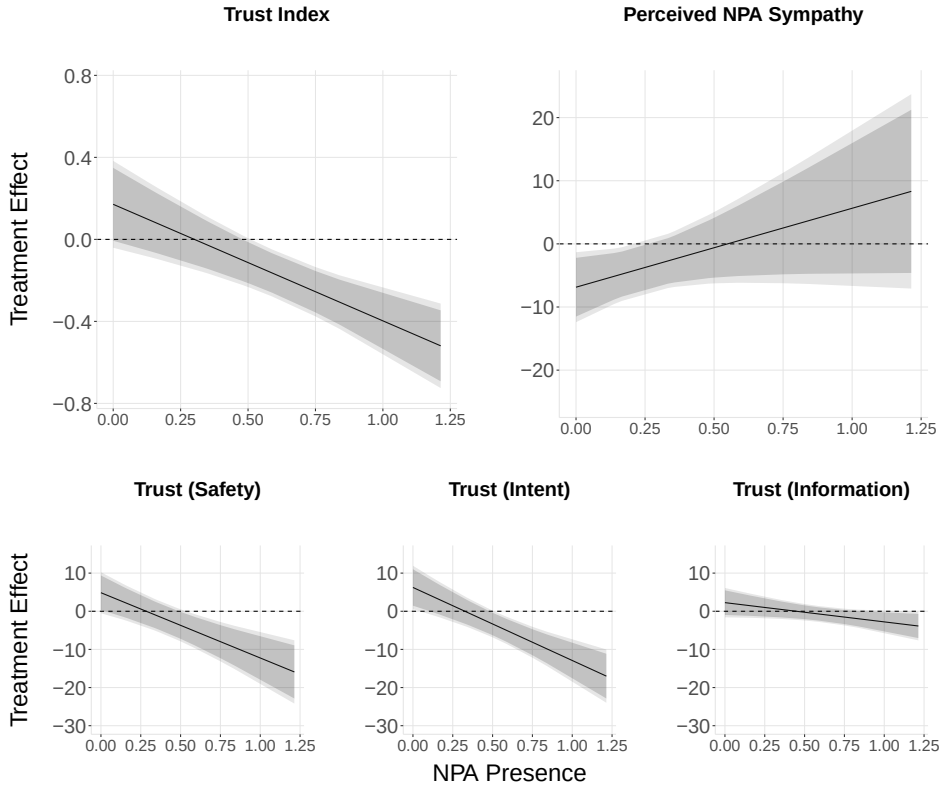
Notably, officers who conducted community policing in NPA-affected areas tended to express frustration with the broader citizenry, even though the individuals who attended community policing events were largely law-abiding, cooperative and respectful (see Hanson, Kronick, and Slough 2024). We think that information about insurgent activity led officers to infer that a larger proportion of citizens must be sympathetic to the NPA, given that the group survives by hiding amongst the population. Officers focused on citizen behavior being shaped by NPA propaganda while minimizing citizens' own fear of retaliation, a well-known factor that impedes citizen information-sharing about insurgents.

These observations led us to investigate whether the effect of community policing was conditioned by the baseline level of NPA presence in the municipality where officers were stationed. To measure NPA presence, we rely on internal military intelligence reports that code barangays on a 3-point scale of "cleared" (0), "influenced" (1), or "controlled" (2) by the NPA.[19] The latter two categories indicate that insurgents regularly take residence or have an active party organization in the barangay. Our measure proxies for NPA activity by calculating the mean value of the 3-point scale for all barangays in a given municipality. While our sample was restricted to "cleared" barangays, NPA personnel still interact with civilians and try to expand their presence into these barangays from their nearby areas of

[19]We use reports from December 2015, the most recent year when they were made available to us. Similar data are used in Crost, Felter, and Johnston (2014) and Haim (2022).

control.[20] In Figure 4, we begin by focusing on the heterogeneous effect of NPA presence on officer *Trust*, the citizen-facing outcome most directly tied to officer safety.

Figure 4: Heterogeneous Effect of Community Policing on Officer Trust, by NPA Presence



All plots display marginal effects of treatment by NPA (rebel) presence in the municipality, using the specification controlling for baseline outcomes (equation 1). The top panel displays results for the main trust index and a separate outcome asking officers to rate civilians' sympathy towards the NPA in their assigned municipality. The bottom panel displays results for the individual survey items that compose the trust index. Confidence intervals are displayed at the 95% level (light gray) and 90% level (dark gray). Full regression results are in SI Table A.7.

Participation in community policing in NPA-adjacent areas substantially decreased officers' trust in civilians (see the *Trust* index in the top-left panel). On the other hand,

[20]We speculate that the detrimental effects of police contact with citizens may be exaggerated in less secure areas.

in municipalities without an active NPA presence, community policing marginally increased trust in civilians. This pattern is especially pronounced for officers' perception of their own safety (*Trust - Safety*) and of citizens' concern for officer well-being (*Trust - Intent*), displayed in the bottom panel of Figure 4. In the SI, we show that the interpretation of the marginal effects is consistent when using the Hainmueller, Mummolo, and Xu (2019) binning estimator, and when including a battery of municipal-level controls.

The heterogeneous effects for the remaining outcome indices are shown in the SI. There was no significant heterogeneous effect on *Shared Citizen Concerns*, for which the average treatment effect is positive across the full range of the *NPA Presence* moderator. This is consistent with the idea that officers learned about the community's concerns regardless of where they were assigned, but the *content* of those concerns differed. To examine this mechanism, we looked to a separate survey item that asked officers to estimate the percentage of citizens in their municipality who are sympathetic to the NPA (top-right panel of Figure 4). Consistent with our qualitative observations, treatment officers in low-NPA areas perceived that citizens were less sympathetic to the NPA (relative to the control group), while treatment officers in high-NPA areas perceived that citizens were more sympathetic to the NPA.[21] We do not find significant heterogeneous effects on officer *Empathy*, the other citizen-facing outcome. This result is not surprising in light of the fact that we operationalized this concept to focus on the seriousness of crimes and complaints that citizens *did* report

[21]Evidence for this pattern using the equation 1 specification is marginal ($p = 0.12$ on the interaction term), but the results are significant at conventional levels when using weighted regression or cluster-bootstrapped standard errors.

to the police. As described above, the most relevant aspect of empathy in the context of the NPA is whether officers understand citizens' choice *not* to report rebel activity.[22]

Based on these patterns, we challenge the idea that the information officers attain through community policing will uniformly improve officers' expectation of personal safety. Although this assumption may hold in "safe" areas, where citizens are likely to raise concerns that pose little threat to officers, we predict that additional information about citizen concerns in "dangerous" areas can increase officers' perceived risks to personal safety. This can generate anxiety and worsen trust in citizens. Our hypothesis is again rooted in intergroup contact theory: "positive" contact can improve outgroup attitudes, while "negative" contact has the reverse effect (Paolini, Harwood, and Rubin 2010; Schäfer et al. 2021; Lowe 2021). Even when officers experience both positive and negative contact during community policing, their overall trust may decrease because negative contact tends to be a more powerful predictor of outgroup attitudes (Barlow et al. 2012). Importantly, this process does not require that direct interactions with citizens during community policing are antagonistic. Rather, negative contact can occur from exposure to information that highlights threats posed by outgroup members, regardless of whether the threat comes from the individuals with whom one interacts (Schäfer et al. 2021). Negative information can raise intergroup anxiety and harden identity divides between officers and citizens, two of the strongest mediators linking negative contact to attitudinal change (Pettigrew and Tropp 2011; Paolini, Harwood, and Rubin 2010).

Our theory is particularly powerful when combined with insights on how "neighborhood context" affects police officers (Terrill and Reisig 2003; Paoline, Myers, and Worden

---

[22]Our theory does not speak to whether NPA presence should affect *Accountability* or *Corruption*, though we did find a positive heterogeneous effect of NPA presence on *Accountability* ($p = .099$). From a purely speculative standpoint, police leaders may be more sensitive to officer misconduct when trying to win "hearts and minds" away from insurgents.

2000). The key insight is that rather than *overcoming* the effects of neighborhood context (as is commonly assumed), community policing can *heighten* the effects of neighborhood context on officers' attitudes and behaviors. This mechanism corresponds to Nix, Pickett, and Wolfe's (2020) claim that working in violent communities may condition officers to expect more noncompliance, disrespect, and violence from citizens. We contend that the information officers acquire during community policing can worsen harmful stereotypes of the citizenry, even if officers primarily interact with cooperative, law-abiding citizens. We focused on the threat posed by insurgency due to its relevance in Sorsogon. The literature on neighborhood context highlights additional threats like organized crime, high gun ownership rates, or a history of citizen antagonism towards the police, which are less relevant in Sorsogon but may yield similar effects in settings where they are present.

One troubling implication of our theory is that community policing may be most likely to fail in places where improvements to officer-citizen relations are most needed. Officers assigned to areas with high safety threats may become more guarded and more likely to think force is necessary, which suggests the need to explore alternative reforms in high risk areas. We do not suggest that the police should become less engaged or more militaristic, which could have negative effects on citizen-facing outcomes. Instead, departments may consider alternative modalities of community policing that retain a citizen-centered approach without immediately ramping up contact on a large scale. The police could also incorporate perspective-taking interventions into officer trainings, which may serve to set appropriate expectations and help officers see the world from citizens' perspectives.

Discussion

This article highlights the importance of studying street-level officers' attitudes towards the communities they serve. We find that community policing can help officers better understand citizens' public safety needs, but it does not bring about overall improvements to officers'

trust, empathy, or views on misconduct. This pattern suggests that much of the literature lacks important nuance when it presumes that community policing uniformly improves officer attitudes. Our inductive theory challenges the foundational assumption that "regular, informal contact with citizens" can consistently minimize officers' tendency "to think that citizens' intentions and safety risks are worse than they are in reality" (Paoline, Myers, and Worden 2000, 7). Although this may be true for officers who are not already embedded in the community, officers who have extensive local ties likely hold entrenched attitudes that are difficult to shift through additional contact with citizens. Moreover, contact with citizens (even in curated, cooperative settings) can raise the salience of threats to officers' personal safety, negatively affecting their views of the average citizen's trustworthiness. In this section, we describe several limitations of our study, then discuss implications for bureaucrats in other contexts.

First, we are limited in our ability to capture whether the intervention translated to changes in officers' *behavior* outside the community policing meetings. Our field staff's observations suggest that embeddedness and threats to officer safety shaped officers' effort and communication style during the POP meetings, but the PNP does not collect officer-level data that would allow us to reliably measure effort, discrimination, or abuse in their day-to-day tasks. For example, one outstanding question that speaks to the policy implications of our results is whether officers became more effective at their jobs as a result of internalizing local communities' needs, even in the absence of broader changes to officer attitudes. Our results are best understood as an important first step towards uncovering the behavioral implications of programs like community policing, given that frontline governments agents' attitudes are often linked to their behaviors (Ingram, Terril, and Pauline 2018; Keiser 2010).

We also acknowledge that there are other moderators and contextual factors that could influence community policing's impact on officers, beyond those we chose to highlight in this piece. For instance, our inductive theory centers on two factors that determine whether the "bottom-up" mechanisms of community policing (driven by citizen contact) take hold.

Evidence for how police leadership can instill a culture of respect and accountability towards citizens from the "top-down" is limited. The largely null top-down results could be explained in part by the powerful norm against "ratting out" fellow officers for misconduct. We caution that in order to change citizen-facing norms, high level leadership buy-in must be paired with broader institutional reforms that improve monitoring and ensure effective implementation by mid-level leadership.[23] In addition, the effects of community policing on officers may depend on the specific bundle of activities that are included in the program, though we reiterate that our intervention included a bundle of activities that is commonly implemented in the real-world.[24] For example, scholars and policymakers might consider whether some of the backlash effects we observed could be mitigated by weighting the intervention more heavily towards contact with "ordinary" citizens or community leaders. Finally, we note that our study was not designed to definitively decipher between the four mechanisms of social contact theory, and that our application of these concepts is an inductive exercise. Future work could test how officers' information acquisition, anxiety reduction, perspective-getting, and shared identity are shaped by citizen contact, and how these mechanisms ultimately shape bureaucrats' attitudes and behaviors.

With those limitations in mind, we now turn to the question of how this article speaks to research on street-level state agents more broadly. Social contact with citizens is a crucial aspect of the job for soldiers and many "conventional" bureaucrats like social workers, healthcare workers, and cadastral officers. Like police officers, these agents hold

---

[23]Anecdotally, the intervention seemed to have a greater effect where station-level leaders were more engaged and did not just treat the program as additional "red tape."

[24]The fact that Blair et al. (2021) also found null results of "locally appropriate" bundles of community policing practices on officer attitudes in other settings suggests that our results are not an anomaly on this front.

high discretion over implementing policy (Lipsky 1980; Brierley et al. 2023), and their actions are shaped by their attitudes towards clients and the bureaucratic mission (Keiser 2010). Yet existing studies tend to rely on static measures of embeddedness and identity to explain bureaucrats' local knowledge, motivations, and prejudices. We think it is instructive to treat bureaucrats' local knowledge and their attitudes towards citizens as being endogenous to the governance strategies they implement while on the job.

Future research could explore how bureaucrats' behavior is shaped by the interaction of their fixed traits and the ties they build with citizens while on the job. For example, variation in whether bureaucrats are racially, religiously, or socioeconomically representative of the communities they serve may shape the effects of community engagement. Perhaps bureaucrats who do not have prior experiences with minority populations will derive greater benefits from social contact with citizens from these groups. On the other hand, it may be the case that certain dimensions of representation cannot be artificially replicated, such as race in communities with deep-seated racial divisions. Since racial and religious divisions were largely absent from our study's context, we cannot speak directly to this dynamic, but we view this as a promising path for future research.

When it comes to applying our theory of local safety context, we note that many state agents face substantial threats to personal safety that can be revealed through citizen contact, though the particular nature of those threats may vary. Our theory might be most applicable to soldiers because, like police officers, they are part of the coercive arm of the state and are more likely to be targeted with violence by criminals or insurgents. "Hearts and minds" counterinsurgency often brings soldiers into contact with community leaders in insecure areas, raising the salience of threats that could increase soldiers' anxieties, stereotypes, and propensity to employ violence. This insight likely also applies to conventional bureaucrats assigned to areas with active non-state armed groups or criminal gangs, given that these groups often target violence at government workers who they view as encroaching into their

territory.[25] Finally, our theory may apply beyond armed conflict and criminal violence. Although the nature and severity of the threat is different, social workers and teachers often report high risks of physical assault, sexual harassment, and verbal abuse, which could have a similar effect on trust towards the populations they serve.

Because the behavior of individual police officers, soldiers, and bureaucrats is crucial to jumpstarting virtuous cycles of citizen-state relations, we hope this study motivates new research on how government policies and political context affect the attitudes and behaviors of the people charged with implementing those policies. Creating the conditions for street-level state agents to develop a sense of trust, empathy, and accountability towards the communities they serve is a key step in building government legitimacy more broadly.

Data Availability Statement

Research documentation and/or data that support the findings of this study are openly available at the American Political Science Review Dataverse: https://doi.org/10.7910/DVN/2LHT5U.

Funding Statement

This research was funded by the Evidence in Governance and Politics (EGAP) Metaketa IV project, and by the UC San Diego Policy Design and Evaluation Lab (PDEL).

Ethics and Conflicts of Interest

The authors declare the human subjects research in this article was reviewed and approved by UC San Diego IRB under protocols 170974 and 170415, with reliance agreements approved

[25]Indeed, an evaluation of the Philippines' flagship development program in conflict zones found that safety concerns caused development workers to shirk their duties, undercutting the program's implementation (Haim, Fernandez, and Cruz 2019).

by the IRBs at Saint Louis University and Florida State University. The authors affirm that this article adheres to the principles concerning research with human participants laid out in APSA's Principles and Guidance on Human Subject Research (2020). The authors declare no ethical issues or conflicts of interest in this research.

References

Adida, Claire, Adeline Lo, Melina Platas, Lauren Prather, and Scott Williamson. 2024. "When Hearts Meet Minds: Complementary Effects of Perspective-Getting and Information on Refugee Inclusion." *Working Paper*. 30

Allport, Gordon W. 1954. *The Nature of Prejudice.* Addison-Wesley. 3, 25, 26

Ba, Bocar A, Dean Knox, Jonathan Mummolo, and Roman Rivera. 2021. "The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago." *Science* 371 (6530): 696–702. 2

Barlow, Fiona Kate, Stefania Paolini, Anne Pedersen, Matthew J Hornsey, Helena RM Radke, Jake Harwood, Mark Rubin, and Chris G Sibley. 2012. "The Contact Caveat: Negative Contact Predicts Increased Prejudice More than Positive Contact Predicts Reduced Prejudice." *Personality and Social Psychology Bulletin* 38 (12): 1629–1643. 36

Barrett-Lennard, Godfrey T. 1981. "The Empathy Cycle: Refinement of a Nuclear Concept." *Journal of counseling psychology* 28 (2): 91. 7

Bhavnani, Rikhil R and Alexander Lee. 2018. "Local Embeddedness and Bureaucratic Performance: Evidence from India." *The Journal of Politics* 80 (1): 71–87. 2, 30

Birzer, Michael L. 2008. "What Makes a Good Police Officer? Phenomenological Reflections from the African-American Community." *Police Practice and Research: An International Journal* 9 (3): 199–212. 7

Blair, Graeme, Fotini Christia, Jeremy Weinstein, Eric Arias, Emile Badran, Robert A. Blair, Ali Cheema, Thiemo Fetzer, Guy Grossman, Dotan Haim, Zulfiqar Hameed, Rebecca Hanson, Ali Hasanain, Ben Kachero, Dorothy Kronick, Benjamin Morse, Robert Muggah, Fatiq Nadeem, Matthew Nanes, Tara Slough, Nico Ravanilla, Jacob N. Shapiro, Barbara Silva, Pedro C. L. Souza, Lily Tsai, and Anna Wilke. 2021. "Community Policing Does Not Build Citizen Trust in Police or Reduce Crime in the Global South." *Science* 374 (6571): eabd3446. 1, 3, 4, 10, 39

Blair, Robert A., Sabrina M. Karim, Michael J. Gilligan, and Kyle Beardsley. 2022. "Policing Ethnicity: Lab-in-the-Field Evidence on Discrimination, Cooperation, and Ethnic Balancing in the Liberian National Police." *Quarterly Journal of Political Science* 17 (2): 141–181. 2

Bracey, Dorothy H. 1992. "Police Corruption and Community Relations: Community Policing." *Police Stud.: Int'l Rev. Police Dev.* 15 : 179. 9

Brierley, Sarah, Kenneth Lowande, Rachel Augustine Potter, and Guillermo Toral. 2023. "Bureaucratic Politics: Blind Spots and Opportunities in Political Science." *Annual Review of Political Science* (Forthcoming). 1, 40

Cordner, Gary W. 1995. "Community Policing: Elements and Effects." *Police Forum* 5 (3). 1, 7, 9

Crank, John P. 2014. *Understanding Police Culture*. Routledge. 7, 26

Crost, Benjamin, Joseph Felter, and Patrick Johnston. 2014. "Aid Under Fire: Development Projects and Civil Conflict." *American Economic Review* 104 (6): 1833–1856. 33

Cruz, Cesi, Julien Labonne, and Pablo Querubin. 2017. "Politician Family Networks and Electoral Outcomes: Evidence from the Philippines." *The American Economic Review* 107 (10): 3006 – 3037. 23, SI 25, SI 26

Fjeldstad, Odd-Helge. 2005. "Corruption in Tax Administration: Lessons from Institutional Reforms in Uganda." *CMI Working Paper* 2005:10 . 30

Gaertner, Samuel L, John F Dovidio, Mary C Rust, Jason A Nier, Brenda S Banker, Christine M Ward, Gary R Mottola, and Missy Houlette. 1999. "Reducing Intergroup Bias: Elements of Intergroup Cooperation." *Journal of personality and social psychology* 76 (3): 388. 26, 30

Gerber, Alan S and Donald P Green. 2012. *Field Experiments: Design, Analysis, and Interpretation.* WW Norton. WW Norton. 20

Gill, Charlotte, David Weisburd, Cody W Telep, Zoe Vitter, and Trevor Bennett. 2014. "Community-Oriented Policing to Reduce Crime, Disorder and Fear and Increase Satisfaction and Legitimacy Among Citizens: A Systematic Review." *Journal of Experimental Criminology* 10 (4): 399 – 428. 1

Goldstein, Herman. 1990. "Excellence in Problem-Oriented Policing." In *New York NY: Police Executive Research Forum*. 14

Gulzar, Saad and Benjamin J Pasquale. 2017. "Politicians, Bureaucrats, and Development: Evidence from India." *American Political Science Review* 111 (1): 162–183. 2

Haarr, Robin N. 2001. "The Making of a Community Policing Officer: The Impact of Basic Training and Occupational Socialization on Police Recruits." *Police Quarterly* 4 (4): 402–433. 7

Haim, Dotan. 2022. "Networks and Insurgency: How Civilian Social Relationships Shape Conflict." *Working Paper*. 33, SI 25

Haim, Dotan, Maria Carmen Fernandez, and Micah Cruz. 2019. "Evaluation of the Payapa at Masaganang Pamayanan (Pamana) Program." *UNDP Development Report*. 41

Haim, Dotan, Matthew Nanes, and Michael W Davidson. 2021. "Family Matters: The Double-Edged Sword of Police Community Connections." *Journal of Politics* 83 (4): 1529–1544. 2, 30

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27 (2): 163–192. 35, SI 19, SI 27

Hanson, Rebecca, Dorothy Kronick, and Tara Slough. 2024. "Preaching to the Choir: A Problem of Participatory Interventions." *Journal of Politics* Forthcoming . 33

Hayeslip, David W and Gary W Cordner. 1987. "The Effects of Community-Oriented Patrol on Police Officer Attitudes." *American Journal of Police* 95 : 95–120. 10

Hewstone, Miles, Mark Rubin, and Hazel Willis. 2002. "Intergroup Bias." *Annual review of psychology* 53 (1): 575–604. 26, 30

Hopkins, Daniel J, John Sides, and Jack Citrin. 2019. "The Muted Consequences of Correct Information About Immigration." *The Journal of Politics* 81 (1): 315–320. 25

Ingram, Jason R, William Terril, and Eugene A Pauline. 2018. "Police Culture and Officer Behavior: Application of a Multi-Level Framework." *Criminology* 56 (4): 780–811. 1, 7, 38

Inzunza, Miguel. 2015. "Empathy from a Police Work Perspective." *Journal of Scandinavian Studies in Criminology and Crime Prevention* 16 (1): 60–75. 7

Kalla, Joshua L and David E Broockman. 2023. "Which Narrative Strategies Durably Reduce Prejudice? Evidence from Field and Survey Experiments Supporting the Efficacy of Perspective-Getting." *American Journal of Political Science* 67 (1): 185–204. 30

Karim, Sabrina. 2020. "Relational State Building in Areas of Limited Statehood: Experimental Evidence on the Attitudes of the Police." *American Political Science Review* 114 (2): 536–551. 25

Kaufman, Herbert. 1956. "Emerging Conflicts in the Doctrines of Public Administration." *American Political Science Review* 50 (4): 1057–1073. 2

Keiser, Lael R. 2010. "Understanding Street-Level Bureaucrats' Decision Making: Determining Eligibility in the Social Security Disability Program." *Public administration review* 70 (2): 247–257. 5, 38, 40

Kennedy, Brandy A, Adam M Butz, Nazita Lajevardi, and Matthew J Nanes. 2017. *Race and Representative Bureaucracy in American Policing*. Springer. 30

Kuipers, Nicholas. 2022. "Failing the Test: The Countervailing Attitudinal Effects of Civil Service Examinations." *American Political Science Review*: 1–18. 2, 6

Lipsky, M. 1980. *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*. New York: Russell Sage Foundation. 1, 6, 40

Lowe, Matt. 2021. "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration." *American Economic Review* 111 (6): 1807–1844. 36

Lyall, Jason. 2010. "Are Coethnics More Effective Counterinsurgents? Evidence from the Second Chechen War." *American Political Science Review* 104 (1): 1–20. 30

Lyall, Jason, Graeme Blair, and Kosuke Imai. 2013. "Explaining Support for Combatants during Wartime: A Survey Experiment in Afghanistan." *American Political Science Review* 107 (4): 679 – 705. 5, 30

Maddox, Lucy, Deborah Lee, and Chris Barker. 2011. "Police Empathy and Victim PTSD as Potential Factors in Rape Case Attrition." *Journal of police and criminal psychology* 26 : 112–117. 7

March, James G. 1994. *A Primer on Decision Making: How Decisions Happen.* Simon and Schuster. 8

Mastrofski, Stephen D., Robert E Worden, and Jeffrey B Snipes. 1995. "Law Enforcement in a Time of Community Policing." *Criminology* 33 (4): 539–563. 1

Maynard-Moody, Steven Williams and Michael Craig Musheno. 2003. *Cops, Teachers, Counselors: Stories from the Front Lines of Public Service.* University of Michigan Press. 5

Meier, Kenneth J and Jill Nicholson-Crotty. 2006. "2
Gender, Representative Bureaucracy, and Law Enforcement: The Case of Sexual Assault." *Public Administration Review* 66 (6): 850–860.

Mo, Cecilia Hyunjung and Katharine M Conn. 2018. "When do the Advantaged See the Disadvantages of Others? A Quasi-Experimental Study of National Service." *American Political Science Review* 112 (4): 721–741. 2

Moon, Byongook and Laren J Zager. 2007. "Police Officers' Attitudes Toward Citizen Support: Focus on Individual, Organizational and Neighborhood Characteristic Factors." *Policing: An International Journal of Police Strategies & Management* 30 (3): 484–497. 7

Moser, Frederick C. 1968. *Democracy and the Public Service.* Oxford University Press. 2

Mummolo, Jonathan. 2018. "Modern Police Tactics, Police-Citizen Interactions, and the Prospects for Reform." *The Journal of Politics* 80 (1): 1–15. 2

Nanes, Matthew and Dotan Haim. 2020. "Self-Administered Field Surveys on Sensitive Topics." *Journal of Experimental Political Science*. 20

Nix, Justin, Justin T Pickett, and Scott E Wolfe. 2020. "Testing a Theoretical Model of Perceived Audience Legitimacy: The Neglected Linkage in the Dialogic Model of Police–Community Relations." *Journal of Research in Crime and Delinquency* 57 (2): 217–259. 37

Paoline, Eugene A. 2003. "Taking Stock: Toward a Richer Understanding of Police Culture." *Journal of Criminal Justice* 31 (3): 199–214. 6

Paoline, Eugene A., Stephanie M. Myers, and Robert E. Worden. 2000. "Police Culture, Individualism, and Community Policing: Evidence from Two Police Departments." *Justice Quarterly* 17 (3): 575–605. 6, 8, 10, 13, 24, 36, 38

Paolini, Stefania, Jake Harwood, and Mark Rubin. 2010. "Negative Intergroup Contact Makes Group Memberships Salient: Explaining Why Intergroup Conflict Endures." *Personality and social Psychology bulletin* 36 (12): 1723–1738. 36

Paolini, Stefania, Jake Harwood, Mark Rubin, Shenel Husnu, Nicholas Joyce, and Miles Hewstone. 2014. "Positive and Extensive Intergroup Contact in the Past Buffers Against the Disproportionate Impact of Negative Contact in the Present." *European Journal of Social Psychology* 44 (6): 548–562. 30

Pepinsky, Thomas B, Jan H Pierskalla, and Audrey Sacks. 2017. "Bureaucracy and Service Delivery." *Annual Review of Political Science* 20 (1). 1, 30

Perry, James and Lois Wise. 1990. "The Motivational Bases of Public Service." *Public Administration Review* 50 (3): 367–373. SI 14

Pettigrew, Thomas F. 1998. "Intergroup Contact Theory." *Annual review of psychology* 49 (1): 65–85. 3, 25, 29, 30

Pettigrew, Thomas F. and Linda R. Tropp. 2008. "How Does Intergroup Contact Reduce Prejudice? Meta-Analytic Tests of Three Mediators." *European Journal of Social Psychology* 38 (6): 922–934. 25, 26, 30

Pettigrew, Thomas F. and Linda R. Tropp. 2011. *When Groups Meet, The Dynamics of Intergroup Contact*. Philadelphia, PA: Psychology Press. 36

Pfaff, Steven, Charles Crabtree, Holger L. Kern, and John B. Holbein. 2021. "Do Street-Level Bureaucrats Discriminate Based on Religion? A Large-Scale Correspondence Experiment among American Public School Principals." *Public Administration Review* 81 (2): 244–259. 2

Rainey, Carlisle. 2014. "Arguing for a Negligible Effect." *American Journal of Political Science* 58 (4): 1083–1091. 21

Ricks, Jacob I. 2016. "Building Participatory Organizations for Common Pool Resource Management: Water User Group Promotion in Indonesia." *World Development* 77 : 34 – 47. 30

Rosenbaum, Dennis P., Amie M. Schuck, Sandra K. Costello, Darnell F. Hawkins, and Marianne K. Ring. 2005. "Attitudes Toward the Police: The Effects of Direct and Vicarious Experience." *Police Quarterly* 8 (3): 343–365. 25

Rosenbaum, Dennis P., Sandy Yeh, and Deanna L. Wilkinson. 1994. "Impact of Community Policing on Police Personnel: A Quasi-Experimental Test." *Crime & Delinquency* 40 (3): 331–353. 10

Sabatier, Paul A, John Loomis, and Catherine McCarthy. 1995. "Hierarchical Controls, Professional Norms, Local Constituencies, and Budget Maximization: An Analysis of US Forest Service Planning Decisions." *American Journal of Political Science*: 204–242. 6

Schäfer, Sarina J, Mathias Kauff, Francesca Prati, Mathijs Kros, Timothy Lang, and Oliver Christ. 2021. "Does Negative Contact Undermine Attempts to Improve Intergroup Relations? Deepening the Understanding of Negative Contact and its Consequences for Intergroup Contact Research and Interventions." *Journal of Social Issues* 77 (1): 197–216. 36

Schuck, Amie M and Dennis P Rosenbaum. 2005. "Global and Neighborhood Attitudes Toward the Police: Differentiation by Race, Ethnicity and Type of Contact." *Journal of quantitative criminology* 21 : 391–418. 9

Skogan, Wesley G. 2004. *Community Policing: Can it Work*. Belmont: Wadsworth Publishing Co. 5

Skogan, Wesley G and Susan M Hartnett. 1997. *Community Policing, Chicago Style*. Oxford University Press. Oxford University Press. 1, 5, 6

Skolnick, Jerome H. 2011. *Justice Without Trial: Law Enforcement in Democratic Society* (4th ed.). Quid pro books. 8

Slough, Tara. 2022. "Squeaky Wheels and Inequality in Bureaucratic Service Provision." *Working Paper*. 2

Sparrow, Mark, Harrison Moore, and David M. Kennedy. 1990. *Beyond 911: A New Era for Policing*. New York: Basic Books. 1, 6, 8

Tajfel, Henri. 1978. *Differentiation Between Social Groups: Studies in the Social Psychology of Intergroup Relations*. psycnet.apa.org. Academic Press. 26

Terrill, William and Michael D. Reisig. 2003. "Neighborhood Context and Police Use of Force." *Journal of Research in Crime and Delinquency* 40 (3): 291–321. 8, 36

Todd, Andrew R and Adam D Galinsky. 2014. "Perspective-Taking as a Strategy for Improving Intergroup Relations: Evidence, Mechanisms, and Qualifications." *Social and Personality Psychology Compass* 8 (7): 374–387. 7

Toral, Guillermo. 2023. "How Patronage Delivers: Political Appointments, Bureaucratic Accountability, and Service Delivery in Brazil." *American Journal of Political Science*. 2, 6

Tsai, Lily L. 2007. *Accountability Without Democracy.* Cambridge University Press. Cambridge University Press. 2, 30

Turgoose, David, Naomi Glover, Chris Barker, and Lucy Maddox. 2017. "Empathy, Compassion, Fatigue, and Burnout in Police Officers Working with Rape Victims." *Traumatology* 23 (2): 205. 7

Twersky-Glasner, Aviva. 2005. "Police Personality: What is it and Why are they Like That?" *Journal of police and criminal psychology* 20 (1): 56–67. 8

Weisburd, David and John E Eck. 2004. "What Can Police do to Reduce Crime, Disorder, and Fear?" *The annals of the American academy of political and social science* 593 (1): 42–65. 4

Weisburd, David, Cody W Telep, Joshua C Hinkle, and John E Eck. 2010. "Is Problem-Oriented Policing Effective in Reducing Crime and Disorder: Findings from a Campbell Systematic Review." *Criminology and Public Policy* 9 (1): 139 – 172. 14

White, Ariel R, Noah L Nathan, and Julie K Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109 (1): 129–142. 2

Zaki, Jamil. 2014. "Empathy: a Motivated Account." *Psychological bulletin* 140 (6): 1608. 7

Supporting Information

How Does Community Policing Affect Police Attitudes?
An Experimental Test and a Theory of Bureaucrat-Citizen Contact

American Political Science Review

Dotan Haim, Matthew Nanes, and Nico Ravanilla

Table of Contents

A.1 Intervention and Sampling Details

The One Sorsogon program we evaluate in this research was a large province-wide community policing initiative rolled out by the Sorsogon PNP. As we discuss in the main text, the impetus behind this initiative was to repair citizen-police relations in Sorsogon in the wake of the violent war against drugs occurring elsewhere in the Philippines. In this section, we begin by describing the nature of community-facing police activities in Sorsogon prior to the intervention. We then describe the One Sorsogon community policing program, which represented a wholesale change in the citizen-facing activities for regular beat officers assigned to treatment. Finally, we include details on sampling for the randomized intervention.

A.1.1 Community Policing Before One Sorsogon

As we describe in detail in SI Section A.6 on research ethics, the Sorsogon Police Chief who worked with us on the experiment was a known reformer who implemented several citizen outreach programs in the two years he was assigned to Sorsogon prior to our experiment. This included a program to engage with at-risk youth at schools, a program to connect students interested in law enforcement careers with similar students in other countries via Facebook, and efforts to improve the PNP's social media image. However, the previous versions of "community policing" in Sorsogon, and in the Philippines generally, were of limited scope, specifically with regard to which officers participated in these programs. Each municipal station had a single officer who was assigned as the "police community relations" (PCR) director, and whose role centered on public outreach. The Sorsogon PNP chief's previous citizen outreach programs were implemented almost exclusively by these specialized PCR officers (who were excluded from our sampling frame). The PNP did hold periodic meetings with barangay officials to identify local issues, referred to as "Enhanced Managing Police Operation" (EMPO) meetings, but these meetings were also led by the specialized PCR officers and were far less frequent than our POP meetings. On our baseline survey of the barangay leaders, nearly 40% reported no contact with the police in the six-month period prior to the intervention. On the plus side, the existence of the EMPO meetings provided a baseline level of familiarity with the model for station leadership and PCR directors that aided their ability to support the wider POP intervention.

The ordinary beat patrol officers in Sorsogon who were eligible for the intervention had very limited involvement in any of these programs. These officers virtually never engaged in the usual components of community policing, such as structured engagement with community leaders, extensive citizen-facing foot patrols, or leading town-hall meetings with ordinary citizens. Officers we spoke with agreed that our version of POP – especially the regularity of meetings and intensity of follow-ups to identified problems – constituted a wholesale change. Therefore, we view it as unlikely that our null findings stem from insufficient differences between existing and newly-implemented practices of community policing.

A.1.2 The "One Sorsogon" Community Policing Program

The core components of the intervention were 1) a two-day training program on community policing, 2) monthly "Problem-Oriented Policing" (POP) meetings with barangay leaders at the municipal station, 3) citizen-facing foot patrols in treatment barangays between the meetings, 4) monthly meetings with the Station Chief and PCR Director to discuss progress, and 5) a culminating town hall with barangay citizens. Treatment officers participated in all these activities, while control officers participated in none of them. Disaggregating the effects of the various aspects of community policing is certainly an important outstanding question in the literature, but we chose to bundle the treatment in order to stay true to how community policing is normally implemented, as well as to implement the strongest version of the program in the face of resource constraints.

Before the community policing program began, PNP officers assigned to POP teams and one LGU representative from each participating barangay attended a two-day training workshop in Sorsogon City. The PIs collaborated

with PNP leadership and a Manila-based consultant to develop the workshop, which aimed to provide officers and local government unit (LGU) representatives with a comprehensive understanding of the theory of change behind community policing, as well as to explain their roles in the POP meetings they would attend over the subsequent seven-month intervention. At the training, treatment officers were paired with the barangay leaders in their assigned barangay, practiced using crime reports and barangay logbooks to identify the most pressing local public safety issues, learned techniques for empathetic communication, and focused on recognizing and reporting officer misconduct, including corruption

From December 2017 to May 2018, POP teams from the participating barangays met on a monthly basis in the municipal police station. During Meeting 1, each team discussed current public safety problems in the barangay, and identified one or two issues that would provide the focus for their intervention. Teams were provided with barangay-level data showing the types of crime that citizens experienced most often, along with anonymized logs of SMS tips sent to the PNP that referenced their barangay. Barangay leaders were encouraged to bring their barangay public safety logbooks to the meeting, which are overseen by the chief tanod and the barangay secretary.

In the second and third meetings, the teams were tasked with crafting a strategy to address their selected issue(s) and developing a proposed budget of up to 5,000 pesos (about $100 US) for implementation. After reviewing the plans for compliance with funder regulations, the research staff provided the requested budget to each POP team. In the fourth, fifth and sixth months of the intervention, teams dedicated their resources to implementing their projects. Some barangays used the opportunity to increase foot patrols and used their budgets for basic equipment like flashlights, whistles and rain ponchos. Other barangays took steps to improve relevant infrastructure, like constructing fences along highways to prevent stray dogs from causing traffic accidents, installing streetlights in strategic locations to reduce incidents of vagrancy and public intoxication, and adding road safety signage. To address more complex public safety issues, such as juvenile delinquency, barangays implemented activities to engage those at risk. One POP team chose to purchase sports equipment for at-risk youth to use after school, while another POP team created a community garden tended to by at-risk youth.

During the intervention period, all participating PNP officers were encouraged by their superiors to travel to their assigned barangay at least once between each POP meeting. Officers often helped the barangay leadership implement the solutions they decided on during POP meetings, and informally met with citizens to get their perspective on the focus public safety issue. Subsequently, participating barangays held culminating town halls to reinforce their progress and discuss ways to continue activities that would promote public safety in the future. The members of the POP team led the town halls by describing their activities and plans, after which there was a long period of Q&A and unstructured discussion. On average, these culminating town halls were attended by nearly 80 citizens and lasted more than three hours. We provide additional details about the intervention in the Supporting Information (SI). The key characteristic of the POP program is that each barangay's POP team decided individually what projects they would take on, but all teams operated within the same structure and shared the same goal of solving community-specific issues through preventive interventions.

Municipal police chiefs and "Police Community Relations" (PCR) directors were instructed by the Provincial Police Office (PPO) to conduct monthly check-ins with each pair of treatment officers assigned to a POP team to discuss progress and surface any citizen reports of police corruption. Officers were aware that the municipal chiefs and PCR directors also met monthly with the Provincial PNP Chief and his staff to discuss the progress of his flagship community policing program. One goal of having officers work in teams was to create the sense that corrupt behavior may be reported by officers' well-meaning colleagues, especially after the common knowledge about reporting practices was generated by the training. Importantly, the One Sorsogon program started shortly after President Duterte and the PNP Director-General stated that they were instituting a nationwide pause to the Drug War to instead focus on tackling police corruption (link).

As described above, a substantial portion of the treated officers' formal interactions as part of One Sorsogon were with barangay leaders rather than "ordinary" citizens. We designed the intervention this way in part to reflect

the types of "citizens" focused on by prior research on community policing. Community policing interventions, and POP in particular, often focus on increasing contact between police officers and "community leaders," who typically have a bigger picture understanding of crime in their community and can better assist the police in implementing reforms. Barangay captains and councilors are elected officials, and tanods are appointed by the barangay captain, but none of these positions is a full-time job, and these individuals are only loosely tied to municipal and national-level governments. For example, only 60% of barangay leaders said they had contacted the police in the *six-month* period prior to the intervention, suggesting that the intervention represented a substantial change to police officers' interactions with barangay leaders. In the context of police-community relations, the barangay leaders on the POP teams are best viewed as advocates for their barangay, rather than an arm of the national government. From the perspective of the PNP, barangay leaders were discussed as "citizen representatives," and they play a similar role to informal village leaders and community watch members in other parts of the Global South. In other words, we view the community leaders involved in the POP intervention as being comparable to a class of "citizens" discussed in much of the prior research on community policing.

In addition to the randomization of PNP officers into the intervention, several aspects of One Sorsogon program were randomized across barangays. The component of the project that is the focus of this paper was part of a larger effort to evaluate community policing, including citizen-facing aspects. From the 298 barangays in our sample, 99 were assigned to receive POP teams that included PNP officers, while 99 received a version of the intervention that was led only by the LGU officials. The remaining 100 barangays were "control" units that received policing-as-usual. In addition to the POP teams, another aspect of One Sorsogon was that the PNP upgraded a previously underutilized hotline that citizens could use to contact the police, paired with a large-scale drive to advertise the hotline. We provide further details on the intervention and the research design we used to evaluate the citizen-facing aspects of the program in the PAP. The officer randomization only occurred *within* the first category of treatment barangays in the larger study (those that received POP teams including PNP officers), meaning we cannot leverage the barangay-level randomization for analyzing the effects on officers. The barangay-level randomization only matters insofar as it increases the representativeness of the study areas to the larger sample of barangays in Sorsogon (which speaks to external validity).

A.1.3 Sampling

The primary sampling unit for the geographical randomization is the barangay, the smallest administrative unit in the Philippines. As discussed in-text, we began by identifying the barangays in Sorsogon Province which have sufficient government control as to make the community policing intervention feasible and safe. From there, we randomly assigned 99 barangays to receive the POP treatment with only LGU members, 99 brangays to receive the POP treatment with both LGU and PNP members, and 100 barangays to receive the control condition of "policing as usual," i.e. largely reactive policing that generates little contact between civilians and police officers in non-emergency situations.

We then surveyed every police officer in Sorsogon province twice, once shortly before the intervention began and a second time shortly after it ended. The survey was conducted via in-person self enumeration: trained research staff visited each municipal police station at a pre-appointed time, explained the study to the police officers, and gave them tablet computers loaded with the survey. Officers completed the survey on their own and then returned the tablets to the staff. Staff remained available to answer any questions during the survey, but officers entered their responses themselves in an effort to ensure privacy.

For the officer-level randomization, we relied on officers who took the baseline officer survey as our sampling frame. We excluded non-uniformed personnel and the upper-level leadership in each municipal police station because the intervention did not fit their professional responsibilities. Officer sampling was blocked on police stations in order to assign each treatment officer to participate on a POP team in the jurisdiction where they were

assigned. Second, we blocked on officer rank. The PNP wanted to assign at least one "senior" officer to each team to ensure that officers had the experience necessary to successfully lead the intervention.

A.2 Descriptive Statistics

A.2.1 Summary Statistics

Tables A.1 and A.2 display descriptive statistics for the officers in our sampling frame and for the municipalities where the intervention was implemented, respectively. Municipality sociodemographic variables are from the 2010 Philippine Census, while the NPA presence variable is from AFP intelligence reports in December 2015. Officer characteristics are from our officer survey.

Table A.1: Summary Statistics - Survey

	Baseline			**Endline**		
Variable	N	Mean	SD	N	Mean	SD
Treated	691	0.28	0.45	705	0.33	0.47
Age	679	35	6.9	678	35	6.8
Male	688	0.76	0.42	704	0.77	0.42
Completed College	687	0.18	0.39	701	0.2	0.4
Catholic	688	0.91	0.29	0		
Rank (Senior)	691	0.2	0.4	705	0.18	0.38
Years Served	691	10	6.5	703	9.8	6.4
Years Served in Mun	684	4.3	4.4	699	4.4	4.6
Lives in Province	691	0.87	0.33	705	0.88	0.33
Lives in Municipality	691	0.54	0.5	705	0.56	0.5
Share Concerns	625	1.5	0.84	698	1.3	0.85
Officer Attitude Index	528	-0.038	0.54	568	-0.0098	0.55
Officer Attitude Idx (alt)	676	-0.021	0.53	698	-0.0066	0.56
Trust Index	665	-0.13	0.83	680	0.0037	0.82
Trust Index (alt)	678	-0.14	0.84	694	-0.0021	0.82
Trust (Safe)	682	71	26	690	75	26
Trust (Intent)	674	76	22	693	80	22
Trust (Info)	671	77	19	693	79	20
Empathy Index	665	-0.065	0.77	676	-0.016	0.8
Empathy Index (alt)	684	-0.058	0.77	700	-0.018	0.8
Empathy (Complaints)	669	62	27	680	62	29
Empathy (Reports)	680	82	18	696	84	19
Accountability Index	581	-0.04	0.67	612	0.0092	0.63
Accountability Idx (alt)	609	-0.039	0.67	641	0.0073	0.63
Accountability (Care)	683	87	17	694	90	15
Accountability (Hyp1 - Punish)	678	3	1.4	695	2.8	1.3
Accountability (Hyp1 - Report Self)	632	77	28	667	80	25
Accountability (Hyp1 - Report Other)	624	76	28	663	78	26
Accountability (Hyp2 - Punish)	680	2.5	1.3	692	2.5	1.3
Accountability (Hyp2 - Report Self)	631	71	29	655	72	28
Accountability (Hyp2 - Report Other)	627	69	29	645	71	28
Corruption Idx	614	0.059	0.8	660	-0.0091	0.76
Corruption Idx (alt)	664	0.048	0.79	690	-0.0059	0.76
Corruption (Hyp1 - Self)	661	93	18	691	94	17
Corruption (Hyp1 - Other	661	90	22	688	91	22
Corruption (Hyp2 - Self)	664	78	27	690	81	27
Corruption (Hyp2 - Other	663	76	28	688	77	30

Table A.2: Summary Statistics - Municipalities

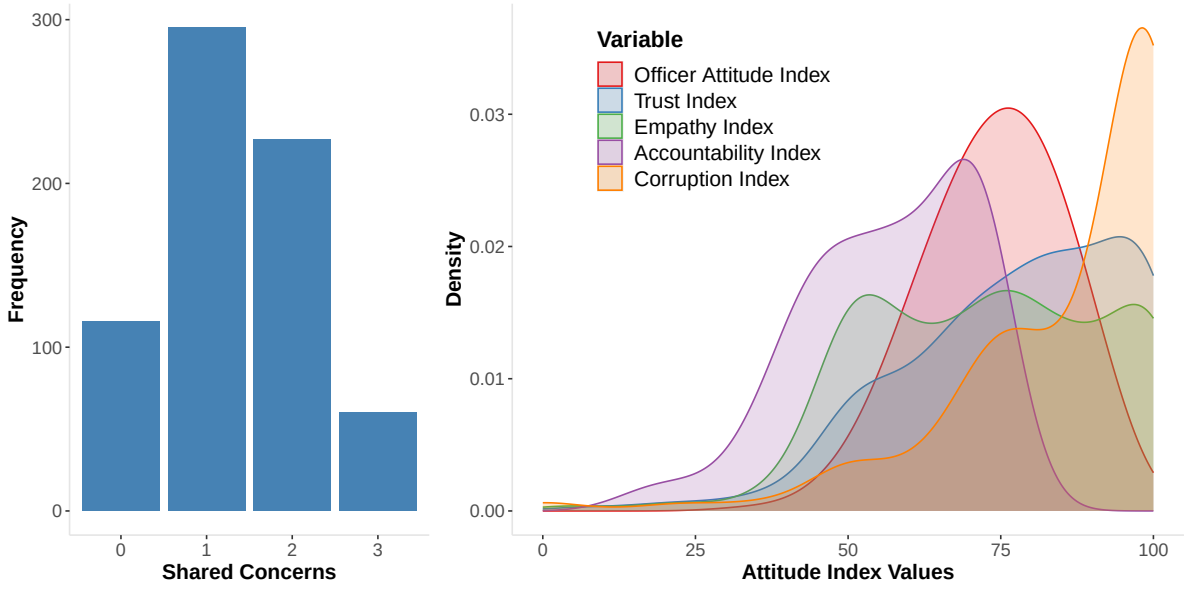
Variable	N	Mean	SD	Min	Max
Population (mun)	17	34,136	27,277	9,338	86,728
Household Size (mun)	17	3.3	0.18	2.8	3.5
Education (mun)	17	1.8	0.13	1.7	2.1
Percent Catholic (mun)	17	0.95	0.011	0.93	0.98
Home Quality (mun)	17	0.38	0.0074	0.37	0.39
Avg. NPA Presence 2015 (mun)	17	0.34	0.36	0	1.2

A.2.2 Outcome Measures and Index Construction

The "Shared Concerns" outcome was calculated in four steps. First, we asked officers: "Below is a list of public safety concerns. What do you consider to be the three most important public safety concerns in the municipality where you are assigned." Officers selected from the 13 public safety concerns displayed in Table 1 in the article. Second, we asked the same question (specific to the barangay) on our survey of the citizen barangay leaders, which included the barangay captain, barangay counselor in charge of peace and order, and the chief tanod. Third, we aggregated the ranking of these issues from the barangay leader survey to the municipality-level. Fourth, for each officer, we counted how many of the officer-identified issues matched the four most common issues from the aggregated barangay leader survey responses in the same municipality. Figure A.1(a) shows the distribution of the *Shared Concerns* measure at baseline. In Section A.4.2, we show that the results are robust to alternative ways of measuring this variable.

Figure A.1: Outcome Distributions

(a) Shared Concerns (b) Attitude Indices



The above figures show the distribution of the outcome variables at baseline. The left panel shows the distribution of the shared concerns (knowledge) variable, which counts how many of the most important public safety issues identified by officers match the issues identified by community leaders in the same municipality. The right panel shows the distribution of the raw attitudinal indices.

Item	Question	Options
	TRUST	
trust_safe	How safe for officers is the municipality where you are assigned.	0-Very Dangerous; 100-Very Safe
trust_intent	Citizens care about the well-being of police officers.	0-Strongly disagree; 100-Strongly agree
trust_info	Information that I receive from civilians is likely to be accurate.	0-Strongly disagree; 100-Strongly agree
Trust Index		Index of: *trust_safe, trust_intent, trust_info*
	EMPATHY	
empathy_complain	When people complain about the police, it's because they have a good reason.	0-Strongly disagree; 100-Strongly agree
empathy_report	Most things that people report to the police are worth taking seriously.	0-Strongly disagree; 100-Strongly agree
Empathy Index		Index of: *empathy_complain, empathy_report*
	HYPOTHETICAL SCENARIOS	
	A police officer stops a motorist for speeding. The officer agrees to accept a personal gift of half of the amount of the fine in exchange for not issuing a citation.	
h2_corrupt_self	Do you consider this behavior to be serious misconduct?	0-Not at all serious; 100-Very serious
h2_corrupt_other	Do most police officers consider this behavior to be serious misconduct?	0-Not at all serious; 100-Very serious
h2_punish	If an officer engaged in this behavior and was discovered doing so, what if any discipline do you think will follow?	0-None; 1-Verbal reprimand; 2-Written reprimand; 3-Period of suspension without pay; 4-Demotion in rank; 5-Dismissal
h2_report_self	Do you think you would report a fellow police officer who engaged in this behavior?	0-Definitely not; 100-Definitely yes
h2_report_other	Do you think most police officers would report a fellow police officer who engaged in this behavior?	0-Definitely not; 100-Definitely yes
	A police officer routinely accepts free meals, cigarettes, and other items of small value from merchants on his beat. He does not solicit these gifts and is careful not to abuse the generosity of those who give gifts to him.	
h3_corrupt_self	Do you consider this behavior to be serious misconduct?	0-Not at all serious; 100-Very serious
h3_corrupt_other	Do most police officers consider this behavior to be serious misconduct?	0-Not at all serious; 100-Very serious
h3_punish	If an officer engaged in this behavior and was discovered doing so, what if any discipline do you think WILL follow?	0-None; 1-Verbal reprimand; 2-Written reprimand; 3-Period of suspension without pay; 4-Demotion in rank; 5-Dismissal
h3_report_self	Do you think you would report a fellow police officer who engaged in this behavior?	0-Definitely not; 100-Definitely yes
h3_report_other	Do you think most police officers would report a fellow police officer who engaged in this behavior?	0-Definitely not; 100-Definitely yes
	ACCOUNTABILITY	
account_care	The police leadership takes citizen complaints about officers seriously.	0-Strongly disagree; 100-Strongly agree
Accountability Index		Index of: *account_care, h2_punish, h2_report_self, h2_report_other, h3_punish, h3_report_self, h3_report_other*
	CORRUPTION	
Corruption Index		Index of: *h2_corrupt_self, h2_corrupt_other, h3_corrupt_self, h3_corrupt_other*
	NPA SYMPATHY	
npa_sympathy	What percentage of civilians in the municipality where you are assigned would you estimate are sympathetic to the NPA?	0%; 100%

The attitudinal outcome indices — *Trust*, *Empathy*, *Accountability*, and *Corruption* — were created by adding the values of individual survey items (displayed above), and then dividing by the number of items. The umbrella *Officer Attitude Index* is an index of indices, created by adding the values of the other attitudinal indices and dividing by the number of constituent indices. Most individual survey items were captured using a slide bar in ISurvey that translated to a 100-point scale when the data were exported. The exceptions were *h2_punish* and *h3_punish*, part of the accountability index, which were measured on a 6-point ordinal scale. Following the PAP, we standardized each survey item before combining them into the index by subtracting the mean and dividing by the standard deviation of the variable at baseline. Figure A.1(b) displays the distribution of the attitudinal indices at baseline. For ease of interpretation, the figure displays versions of the indices created using the raw, non-standardized item responses.

In Table A.12, we use an alternative version of the main indices where values for missing items were imputed using other items from the same respondent. For example, the trust index is made up of three sub-items: *trust_info*,

trust_safe and *trust_intent*. To impute values of *trust_info*, we regress *trust_info* on *trust_safe* and *trust_intent* and extract the predicted values. We then use the same procedure to impute values for *trust_safe* and *trust_intent* before combining the items into the index. This method is described in the PAP.

A.2.3 Balance and Attrition

Table A.3 displays the balance of officer characteristics based on their treatment status. To account for the blocked randomization, we display outputs of regressions that predict treatment status based on each pre-treatment officer characteristic, including fixed effects for the blocking variables. In other words, the displayed estimates compare treatment officers to control officers in their same station and rank. None of these differences are statistically significant at conventional levels, though treatment officers were marginally less educated than control officers. In SI section A.4, we show that the main results hold when controlling for officer education.

In Table A.4, we analyze attrition. Officer turnover over the 8 months between our baseline survey and endline survey was relatively high. Of the 705 eligible officers we interviewed at baseline (rank PO or SPO), 190 dropped out from the sample before endline. The majority of attrition was due to officers being reassigned to different provinces, being assigned to training in the regional/national capitol, and retirements. Between baseline and endline, 183 new officers who would otherwise have been eligible to participate on a community policing team were hired or transferred to Sorsogon.

Table A.3: Balance Table

Variable	**Estimate**	**p-value**
Age	0.004	0.208
Male	0.026	0.466
Education	-0.071	0.041
Catholic	-0.047	0.564
Lives in Sorsogon	0.059	0.262
Commute	-0.013	0.336
Years Served	-0.001	0.836
Share Citizen Concerns	0.034	0.105
Officer Attitude Index	-0.030	0.286
Trust Index	-0.009	0.687
Empathy Index	-0.020	0.246
Accountability Index	-0.017	0.645
Corruption Index	0.011	0.692

Each row reports the estimates and p-values from regressions that predict treatment status based on pre-treatment officer attributes (demographic and attitudinal). All regressions include station and rank fixed effects to account for blocking. Standard errors are clustered at the station level.

As a result of this turnover, two comparisons are particularly important. First, were the officers who dropped out of our sample after baseline significantly different from the officers who remained in the sample? We find that these two groups were very similar. Most importantly, officers' treatment status did not significantly predict whether they dropped out (estimate=-0.06, p=.29). In Table A.4, this comparison can be seen for all demographic covariates where the Attrition Category is "Both Surveys" (the comparison group in these rows is officers who only took the baseline survey) and all survey indices where the Attrition Category is "Baseline Only" (the comparison group in these rows is officers who took both surveys). Most of these differences are insignificant. Second, because we include the group of newly hired officers in our analysis, it is important that this group of new officers who joined the sample does not differ significantly from the group of officers who left the sample. In Table A.4, this comparison can be seen for all demographic covariates where the Attrition Category is "Endline Only" (the comparison group

Table A.4: Attrition Table

Variable	Attrition Category	Estimate	p-value
Male	Endline Only	-0.029	0.30
Male	Both Surveys	0.002	0.97
Age	Endline Only	-3.504	0.08
Age	Both Surveys	-0.317	0.70
College Educated	Endline Only	0.014	0.57
College Educated	Both Surveys	-0.025	0.47
Catholic	Both Surveys	-0.008	0.59
Years Served	Endline Only	-3.030	0.10
Years Served	Both Surveys	-0.310	0.73
Lives in Sorsogon	Endline Only	0.009	0.87
Lives in Sorsogon	Both Surveys	0.097	0.01
NPA Presence 2015	Endline Only	0.016	0.87
NPA Presence 2015	Both Surveys	0.042	0.23
Crime Rate	Endline Only	0.023	0.72
Crime Rate	Both Surveys	-0.029	0.24
Attitude Idx (endline)	Endline Only	-0.055	0.20
Trust Idx (endline)	Endline Only	-0.124	0.26
Empathy Idx (endline)	Endline Only	-0.052	0.31
Account Idx (endline)	Endline Only	0.148	0.00
Corrupt Idx (endline)	Endline Only	-0.226	0.00
Attitude Idx (baseline)	Baseline Only	0.067	0.24
Trust Idx (baseline)	Baseline Only	-0.101	0.28
Empathy Idx (baseline)	Baseline Only	-0.001	0.99
Account Idx (baseline)	Baseline Only	0.057	0.31
Corrupt Idx (baseline)	Baseline Only	0.074	0.31

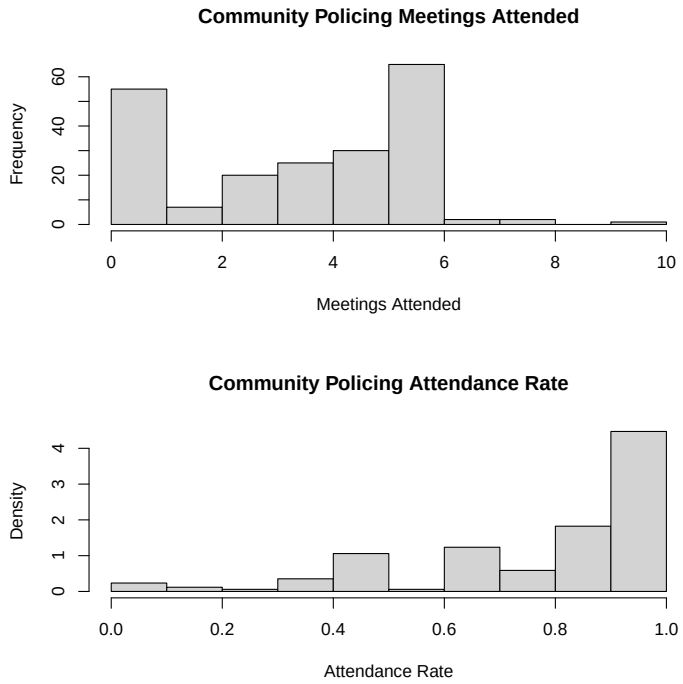
Each row reports the estimates and p-values from regressions that predict covariates based on officers' attrition status. These estimates are calculated without respect to the blocks unlike in Table A.3, since only treatment was assigned with respect to the blocks. Attrition Category is a categorical variable with the following levels: *Baseline Only* (officers who were assigned in Sorsogon at baseline but attritioned before endline), *Endline Only* (officers who were newly assigned in Sorsogon after baseline), and *Both Surveys* (officers who were assigned in Sorsogon through the duration of the experiment). The comparison group for the demographic variables is *Baseline Only* (with the exception of Catholic, which was measured only at endline). The comparison group for all survey indices is *Both Surveys*. Because the endline survey indices (last four rows) are measured post-treatment, these regressions are subset to officers in the control group.

in these rows is officers who only took the baseline survey) and all survey indices where the Attrition Category is "Endline Only" (the comparison group in these rows is officers in the control group who took both surveys).

A.2.4 Treatment Compliance

Figure A.2 displays 1) the number of POP meetings that each treatment officer attended (regardless of whether they later had to drop off a POP team due to reassignment) and 2) the POP meeting attendance rate (subset to the officers who were still active members of a POP team). Both histograms are subset to officers who were randomly selected to participate in the community policing intervention. The first histogram shows that the modal officer attended all six POP meetings, but a number of treatment officers attended zero community policing meetings. This occurred because, right after the intervention began, a number of officers were assigned to a training program in Manila, which also corresponded with a batch of officer reassignments to different stations. Because these officers underwent the community policing training, we consider them "treated" when calculating the ITT. A few officers attended more than six meetings because they temporarily stepped in for a colleague on another barangay's POP team. The second histogram displays the "attendance rate," which is calculated as the number of meetings attended

Figure A.2: Treatment Compliance and Intensity



The top histogram displays the number of POP meetings attended by each treatment offier, regardless of whether they later had to drop off a POP team due to reassignment. The bottom histogram is subset to officers who were still active members of a POP team, excluding reassigned officers.

divided by the number of meetings to which an officer was *assigned*. If, during the intervention, an officer was removed from a POP team, we no longer categorized them as being assigned to attend future meetings.

The above data only allow us to measure compliance with the meetings with barangay leaders, but not interactions with "ordinary" citizens. We do not have systematic data on the number of citizen-facing foot patrols officers participated in between POP meetings, but we note the following relevant facts:

- The intervention concluded with a culminating town hall to which all barangay citizens were invited, and this was officers' last assigned activity before the endline survey. On average, these culminating town halls were attended by nearly 80 citizens and lasted more than three hours. Attendance at the citizen-facing culminating town hall among officers assigned to POP was over 90%.
- Our field officers, who attended each POP meeting, reported that most officers visited their assigned barangay at least once between each POP meeting. These barangay visits typically included a conversation with the barangay leaders, but also included foot patrols targeted at increasing contact with ordinary citizens.
- On our representative endline survey of "ordinary" citizens, we asked respondents if they had interacted with a PNP officer as part of the One Sorsogon program. In treatment villages, 15% of respondents indicated that they had interacted with an officer as part of the program, while only 5% of respondents in control villages reported such an interaction.
- Between the fourth and fifth POP meetings, our field officers conducted an "audit" of whether the assigned officers had interacted with citizens and whether citizens were aware of details of the One Sorsogon program. Field officers were tasked with speaking with one tanod and one barangay counselor who were not on the POP team, plus three additional citizens in the area around the barangay center, using a simple random

walk protocol. We asked citizens to identify the public safety issue focused on by the POP team, to identify actual awareness. In this audit, 40% of ordinary barangay residents correctly identified the issue that the POP team was working on, without any prompting.

A.3 Main Results

In this section, we show the full regression tables underlying the figures in the manuscript. Table A.5 shows the regression results displayed in Figure 1. Table A.6 shows the regression results displayed in Figure 2. Table A.7 shows the regression results displayed in Figure 4.

Table A.5: Main Treatment Effects

	Share Concerns		Attitude Idx		Trust		Empathy		Accountability		Corruption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated	0.21*	0.25**	0.06	−0.01	−0.06	−0.02	−0.12	−0.08	−0.00	0.01	0.12	0.06
	(0.10)	(0.10)	(0.06)	(0.04)	(0.06)	(0.06)	(0.10)	(0.09)	(0.07)	(0.07)	(0.08)	(0.09)
Base Outcome	0.27***		0.26***		0.34***		0.28***		0.43***		0.26***	
	(0.03)		(0.06)		(0.05)		(0.04)		(0.04)		(0.06)	
Model	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross
R^2	0.13	0.07	0.10	0.04	0.21	0.13	0.10	0.04	0.16	0.05	0.10	0.05
Observations	627	632	460	534	656	671	652	672	536	587	589	621
Clusters	18	18	18	18	18	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Panel models include a dummy variable for missing baseline.
All models include station (municipality) and rank fixed effects for the blocking variables.
Standard errors are clustered at the station-level.

Table A.6: Heterogeneous Treatment Effects - Embeddedness

	Share Concerns	Attitude Index	Trust	Empathy	Account	Corrupt	Attitude Idx (alt)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.28*	0.31**	0.29	0.28	0.08	0.12	0.21
	(0.16)	(0.13)	(0.26)	(0.20)	(0.17)	(0.30)	(0.14)
Lives in Province	0.19*	0.17***	0.20***	0.23	0.11	−0.05	0.11***
	(0.09)	(0.03)	(0.07)	(0.17)	(0.09)	(0.14)	(0.02)
Treat*InProvince	−0.09	−0.28**	−0.38	−0.43*	−0.09	−0.00	−0.21
	(0.16)	(0.12)	(0.28)	(0.21)	(0.18)	(0.31)	(0.14)
Base Outcome	0.27***	0.26***	0.34***	0.27***	0.43***	0.26***	0.24***
	(0.03)	(0.06)	(0.05)	(0.04)	(0.04)	(0.06)	(0.05)
R^2	0.13	0.10	0.21	0.11	0.16	0.10	0.08
Observations	627	460	656	652	536	589	679
Clusters	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. All models also include station (municipality) and rank fixed effects for the blocking variables. Standard errors are clustered at the station-level.

Table A.7: Heterogeneous Treatment Effects - Rebel Presence (Trust Outcomes)

	Trust Index	Trust (Safe)	Trust (Intent)	Trust (Info)	Perceived NPA Sympathy
	(1)	(2)	(3)	(4)	(5)
Treated	0.17	4.86*	6.24**	2.26	−6.86**
	(0.11)	(2.75)	(2.90)	(1.96)	(2.83)
NPA Presence 2015	0.23	2.95	8.72**	4.32*	−5.11*
	(0.14)	(4.54)	(3.70)	(2.29)	(2.77)
Treated*NPA	−0.57***	−17.10***	−19.14***	−5.04*	12.50
	(0.15)	(4.49)	(4.42)	(2.69)	(7.82)
Base Outcome	0.40***	0.33***	0.35***	0.28***	0.41***
	(0.04)	(0.05)	(0.06)	(0.03)	(0.04)
Base Outcome	0.40***				
	(0.04)				
R^2	0.17	0.12	0.13	0.11	0.16
Observations	612	635	628	628	636
Clusters	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. All models also include rank fixed effects for the blocking variables, though they do not include station fixed effects because NPA Presence is measured at the municipality (station) level. Standard errors are clustered at the station-level.

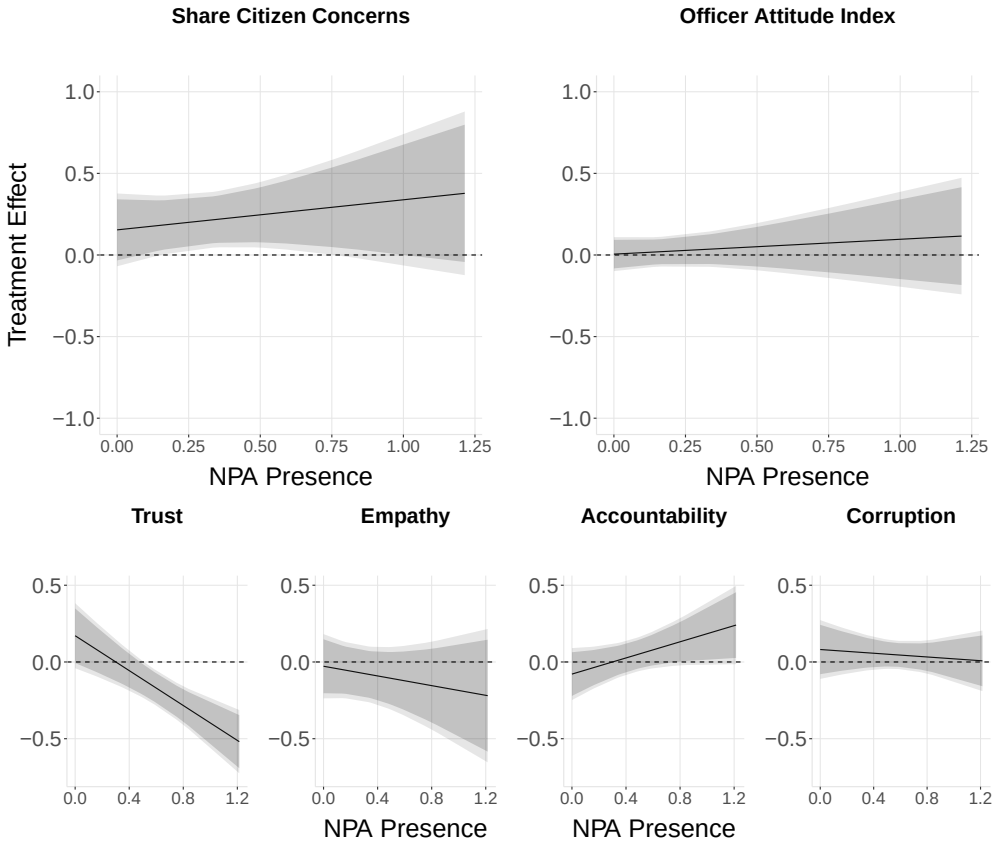
Table A.8: Heterogeneous Treatment Effects - Rebel Presence

	Share Concerns	Attitude Index	Trust	Empathy	Account	Corrupt	Trust Idx (alt)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.15	0.01	0.17	−0.03	−0.08	0.08	0.17
	(0.11)	(0.05)	(0.11)	(0.11)	(0.09)	(0.10)	(0.10)
NPA Presence 2015	0.09	−0.20**	0.23	0.09	−0.21**	−0.21***	0.25*
	(0.11)	(0.09)	(0.14)	(0.10)	(0.09)	(0.06)	(0.14)
Treated*NPA	0.18	0.09	−0.57***	−0.16	0.26*	−0.06	−0.58***
	(0.26)	(0.17)	(0.15)	(0.21)	(0.15)	(0.14)	(0.14)
Base Outcome	0.28***	0.26***	0.40***	0.27***	0.46***	0.25***	0.41***
	(0.03)	(0.07)	(0.04)	(0.04)	(0.04)	(0.06)	(0.05)
Base Outcome			0.40***				0.41***
			(0.04)				(0.05)
R^2	0.09	0.07	0.17	0.08	0.14	0.08	0.17
Observations	627	431	612	610	500	554	625
Clusters	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. All models also include rank fixed effects for the blocking variables, though they do not include station fixed effects because NPA Presence is measured at the municipality (station) level. Standard errors are clustered at the station-level.

The heterogeneous effects relating to NPA presence that we present in Figure 4 and Table A.7 zoom in on the Trust index and the additional *Perceived NPA Sympathy* outcome. Figure A.3, which displays the marginal effects of regressions in Table A.8, displays the heterogeneous effect of NPA presence on all our main outcomes. The only other CATE of note is a marginally significant increase in officers' perception of accountability in areas with high NPA presence. The coefficient is only marginally significant ($p = 0.099$), we did not have qualitative evidence that pointed us to look for this heterogeneous effect, and the result is insignificant when including station and officer controls. We thus advise caution when interpreting the result. From a purely exploratory standpoint, we think it is possible that the treatment had a positive impact on officers' perceptions of accountability in high-NPA areas because police leaders in these areas are more sensitive to improving citizen-officer relationships as part of the effort to win "hearts and minds" away from the insurgents.

Figure A.3: Heterogeneous Treatment Effects - Rebel Presence (Main Outcomes)



All plots display marginal effects of treatment by NPA (rebel) presence in the municipality, using the panel specification (equation 1). The top panel displays results for shared concerns (knowledge) and the umbrella officer attitude index. The bottom panel displays results for the other main outcome indices. Confidence intervals are displayed at the 95% level (light gray) and 90% level (dark gray). Full regression results are in Table A.8.

A.4 Robustness

A.4.1 Heterogeneous Treatment Effects with Controls

In this section, we show the results of the robustness tests described in the manuscript.Table A.9 probes whether the embeddedness findings are robust to including controls for potential confounders. We include different combinations of control variables when predicting the umbrella Attitude Index in Models 1-3, and then apply the specification in the full model to the sub-index outcomes in Models 4-7. Model 1 retains the station fixed effects and includes controls for officer-level characteristics, including their rank, education, religion, age, gender, years served in the municipality, average daily commute to their assigned police station, and their public service motivation (PSM). Officers' PSM score is based on a standard battery of questions described in Perry and Wise (1990). Model 2 drops the station fixed effects and replaces them with municipality-level control variables from the 2010 census. Models 3-7 includes the full battery of controls, including their interactions with the treatment indicator. All coefficients and standard errors on the interaction term of interest remain consistent across these different specifications.

Table A.10 follows a similar format for the Rebel Presence findings. As a reminder, these models do not include station fixed effects because the NPA Presence moderator is measured at the municipality-level. Models 1-3 display the results on the Trust Index using station-level, officer-level, and interacted controls. The results are highly consistent when including non-interacted controls, including for the alternative (imputed) version of the trust index and all the individual survey items that compose the trust index (Models 4-7). The one coefficient that loses significance using the full battery of controls is the base term on the *Perceived NPA Sympathy* outcome (Model 8), though Model 9 shows that this remains significant when including only the municipality-level controls. Because the moderator is measured at the station-level, the highly-saturated model with interactive controls (Model 3) substantially inflates the standard errors in this case. Though the coefficient on the *Treated*NPA* interaction term remains consistent in this model, it is barely insignificant at conventional levels (p=0.12).

The models with controls (as well as the station fixed effects in the original models) helps address the concern that officers from inside and outside of Sorsogon are systematically assigned to police stations with different observable characteristics. That said, we explore whether these assignments appear to be balanced on municipality characteristics in Table A.11, and we do not find significant differences of municipality characteristics by officer embeddedness.

Table A.9: Heterogeneous Treatment Effects - Embeddedness (with Controls)

	Attitude Index			Trust	Empathy	Account	Corrupt
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.30**	0.29**	14.38	16.23*	7.70	9.56	−7.60
	(0.13)	(0.13)	(11.75)	(9.07)	(10.27)	(10.31)	(12.26)
Lives in Province	0.18***	0.16**	0.16**	0.11*	0.17	0.15	−0.01
	(0.02)	(0.06)	(0.06)	(0.06)	(0.13)	(0.12)	(0.14)
Treated*InProvince	−0.27*	−0.26*	−0.28**	−0.27	−0.38*	−0.06	−0.13
	(0.13)	(0.13)	(0.12)	(0.24)	(0.21)	(0.20)	(0.31)
Base Outcome	0.26***	0.25***	0.26***	0.32***	0.21***	0.43***	0.21***
	(0.07)	(0.07)	(0.07)	(0.06)	(0.04)	(0.05)	(0.06)
Rank (officer)	0.01	−0.00	0.05	−0.01	0.13	−0.02	−0.15
	(0.09)	(0.07)	(0.14)	(0.12)	(0.12)	(0.10)	(0.09)
Education (officer)	−0.01	−0.02	−0.04	−0.02	−0.09	0.01	0.11**
	(0.06)	(0.05)	(0.09)	(0.08)	(0.08)	(0.07)	(0.05)
Catholic (officer)	0.05	0.06	0.11*	0.13	0.30***	−0.05	0.12**
	(0.08)	(0.08)	(0.06)	(0.18)	(0.09)	(0.07)	(0.05)
Age (officer)	−0.01	−0.00	−0.01	0.02**	0.01	−0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Gender (officer)	−0.05	−0.03	−0.04	−0.02	−0.12*	−0.07	0.10
	(0.05)	(0.05)	(0.07)	(0.07)	(0.06)	(0.08)	(0.13)
Years Served (officer)	−0.00	−0.00	−0.00	0.01	−0.00	−0.01	−0.01
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Commute (officer)	−0.02	−0.01	−0.02	−0.01	0.02	−0.02	0.01
	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.05)
PSM (officer)	0.02	0.06	−0.01	0.19*	0.31***	0.02	−0.42***
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.08)	(0.14)
Education (mun)		0.13	0.65	1.55	0.97	−0.33	−1.70**
		(0.73)	(0.79)	(0.91)	(1.06)	(1.07)	(0.77)
Population (mun)		0.00	−0.00	−0.00*	−0.00	0.00	0.00**
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Household Size (mun)		−0.13	0.30	2.34***	0.51	−0.11	−0.66
		(0.41)	(0.32)	(0.55)	(0.39)	(0.32)	(0.60)
Percent Catholic (mun)		−0.18	0.28	5.11	−3.97	1.25	9.74**
		(3.92)	(4.68)	(5.52)	(7.62)	(8.86)	(3.85)
Crime Rate (mun)		0.34*	0.71***	−0.19	0.66***	0.26	0.47
		(0.18)	(0.21)	(0.34)	(0.20)	(0.26)	(0.35)
NPA Presence 2015 (mun)		0.05	0.11	−0.44*	0.20	0.06	0.32
		(0.17)	(0.21)	(0.21)	(0.20)	(0.15)	(0.22)
Treat*Rank			−0.06	0.29	−0.05	0.02	0.03
			(0.21)	(0.22)	(0.26)	(0.17)	(0.18)
Treat*Education			0.09	−0.01	0.21	0.00	0.02
			(0.18)	(0.18)	(0.14)	(0.23)	(0.19)
Treat*Catholic			−0.15	−0.39	−0.32	−0.11	−0.04
			(0.20)	(0.24)	(0.28)	(0.15)	(0.24)
Treat*Age			0.00	−0.03	−0.01	−0.01	0.01
			(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Treat*Gender			0.06	−0.04	−0.01	0.26	−0.23
			(0.13)	(0.13)	(0.18)	(0.15)	(0.16)
Treat*YearsServed			0.00	0.01	0.00	0.02	−0.02
			(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
Treat*Commute			0.03	0.04	−0.07	0.04	0.05
			(0.06)	(0.05)	(0.05)	(0.04)	(0.07)
Treat*PSM			0.08	0.01	0.14	0.12	0.18
			(0.09)	(0.14)	(0.13)	(0.12)	(0.24)
R^2	0.12	0.10	0.13	0.24	0.16	0.18	0.14
Observations	455	427	427	603	601	496	546
Clusters	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$

Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. Standard errors are clustered at the station-level.

Table A.10: Heterogeneous Treatment Effects - Rebel Presence (with Controls)

	Trust Index			Trust (Safe)	Trust (Intent)	Trust (Info)	Trust Idx (alt)	Perceived NPA Sympathy	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treated	0.04	0.01	16.23*	0.20	1.38	−0.03	0.03	−4.15	−5.91**
	(0.08)	(0.08)	(9.07)	(1.85)	(1.89)	(1.71)	(0.07)	(2.55)	(2.51)
NPA Presence 2015	−0.32	−0.38*	−0.44*	−17.53**	−6.73	−6.63**	−0.38*	4.03	2.41
	(0.21)	(0.20)	(0.21)	(6.12)	(5.22)	(2.80)	(0.20)	(7.46)	(6.67)
Treated*NPA	−0.37***	−0.33***	−0.35	−9.62***	−11.16***	−1.81	−0.31***	8.15	10.64
	(0.09)	(0.08)	(0.22)	(3.20)	(2.57)	(2.27)	(0.08)	(7.72)	(7.59)
Base Outcome	0.35***	0.31***	0.32***	0.27***	0.29***	0.20***	0.31***	0.38***	0.40***
	(0.05)	(0.05)	(0.06)	(0.05)	(0.07)	(0.04)	(0.06)	(0.04)	(0.04)
Education (mun)	1.27	1.32	1.55	34.23	28.25	36.39**	1.29	−2.99	−13.59
	(0.85)	(0.83)	(0.91)	(27.23)	(21.65)	(12.75)	(0.82)	(35.02)	(34.47)
Population (mun)	−0.00	−0.00	−0.00*	−0.00*	−0.00	−0.00**	−0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Household Size (mun)	1.52***	1.68***	2.34***	46.71***	43.78***	32.92***	1.67***	−10.27	−10.04
	(0.50)	(0.49)	(0.55)	(15.39)	(11.53)	(7.81)	(0.49)	(16.74)	(15.90)
Percent Catholic (mun)	3.11	5.30	5.11	−27.34	262.27**	117.32	5.46	−248.03	−136.22
	(5.39)	(5.11)	(5.52)	(187.11)	(123.18)	(100.84)	(5.14)	(265.16)	(245.20)
Crime Rate (mun)	−0.19	−0.22	−0.19	−7.76	−7.55	−3.66	−0.22	7.68	6.79
	(0.28)	(0.29)	(0.34)	(8.52)	(6.20)	(5.60)	(0.29)	(8.24)	(6.78)
Rank (officer)	0.23**	0.09	−0.01	3.37	0.66	1.80	0.07	3.42	−3.15
	(0.09)	(0.10)	(0.12)	(2.41)	(3.09)	(2.49)	(0.10)	(2.51)	(2.59)
Education (officer)		−0.05	−0.02	−2.99	−1.51	−0.23	−0.04	1.96	
		(0.07)	(0.08)	(2.82)	(2.17)	(1.02)	(0.08)	(1.35)	
Catholic (officer)		0.02	0.13	1.32	−0.78	−0.75	−0.02	0.05	
		(0.15)	(0.18)	(3.45)	(2.56)	(3.44)	(0.13)	(5.55)	
Age (officer)		0.01	0.02**	0.15	0.24	0.25	0.01	−0.49*	
		(0.01)	(0.01)	(0.17)	(0.20)	(0.18)	(0.01)	(0.25)	
Gender (officer)		−0.03	−0.02	−1.97	−1.15	0.62	−0.03	−0.24	
		(0.07)	(0.07)	(2.40)	(1.46)	(1.71)	(0.07)	(2.78)	
Years Served (officer)		0.01	0.01	0.48*	0.26	0.19	0.01	−0.47	
		(0.01)	(0.01)	(0.23)	(0.26)	(0.18)	(0.01)	(0.33)	
Commute (officer)		0.00	−0.01	0.05	−0.44	0.05	0.00	1.42	
		(0.02)	(0.03)	(0.75)	(0.43)	(0.37)	(0.02)	(0.99)	
PSM (officer)		0.23***	0.19*	4.41	3.60**	9.27***	0.21**	−6.70**	
		(0.07)	(0.10)	(2.74)	(1.53)	(2.31)	(0.08)	(2.34)	
Controls Interacted	No	No	Yes	No	No	No	No	No	No
R^2	0.20	0.22	0.24	0.17	0.19	0.17	0.22	0.20	0.17
Observations	612	603	603	625	617	618	623	625	636
Clusters	18	18	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline.
Models do not include station fixed effects because NPA Presence is measured at the municipality (station) level.
Standard errors are clustered at the station-level.

Table A.11: Balance of Municipality Characteristics by Officer Embeddedness

Outcome Variable	**Estimate**	**p-value**
Municipality Education	0.031	0.44
Municipality Population	-332.162	0.97
Municipality Home Quality	-0.001	0.56
Municipality Household Size	-0.004	0.87
Municipality Percent Catholic	0.000	0.97
Municipality Crime Rate	-0.075	0.14
Municipality NPA Presence	0.078	0.39

Each row reports the estimates and p-values from regressions that predict municipal characteristics based on officers' embeddedness, which is measured with the dummy variable for whether they live in Sorsogon.

A.4.2 Alternative Operationalization

In this section, we probe whether the results are robust to operationalizing the main concepts in alternative ways. Table A.12 replicates the main average treatment effects (Table A.5) but uses alternative versions of the main attitude indices. In this alternative version, we imputed missing responses for individual items in the indices, using predictions based on non-missing responses to the other questions in the same index. We describe this method in the PAP and SI Section A.2.2 above. Where relevant, results with these alternative indices are also included in heterogeneous effects tables in the Main Results (SI Section A.3).

Table A.12: Treatment Effects - Alternate Indices

	Attitude Idx (Alternative)		Trust (Alternative)		Empathy (Alternative)		Accountability (Alternative)		Corruption (Alternative)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treated	0.01	0.01	−0.04	−0.02	−0.07	−0.08	0.03	0.01	0.07	0.06
	(0.06)	(0.05)	(0.05)	(0.05)	(0.10)	(0.09)	(0.08)	(0.07)	(0.09)	(0.09)
Base Outcome	0.24***		0.34***		0.28***		0.44***		0.27***	
	(0.05)		(0.05)		(0.04)		(0.03)		(0.06)	
Model	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross
R^2	0.08	0.03	0.21	0.13	0.10	0.04	0.16	0.04	0.10	0.05
Observations	679	683	678	684	689	691	582	615	659	671
Clusters	18	18	18	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Panel models include a dummy variable for missing baseline.
All models include station (municipality) and rank fixed effects for the blocking variables.
Standard errors are clustered at the station-level.

Table A.13 displays results on alternative versions of the "Shared Citizen Concerns" outcome. In the main regressions presented in the article (Table A.5), we measure the outcome by counting the number of matches between officers' selected issues and the four most commonly mentioned issues on the same question posed to community leaders (LGU). These results are reproduced in Models 1-2. In Models 3-4, we construct the outcome by counting the number of matches between officers' selected issues and the *three* most commonly mentioned issues on the same question posed to community leaders (LGU). The results remain significant at the .05 level when using this construction.

In Table A.13, Models 5-6, we calculate the matches with community leaders' (LGU) responses, but instead use a version of the question to officers phrased: "Now, what do you think CITIZENS in the municipality where you are assigned consider to be the three most important public safety concerns in their community." We asked the question in this way to try to capture differences between officers' firsthand beliefs and their perceptions of what citizens thought was important. However, because we asked about officers' firsthand beliefs before asking about what they thought citizens believe, officers' answers to the second question may have been colored by their answers to the first question. The correlation between responses to the two versions of the question is $\rho = .83$, leaving us unable to clearly distinguish between these subtly different outcomes. The coefficients remain substantively similar to those in the main models and, in the cross-sectional model, remain significant at the .1 level. In Models 7-8, we instead calculate the matches of officers' responses with the aggregate responses of ordinary citizens (rather than the LGU leaders) in their assigned municipality. The coefficients remain substantively similar in these models, though they are barely insignificant in this case (p=.11 in the cross-sectional model). The correlation between the officer match with 1) LGU and 2) ordinary citizen responses was $\rho = .71$. Once again, this left us unable to clearly distinguish between these subtly different outcomes.

Table A.13: Treatment Effects: Alternative Knowledge Measures

	Top 4 LGU/own		Top 3 LGU/own		Top 4 LGU/their		Top 4 CIV/own	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	0.21*	0.25**	0.16**	0.19**	0.15	0.18*	0.14	0.16
	(0.10)	(0.10)	(0.07)	(0.07)	(0.10)	(0.10)	(0.10)	(0.10)
Base Outcome	0.27***		0.21***		0.25***		0.15***	
	(0.03)		(0.05)		(0.04)		(0.04)	
Model	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross
R^2	0.13	0.07	0.12	0.09	0.11	0.07	0.08	0.06
Observations	627	632	627	632	627	632	627	632
Clusters	18	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Panel models include a dummy variable for missing baseline.
All models include station (municipality) and rank fixed effects for the blocking variables
Models labeled with "Top 4" indicate that the outcome was measured as the match of officer responses with the top 4 issues mentioned by LGU/civilian respondents in the same municipality, while models labeled "Top 3" indicate that we matched with only the top 3 issues mentioned. Models labeled "LGU" indicate that we matched officer responses with the aggregate responses of the LGU community leaders, while models labeled "CIV" indicate that we matched with the aggregate responses of ordinary civilians. Models labeled "own" indicate that we matched using the issues that officers themselves thought were important, while models labeled "their" indicate that we matched using the issues that officers thought that LGU/civilian respondents thought were important.

In Table A.14, we use an alternative measure of the embeddedness moderator. Instead of an indicator for whether the officer lived in Sorsogon province, we use an indicator for whether the officer lived in the same *municipality* where they were assigned (see Table A.14). For reference, 87% of the officers in our sample lived in Sorsogon Province, but only 53% of officer lived in their assigned municipality. The same patterns we picked up in the original test also show up using this alternative measure of embeddedness. The heterogeneous effect on the main attitude index is insignificant in this specification, but we find significant heterogeneous effects on *Trust* and *Empathy*, the same sub-indices that were driving the effects using the original operationalization. Moreover, the results are significant when using the alternate version of the attitude index that imputed missing values for individual survey items.

Table A.14: Heterogeneous Treatment Effects - Embeddedness (Lives In Municipality)

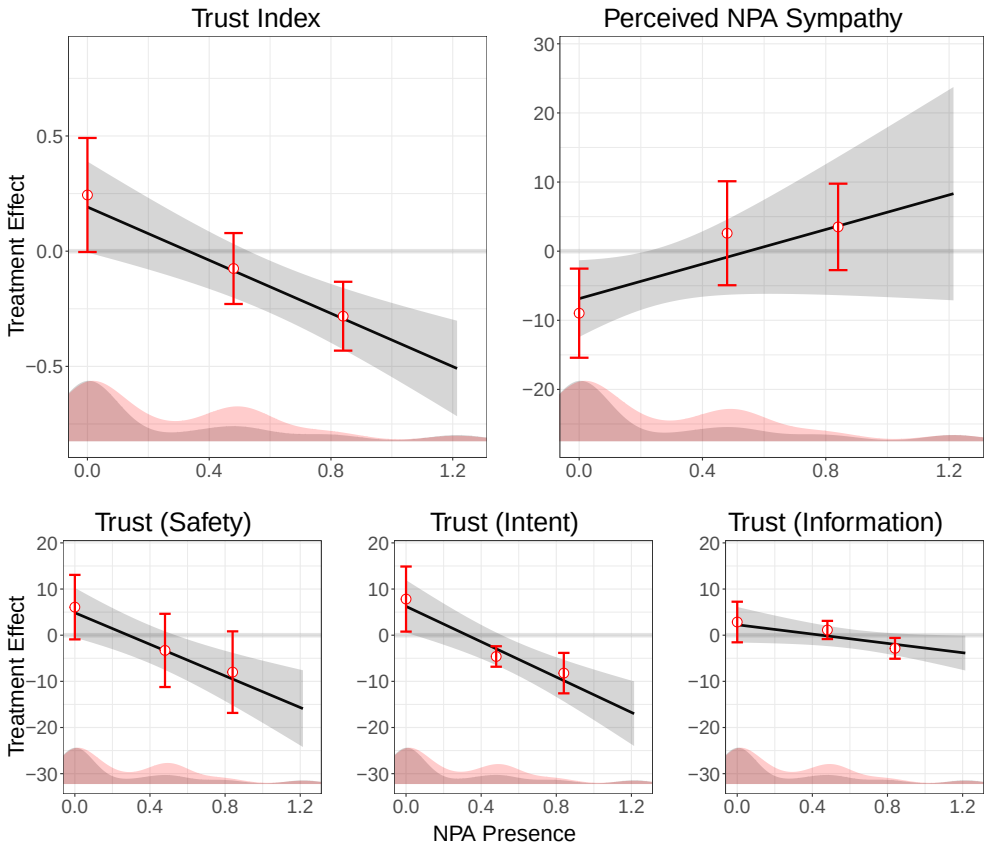
	Share Concerns	Attitude Index	Trust	Empathy	Account	Corrupt	Attitude Idx (alt)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.34***	0.10	0.05	0.03	0.03	0.05	0.09
	(0.11)	(0.09)	(0.07)	(0.12)	(0.15)	(0.13)	(0.09)
Lives in Municipality	0.11**	0.05	0.15**	0.08	0.00	0.04	0.05
	(0.04)	(0.06)	(0.07)	(0.05)	(0.09)	(0.07)	(0.04)
Treated*InMunicipality	−0.19	−0.08	−0.19*	−0.27*	−0.07	0.11	−0.14*
	(0.11)	(0.10)	(0.10)	(0.13)	(0.14)	(0.18)	(0.08)
Base Outcome	0.26***	0.26***	0.29***	0.22***	0.42***	0.21***	0.23***
	(0.03)	(0.07)	(0.06)	(0.04)	(0.05)	(0.06)	(0.05)
Rank (officer)	−0.17	0.01	0.06	0.09	0.01	−0.11	0.03
	(0.11)	(0.09)	(0.08)	(0.08)	(0.12)	(0.09)	(0.06)
Education (officer)	−0.00	−0.01	−0.12*	−0.07	−0.03	0.16*	0.02
	(0.08)	(0.06)	(0.06)	(0.06)	(0.07)	(0.08)	(0.03)
Catholic (officer)	0.12	0.06	−0.03	0.20*	−0.06	0.06	0.10
	(0.09)	(0.08)	(0.13)	(0.11)	(0.09)	(0.07)	(0.07)
Age (officer)	0.01**	−0.01	0.01	−0.00	−0.01*	0.00	−0.01
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Gender (officer)	0.06	−0.04	0.01	−0.07	0.04	0.01	0.01
	(0.10)	(0.05)	(0.07)	(0.06)	(0.06)	(0.14)	(0.04)
Years Served (officer)	−0.02***	−0.00	0.01	0.00	−0.00	−0.01*	−0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Commute (officer)	0.00	−0.02	0.01	0.01	−0.02	0.02	−0.00
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.04)	(0.02)
PSM (officer)	0.06	0.01	0.24***	0.34***	0.06	−0.39***	0.10
	(0.11)	(0.10)	(0.07)	(0.09)	(0.08)	(0.10)	(0.09)
R^2	0.15	0.11	0.23	0.13	0.17	0.13	0.09
Observations	616	455	645	641	531	579	667
Clusters	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. Models include station and rank fixed effects. Standard errors are clustered at the station-level.

A.4.3 Marginal Effects with Binning Estimator

Figure A.4 replicates Figure 4 in the article using the binning estimator described in Hainmueller, Mummolo, and Xu (2019), implemented with the *Interflex* package in R. The binning estimator helps establish whether there is substantial non-linearity in the interaction effects, as well as whether there is sufficient data across the range of the moderator to avoid excessive extrapolation. We rely on this estimator to provide additional support for two main types of claims related to the interaction terms. First, we use the binning estimator to identify whether the intervention had a significant effect on officer outcomes at specific levels of the moderating variable. For this purpose, we look at whether the conditional average treatment effect (CATE) of the intervention is significantly different from zero at "high," "medium," and "low" levels of the moderator, which are defined as the median values of each tercile of the running variable (the default in the *Interflex* package). These CATEs are indicated by the red point estimates and confidence intervals. Second, we use the binning estimator to verify whether the intervention had significantly *different* effects on officer outcomes at high vs. low levels of the moderator. For this purpose, the relevant comparison is whether the confidence intervals of the marginal effects at the "high" level of the moderator overlap with the point estimate of the marginal effects at the "low" level of the moderator (and vice versa).

Figure A.4: Heterogeneous Effect of Community Policing on Officer Trust, by NPA Presence



The displayed marginal effects are produced by a model interacting the baseline crime victimization rate in an officer's municipality with the treatment status, using the panel model specification described in the manuscript. The distribution of the running variable is represented at the bottom of the chart with density plots for treatment officers (red) and control officers (gray). The bin locations were specified using the default settings in the *Interflex* R package, which uses the median value within each tercile of the running variable.

Consistent with the interpretation in the paper, the CATE on officers' trust levels is negative and significant in areas with a high level of NPA presence (Figure A.4). Moreover, the CATE in high-NPA areas is significantly different from the CATE in low-NPA areas. There does not appear to be substantial non-linearity in the heterogeneous effects: the CATE estimates produced by the binning estimator are all close to the predicted marginal effects produced by the linear interaction model. A similar pattern presents for the sub-index items of the trust index. Consistent with the proposed mechanism, officers assigned to the intervention in low-NPA areas perceived that citizens were less sympathetic to the NPA, relative to their control group counterparts. Again, in this model, the CATE among officers assigned to low-NPA areas was negative and significantly different from zero, as well as significantly different from the CATE among officers assigned to high-NPA areas.

A.4.4 Block Bootstrapped Standard Errors

Tables A.15, A.16, A.17, and A.18 replicate the four main results tables in SI Section A.3 using block-bootstrapped standard errors, to account for the relatively small number of clusters (in this case, stations). The robustness tests suggest that, if anything, the standard errors in our pre-registered specification were biased upwards due to the small number of clusters. Using the block-bootstrapped standard errors, the p-values on the coefficients of interests tend to be slightly smaller compared to the original models.

Table A.15: Main Treatment Effects with Block Bootstrapped Standard Errors

	Share Concerns		Attitude Idx		Trust		Empathy		Accountability		Corruption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated	0.21*	0.25**	0.06	−0.01	−0.06	−0.02	−0.12	−0.08	−0.00	0.00	0.12*	0.07
	(0.11)	(0.10)	(0.05)	(0.06)	(0.06)	(0.05)	(0.11)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Base Outcome	0.27***		0.26***		0.34***		0.28***		0.43***		0.26***	
	(0.04)		(0.07)		(0.06)		(0.04)		(0.04)		(0.06)	
Model	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross
R^2	0.13	0.07	0.10	0.04	0.21	0.13	0.10	0.04	0.16	0.05	0.10	0.05
Observations	627	632	460	534	656	671	652	672	536	587	589	621
Clusters	18	18	18	18	18	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
The results in this table correspond to those shown in Table A.5. Panel models include a dummy variable for missing baseline.
All models include station (municipality) and rank fixed effects for the blocking variables.
Standard errors are clustered and block-bootstrapped at the station-level.

Table A.16: Heterogeneous Treatment Effects - Embeddedness with Block Bootstrapped Standard Errors

	Share Concerns	Attitude Index	Trust	Empathy	Account	Corrupt
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.28*	0.31**	0.29	0.28*	0.08	0.12
	(0.17)	(0.14)	(0.25)	(0.17)	(0.21)	(0.36)
Lives in Province	0.19*	0.17***	0.20***	0.23**	0.11	−0.05
	(0.10)	(0.02)	(0.06)	(0.11)	(0.08)	(0.09)
Treat*InProvince	−0.09	−0.28*	−0.38	−0.43**	−0.09	−0.00
	(0.20)	(0.14)	(0.26)	(0.17)	(0.21)	(0.34)
Base Outcome	0.27***	0.26***	0.34***	0.27***	0.43***	0.26***
	(0.04)	(0.07)	(0.04)	(0.04)	(0.03)	(0.06)
R^2	0.13	0.10	0.21	0.11	0.16	0.10
Observations	627	460	656	652	536	589
Clusters	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
The results in this table correspond to those shown in Table A.6. Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. All models also include station (municipality) and rank fixed effects for the blocking variables. Standard errors are clustered and block-bootstrapped at the station-level.

Table A.17: Heterogeneous Treatment Effects - Rebel Presence (Trust Outcomes) with Block Bootstrapped Standard Errors

	Trust Index	Trust (Safe)	Trust (Intent)	Trust (Info)	Perceived NPA Sympathy
	(1)	(2)	(3)	(4)	(5)
Treated	0.17	4.86	6.24***	2.26	−6.86**
	(0.10)	(3.01)	(2.37)	(1.37)	(2.71)
NPA Presence 2015	0.23**	2.95	8.72***	4.32**	−5.11**
	(0.10)	(2.90)	(2.77)	(1.94)	(2.50)
Treated*NPA	−0.57***	−17.10***	−19.14***	−5.04***	12.50
	(0.13)	(5.59)	(3.95)	(1.94)	(7.65)
Base Outcome	0.40***	0.33***	0.35***	0.28***	0.41***
	(0.04)	(0.06)	(0.07)	(0.03)	(0.03)
R^2	0.17	0.12	0.13	0.11	0.16
Observations	612	635	628	628	636
Clusters	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$

The results in this table correspond to those shown in Table A.7. Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. All models also include rank fixed effects for the blocking variables, though they do not include station fixed effects because NPA Presence is measured at the municipality (station) level. Standard errors are clustered and block-bootstrapped at the station-level.

Table A.18: Heterogeneous Treatment Effects - Rebel Presence with Block Bootstrapped Standard Errors

	Share Concerns	Attitude Index	Trust	Empathy	Account	Corrupt	Trust Idx (alt)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.15*	0.01	0.17*	−0.03	−0.08	0.08	0.19**
	(0.08)	(0.04)	(0.09)	(0.10)	(0.06)	(0.08)	(0.08)
NPA Presence 2015	0.09	−0.20**	0.23**	0.09	−0.21***	−0.21***	0.25**
	(0.09)	(0.08)	(0.10)	(0.09)	(0.08)	(0.06)	(0.10)
Treated*NPA	0.18	0.09	−0.57***	−0.16	0.26***	−0.06	−0.58***
	(0.20)	(0.17)	(0.11)	(0.23)	(0.10)	(0.10)	(0.15)
Base Outcome	0.28***	0.26***	0.40***	0.27***	0.46***	0.25***	0.40***
	(0.03)	(0.07)	(0.06)	(0.03)	(0.03)	(0.06)	(0.05)
R^2	0.09	0.07	0.17	0.08	0.14	0.08	0.17
Observations	627	431	612	610	500	554	634
Clusters	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$

The results in this table correspond to those shown in Table A.8. Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. All models also include rank fixed effects for the blocking variables, though they do not include station fixed effects because NPA Presence is measured at the municipality (station) level. Standard errors are clustered and block-bootstrapped at the station-level.

A.4.5 Weighted Regression

Tables A.19, A.20, A.21, and A.22 replicate the four main results tables in SI Section A.3 using weighted regression, to account for officers' different probability of being assigned to treatment. Weights were calculated based on officers' station and rank, as well as the number of treatment barangays in the municipality. For example, the Prieto Diaz station had eleven eligible junior officers, nine eligible senior officers, and four barangays assigned to the One Sorsogon intervention. Because each of the four barangays was assigned one junior officer and one senior officer, this translated to a 36% probability of assignment for junior officers and a 44% probability of assignment for senior officers employed by that station. These weights were incorporated into the models using inverse-probability weighting. The results are similar to the main specifications. The ATE on *Shared Concerns* remains marginally significant in the cross-sectional model, though barely misses significance in the panel model (Table A.19). The heterogeneous effects of embeddedness appear stronger in the weighted models, with the base terms and interaction terms all being significant at conventional levels for the *Attitude Index*, *Trust*, and *Empathy* outcomes (Table A.20). The heterogeneous effects of NPA presence on the *Trust* index and its constituent survey items also appear stronger in the weighted models, except for *Trust (Info)*. The heterogeneous effect of NPA presence on the *Perceived NPA Sympathy* outcome is statistically significant at conventional levels using this specification, whereas they were barely insignificant in the main specification (Table A.21). Finally, we show that the results remain significant for the main outcome of interest in the heterogeneity analyses when incorporating the probability of being assigned as a replacement for one of the original officers assigned to treatment, an additional 19% likelihood. The results using this alternative weighting scheme are displayed in Model 7 of Table A.20 and Model 6 of Table A.21.

Table A.19: Main Treatment Effects Using Weighted Regression

	Share Concerns		Attitude Idx		Trust		Empathy		Accountability		Corruption	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated	0.18	0.21*	0.07	0.02	−0.10	−0.06	−0.08	−0.05	−0.04	−0.03	0.16	0.10
	(0.12)	(0.12)	(0.05)	(0.05)	(0.08)	(0.08)	(0.10)	(0.09)	(0.13)	(0.12)	(0.14)	(0.15)
Base Outcome	0.27***		0.26***		0.34***		0.28***		0.43***		0.26***	
	(0.03)		(0.06)		(0.05)		(0.04)		(0.04)		(0.06)	
Model	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross	Panel	Cross
R^2	0.10	0.07	0.08	0.03	0.14	0.11	0.06	0.02	0.09	0.05	0.07	0.04
Observations	627	632	460	534	656	671	652	672	536	587	589	621
Clusters	18	18	18	18	18	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Panel models include a dummy variable for missing baseline.
All models include station (municipality) and rank fixed effects for the blocking variables.
Standard errors are clustered at the station-level.

Table A.20: Heterogeneous Treatment Effects - Embeddedness Using Weighted Regression

	Share Concerns	Attitude Index	Trust	Empathy	Account	Corrupt	Attitude Index (Alt)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.09	0.16***	0.43*	0.50***	0.05	−0.23	0.17***
	(0.19)	(0.05)	(0.21)	(0.12)	(0.15)	(0.15)	(0.05)
Lives in Province	0.13**	0.14***	0.25***	0.36***	0.07**	−0.20***	0.14***
	(0.06)	(0.01)	(0.03)	(0.09)	(0.03)	(0.07)	(0.01)
Treat*InProvince	0.09	−0.10**	−0.59**	−0.66***	−0.10	0.43*	−0.10**
	(0.18)	(0.04)	(0.24)	(0.12)	(0.23)	(0.21)	(0.04)
Base Outcome	0.27***	0.26***	0.34***	0.27***	0.43***	0.26***	0.26***
	(0.03)	(0.06)	(0.05)	(0.04)	(0.04)	(0.06)	(0.06)
R^2	0.10	0.08	0.16	0.09	0.09	0.08	0.08
Observations	627	460	656	652	536	589	460
Clusters	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. All models also include station (municipality) and rank fixed effects for the blocking variables. Standard errors are clustered at the station-level.

Table A.21: Heterogeneous Treatment Effects - Rebel Presence (Trust Outcomes) Using Weighted Regression

	Trust Index	Trust (Safe)	Trust (Intent)	Trust (Info)	Perceived NPA Sympathy	Trust Index (Alt)
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.16*	5.50**	6.28***	1.83	−7.19***	0.15
	(0.09)	(2.48)	(1.93)	(1.95)	(2.07)	(0.09)
NPA Presence 2015	0.27***	7.20**	8.98***	4.02***	−5.39***	0.26***
	(0.06)	(2.49)	(1.93)	(1.06)	(1.23)	(0.07)
Treated*NPA	−0.68***	−24.61***	−22.38***	−5.14	16.13**	−0.67***
	(0.15)	(4.71)	(5.20)	(3.00)	(7.53)	(0.15)
Base Outcome	0.40***	0.33***	0.35***	0.28***	0.41***	0.40***
	(0.04)	(0.05)	(0.06)	(0.03)	(0.04)	(0.04)
R^2	0.14	0.11	0.09	0.09	0.13	0.14
Observations	612	635	628	628	636	612
Clusters	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. All models also include rank fixed effects for the blocking variables, though they do not include station fixed effects because NPA Presence is measured at the municipality (station) level. Standard errors are clustered at the station-level.

Table A.22: Heterogeneous Treatment Effects - Rebel Presence Using Weighted Regression

	Share Concerns	Attitude Index	Trust	Empathy	Account	Corrupt	Trust Idx (alt)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.16	0.04	0.16*	−0.02	−0.13	0.16	0.19**
	(0.13)	(0.05)	(0.09)	(0.09)	(0.13)	(0.15)	(0.08)
NPA Presence 2015	0.17	−0.28***	0.27***	0.07*	−0.35***	−0.17***	0.30***
	(0.12)	(0.05)	(0.06)	(0.04)	(0.07)	(0.03)	(0.06)
Treated*NPA	0.05	0.09	−0.68***	−0.16	0.36*	−0.19	−0.72***
	(0.32)	(0.15)	(0.15)	(0.16)	(0.19)	(0.21)	(0.15)
Base Outcome	0.28***	0.26***	0.40***	0.27***	0.46***	0.25***	0.40***
	(0.03)	(0.07)	(0.04)	(0.04)	(0.04)	(0.06)	(0.05)
R^2	0.07	0.07	0.14	0.06	0.08	0.07	0.13
Observations	627	431	612	610	500	554	634
Clusters	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline. All models also include rank fixed effects for the blocking variables, though they do not include station fixed effects because NPA Presence is measured at the municipality (station) level. Standard errors are clustered at the station-level.

A.4.6 Spillovers

In this section, we consider the implications of spillovers for the interpretation of our results, and estimate the presence of spillovers shaped by officers' family ties and beat patrol partner assignments. The most plausible type of spillovers in the context of our study are "positive" spillovers that would bias the treatment effects towards null, driven by two mechanisms: 1) officers sharing experiences and opinions with each other and 2) increased commitment to community policing by police leadership. First, treatment officers may relay their experiences with citizens to control officers, leading control officers to learn about citizens' public safety concerns. Interactions between officers may also create a pathway for the diffusion of trust or empathy towards citizens. Second, it is possible that the "top down" mechanism we theorized – the effect of police leadership's commitment to improve citizen interactions – may have also affected the attitudes of control officers.

In an ideal world, we would have designed the experiment to include several "pure control" stations, where none of the officers engaged in community policing. This design would have allowed us to compare attitudes of control officers in control stations to control officers in treatment stations. However, we determined that this was not a feasible approach given power considerations associated with the number of police stations in Sorsogon. In the absence of a design-based approach to spillovers, we conducted several individual-level and station-level analyses to investigate the likelihood that significant spillovers took place.

We attempt to capture control officers "exposure" to treatment by measuring two aspects of officers' social networks that shape which colleagues officers are most likely to interact with: 1) family ties between officers and 2) officers' beat patrol assignments. We coded family ties using matches between officers' surnames within each station. In the Philippines (especially in non-urban areas), kinship networks have been shown to be a highly salient pathway for information transmission and opinion formation (Cruz, Labonne, and Querubin 2017; Haim 2022). Moreover, the nature of surnames in the Philippines provides an opportunity to measure family networks. In the mid-19th century, the Spanish colonial regime assigned unique Hispanic surnames to each family within each province, with the goal of increasing legibility of the population. Surnames are passed down from both the mother's and father's side of the family using Spanish naming conventions. The legacy of this policy is that two

individuals within a locality that share at least one surname are very likely to be related to each other, even nearly two centuries later (Cruz, Labonne, and Querubin 2017).

The second way we measured officers' ties was using data on daily beat patrol schedules. There is no standardized format through which stations record their beat schedules, but eight of the 17 stations in our sample recorded data on daily beat assignments. In these stations, the average size of beat teams was just under three officers, ranging from two to five officers per team. The remaining stations either did not record beat patrol schedules or did so using a level of aggregation that did not allow us to decipher which officers went on beat patrols with each other. Beat data were provided to us for the first two months of the POP intervention, as well as the year prior to the intervention. We denoted a tie between all officers who went on at least one patrol together during the treatment period. Most stations tended to keep the same beat partners together until an officer was transferred to another station, with the average officer-pair patrolling together for 30 days within the first two months of the intervention. Most of the officers in our sample had the same beat partners during the treatment period that they had a year prior to the intervention.

In our context, approximately 20% of control officers shared a surname with at least one officer assigned to treatment in their station and 30% of control officers (in the stations with beat data) were assigned to patrols with treatment officers during the period of the intervention. Although there certainly exist other relevant ties between officers that we were unable to measure, family relationships and beat partnerships are a highly salient pathway of information transmission between officers, and if spillovers existed, we would expect to see that they are structured to some extent by these ties.

Table A.23: Spillover Analysis - Individual-Level

	Shared Concerns			Attitude Index			Trust	Empathy	Account	Corrupt
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Exposure to treated (Any)	−0.07			−0.05			−0.05	−0.06	0.11	−0.26**
	(0.10)			(0.09)			(0.09)	(0.14)	(0.08)	(0.11)
Exposure to treated (Family)		0.04			−0.09					
		(0.15)			(0.09)					
Exposure to treated (Beat)			−0.04			0.03				
			(0.18)			(0.09)				
R^2	0.09	0.09	0.05	0.08	0.08	0.07	0.21	0.06	0.06	0.05
Observations	452	452	140	343	343	109	431	433	377	400
Clusters	17	17	8	17	17	8	17	17	17	17

* $p < .10$, ** $p < .05$, *** $p < .01$
All models include station (municipality) and rank fixed effects for the blocking variables
Regressions are subset to officers in the control group. The independent variables indicate whether the officer had a family member in the treatm
(Family), a beat patrol partner in the treatment group (Beat), or either a family member or beat patrol partner in the treatment group (Any).

We conducted the first set analyses at the individual-level. For each officer in our sample, we create several dummy variables for their exposure to treatment. We code 1) whether officers had at least one family member assigned to treatment – *Exposure (Family)*, 2) whether officers had at least one beat patrol partner assigned to treatment – *Exposure (Beat)*, and 3) whether officers had at least one family member or at least one beat partner assigned to treatment – *Exposure (Any)*. Among the subset of officers in the control group, we run OLS regressions to see whether any of these exposure variables predict officers' shared concerns with citizens or their attitudes towards citizens. We find that none of the *Shared Concerns*, *Officer Attitude Index*, *Trust*, *Empathy*, or *Accountability* outcomes are significantly predicted by exposure to treatment using these measures (Table A.23). The only sub-index significantly predicted by control officers' exposure to treatment is *Corruption* – control officers with exposure

to treatment officers were less likely to report that corrupt activities were acceptable, relative to control officers without exposure. We thus advise caution when interpreting a substantive null effect of the intervention on this outcome.

In additional to the individual-level spillover analysis, we explore spillovers using heterogeneous effects at the station-level, coding which stations may be the most conducive to generating spillovers. If spillovers are present, we expect that knowledge and attitudes may be more likely to transfer between officers 1) in smaller stations (where a larger proportion of officers were treated) and 2) in stations with denser overall family networks. In these stations, there likely existed more interactions between treatment and control officers, which would result in smaller treatment effects if spillovers occurred. We measure station-level family network density using a standard network density measure: the number of *observed* family ties between officers divided by the *possible* number of ties in a station, given the total number of officers.

Table A.24: Spillover Analysis - Station-Level

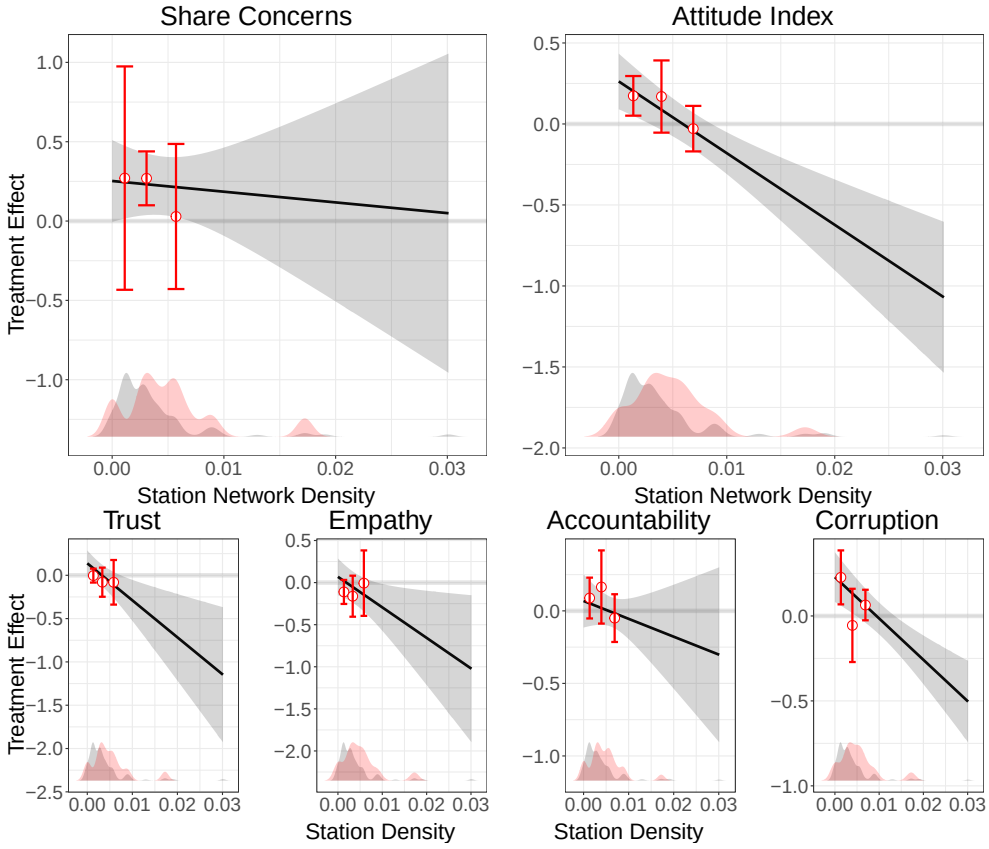
	Share Concerns	Attitude Index	Trust	Empathy	Account	Corrupt	Share Concerns	Attitude Index	Trust	Empathy	Account	Corrupt
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated	0.070	0.096	−0.200*	−0.072	0.097	−0.002	0.253*	0.263***	0.138*	0.068	0.068	0.229***
	(0.151)	(0.082)	(0.107)	(0.120)	(0.075)	(0.076)	(0.131)	(0.087)	(0.076)	(0.111)	(0.093)	(0.075)
StationSize	−0.001	−0.001	−0.003*	−0.001	−0.002**	0.001*	−0.000	−0.001	−0.003*	−0.002	−0.002**	0.001**
	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Treated*Size	0.003*	−0.001	0.002*	−0.001	−0.001	0.002**						
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)						
StationDensity							21.639*	14.069**	9.167	5.847	13.046	7.850*
							(11.043)	(5.691)	(14.515)	(8.524)	(8.698)	(4.130)
Treated*Density							−6.750	−44.266***	−42.704***	−36.265**	−12.317	−24.337***
							(19.770)	(9.839)	(14.399)	(16.370)	(12.629)	(5.625)
Base Outcome	0.283***	0.256***	0.353***	0.258***	0.443***	0.254***	0.278***	0.257***	0.357***	0.265***	0.437***	0.257***
	(0.033)	(0.063)	(0.051)	(0.039)	(0.036)	(0.059)	(0.035)	(0.065)	(0.051)	(0.039)	(0.037)	(0.057)
Education	1.044	0.134	1.068	0.709	−0.077	−1.289**	0.602	0.005	1.154	0.757	−0.411	−1.268*
	(1.172)	(0.721)	(0.946)	(0.952)	(0.978)	(0.486)	(1.039)	(0.599)	(1.022)	(0.891)	(1.160)	(0.613)
Population	−0.000	0.000	−0.000	−0.000	0.000	0.000**	−0.000	0.000	−0.000	−0.000	0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
HH Size	−0.436	−0.144	1.479**	0.191	−0.220	−0.540	−0.919*	−0.158	1.576**	0.317	−0.458	−0.504
	(0.492)	(0.371)	(0.518)	(0.435)	(0.319)	(0.338)	(0.509)	(0.288)	(0.654)	(0.428)	(0.476)	(0.323)
Catholic	−5.479	1.173	5.339	−4.083	−0.042	9.059**	−5.268	0.793	5.327	−4.014	0.050	8.878**
	(7.819)	(4.200)	(5.718)	(5.658)	(7.255)	(3.877)	(6.979)	(3.726)	(4.931)	(4.682)	(7.475)	(4.049)
Crime Rate	0.219	0.344*	−0.160	0.234	0.145	0.337*	0.080	0.443**	−0.059	0.339	0.105	0.378*
	(0.295)	(0.175)	(0.298)	(0.250)	(0.182)	(0.177)	(0.303)	(0.166)	(0.254)	(0.201)	(0.197)	(0.196)
R^2	0.113	0.084	0.206	0.089	0.155	0.099	0.114	0.105	0.212	0.097	0.156	0.100
Observations	627	431	612	610	500	554	627	431	612	610	500	554
Clusters	18	18	18	18	18	18	18	18	18	18	18	18

* $p < .10$, ** $p < .05$, *** $p < .01$
Models in this table use the panel specification denoted in equation 1 and include a dummy variable for missing baseline.
Models include rank fixed effects, but not station fixed effects because the moderating variables are measured at the station level.
Standard errors are clustered at the station-level.

Table A.24 displays results of regressions where we interact officers' treatment status with the size and density of their station, controlling for the other station-level characteristics. Looking at these results, it initially appears as if there were significant positive spillovers – treatment effects on a number of the outcomes were significantly more positive in larger and less densely connected stations, where spillovers are of less concern. But upon further investigation, these results are driven entirely by outliers in the running variables. This pattern is visible in Figure A.5, which uses the Hainmueller, Mummolo, and Xu (2019) binning estimator for the models that interact treatment

with station density. For every outcome, the models are excessively extrapolating the results to a very small number of officers in stations that exhibit extraordinarily high network density. Within the range where there is a sufficient number of officers to support the substantive interpretation of marginal effects, none of the interaction terms are significant. Although the station size moderator has a bimodal distribution rather than a long tail, the binning estimator for these models also fails to reveal a consistent pattern that would indicate spillovers are at play (Figure A.6).

Figure A.5: Spillover Marginal Effects - Station Level Network Density



In summary, the above analyses mitigate the concern that spillovers explain why we found null effects of community policing on officers' attitudes towards citizens. That said, in the absence of a design-based approach to spillovers, we advise some caution in interpreting the results. Beyond the tests presented in this section, we note that if spillovers took place, the heterogeneous effects presented in the inductive theory section would also be biased towards null. For example, if treatment officers developed worse attitudes towards citizens as a result of community policing in high-insurgency areas and proceeded to kickstart a process of diffusion of these attitudes to other officers in their station, the heterogeneous effects would be biased towards null.

A.5 Qualitative Methodology

To develop our inductive theory and interpret our quantitative results, we drew on a rich trove of qualitative information gathered before, during, and after our intervention. For the most part, we viewed this qualitative information-gathering exercise as an effort to uncover nuances and mechanisms that we may or may not have been aware of previously and which could help make sense of seemingly unexpected or counter-intuitive results. We also

Figure A.6: Spillover Marginal Effects - Station Level Size

designed our qualitative investigation to identify whether the intervention was being implemented as we expected, and if there were deviations, to determine why they occurred and how they impacted our results. The two main types of qualitative evidence collection were author interviews and field staff observations, which we describe below.

A.5.1 Author Interviews

First, the authors conducted semi-structured key informant interviews for over a year leading up to the start of the intervention. Given that the idea for this project came about while we were already in the field working on an unrelated project, we were well-positioned to interview mid- and high-ranking police officers, community leaders, bureaucrats, and ordinary citizens. In most cases the author(s) took hand-written notes during these interviews, though the primary emphasis was on contextualizing the policing a public safety situation in the region to understand local circumstances and design an appropriate intervention, rather than on recording direct quotes to use as evidence. A priority was also placed on using these interviews to ensure that the intervention could be carried out safely and ethically, as we discuss in Section A.6.

Next, the authors had hundreds of informal conversations with citizens, community leaders, government officials, and police officers of all ranks before, during, and after the intervention, which contributed to our understanding of the policing and public safety contexts in the region. Oftentimes what began as an informal chat with a taxi driver, bureaucrat, or police desk sergeant ended up yielding relevant information that helped contextualize our systematic quantitative results, even though we did not intend for the exchange to be an "interview" when it began. Because the results of these conversations are neither systematic nor reproducible, we think of them less as

evidence in their own right and more as important background information that helps us ask the right questions of our quantitative data, better interpret the results, and serve as the basis for inductive theory development.

A.5.2 Field Staff Observations

Individuals from our team of approximately 20 local field researchers were assigned to thoroughly monitor the intervention. All of our field staff came from the study area, and most had been involved in evaluations of previous randomized interventions. A member of our field staff was in attendance at every POP meeting and culminating town hall during the intervention, and traveled to the treatment barangays to observe the implementation of solutions developed by the POP teams. In addition, field officers conducted a "barangay audit" after the fourth POP meeting, in order to gather perspectives on the program from individuals who were not on the POP team. In total, our field staff attended nearly two thousand meetings and events associated with the intervention. We view these activities as embedded research, as the field researchers were present for a large portion of the intervention but did not actively participate in it. Only the local field researchers, and not the authors, were present during intervention activities because we did not want our presence as foreigners to influence the intervention.

Our field researchers – particularly the research manager and senior field officers (supervisors) – also conducted hundreds of informal conversations about the public safety context in Sorsogon similar to those we conducted ourselves and describe above. For example, field researchers often arrived early to POP meetings and struck up informal conversations with barangay officials while waiting for other meeting participants to arrive, or discussed the intervention with curious friends or family members who became aware of the PNP's initiatives. Our more senior field researchers also frequently met with mid-ranking police leadership and the officers involved in the intervention to coordinate intervention logistics, providing numerous opportunities to learn about their reactions to and perspectives on the intervention.

We instructed our field staff to take notes at all the activities they attended, and report back to us on the intervention's implementation, progress, and any potential deviations. We also tasked them to think critically and take notes on the factors they were enhancing or inhibiting the success of the program. As with our own conversations, since the focus was primarily on monitoring the implementation of the intervention and understanding contextual nuances, we did not ask the field researchers to capture direct quotations. The field staff were grouped into three teams, each headed by a senior field officer, and overseen by an experienced field manager. Each week, each team of field officers met with the field manager to discuss their notes and elaborate on any insights they had about the intervention. The authors then had weekly meetings with the field manager to discuss the main points raised by the field staff. In addition, at least one of the authors was present at the field site during most times during the intervention, and we often discussed the intervention directly with the field staff.

A.5.3 Interpretation of Qualitative Evidence

Given the informal nature of most of this qualitative information collection, learning from it was primarily an exercise in triangulating big-picture context from thousands of distinct interactions. Most of these interactions painted complementary, if incomplete, pieces of a unified picture. As qualitative information accumulated to paint a more complete picture of the safety and policing context in Sorsogon, and later our intervention's effects, we used it in the following ways:

1. Understand local threats to safety that could affect research on community policing, both in terms of threats to the intervention itself and threats to data collection, for the purposes of mitigating ethical concerns.
2. Design a locally-appropriate community policing intervention that could be realistically implemented. This task relied heavily on conversations with PNP leadership, who had strong opinions about activities they

expected to be effect and resources they could deploy, but also with community leaders who had a useful perspective on public safety needs that were not being served by the existing policing model.

3. Understand and contextualize the quantitative results of our main pre-registered tests. In particularly, we used qualitative information to confirm that the intervention was implemented as expected and null results were not caused by non-compliance. We also gained insights about the challenges that inhibited certain mechanisms from taking hold, such as factors that inhibited police leadership's ability to improve accountability.
4. Identify mediating factors which could have obscured effects in our primary tests, with the goal of motivating inductive theory building and revealing useful additional quantitative tests. Insights from our field officers revealed that officers' home locality, as well as the presence of nearby NPA activity, were two particularly important potential mediators. We also gathered information on a number of other, less developed, mediators that might be shaping program effectiveness (such as the role of mid-level station leadership).

A.6 Ethics and Permissions

All procedures were approved by the University of California San Diego (UCSD) IRB under protocol numbers 170974 and 170415.

Partnering with the Sorsogon PNP on the One Sorsogon community policing initiative required careful consideration of several important ethical issues. In the manuscript, we discuss two concerns that were particularly challenging in the context of our study: the broader Philippine Drug War and local insurgent activity. Here, we further describe our process for assessing these issues and discuss several other ethical issues that were particularly pressing in the context of our study.

An important piece of context is that One Sorsogon was a home-grown initiative of the Provincial Police Chief, who wished to de-emphasize the aggressive tactics of the Drug War and instead focus on non-drug related public safety issues. This does not mean that we, as researchers, absolve ourselves of the risks involved in implementing the program. We went through an extended process of due diligence to assess the risks of the intervention. That said, the homegrown nature of the program does reduce the risk that we as outsiders would introduce harms to the community by introducing a policy that was not locally appropriate. In addition, because the policy would have been implemented with or without our involvement, it was unlikely that we would introduce any additional harms that would not have already been experienced. If anything, the system we put in place to monitor abuses should reduce harms to citizens, in expectation.

Given this context, we saw substantial ethical benefits stemming from the potential for catalyzing broader reform within the PNP, paired with conditions that allowed us to implement the intervention safely. Policing in the Philippines is highly decentralized; regional and provincial police chiefs have a great degree of autonomy over the tactics used by their units. As a result, the Sorsogon Police Chief's community policing initiative was likely to have a large impact on day-to-day policing practices across the province. The chief hoped to overhaul the Sorsogon PNP's approach to focus on community policing before his potential promotion to the regional office so that he could use it as a model for more widespread reform. When he learned of our interest to partner on an evaluation of One Sorsogon, he was adamant that, if the program was successful, a credible external evaluation could help sell this "softer" approach to policing to higher ups in the PNP. Indeed, the model of One Sorsogon helped spur a much larger region-wide community outreach initiative.

Although we saw significant potential benefits to the research, the primary factor in our ethical calculus was whether we could participate in the research without putting subjects or research staff in harm's way. During the project planning period, we held frequent meetings (both with and without the PNP leadership) to discuss how to mitigate risks associated with the intervention. One particularly important consideration was that the level of decentralization in the PNP could backfire. Despite the leadership-level buy-in with community policing practices,

we were concerned that street level officers may not adopt a similar perspective. To guard against the risk that officers may commit abuses as a result of the intervention, we put in place an extensive monitoring system for our research team to observe potential abuses. Multiple staff members from our experienced local research team acted as "community observers" during all town halls conducted as part of the intervention. This allowed us to monitor the vast majority of police-citizen interactions that took place as part of the intervention. A member of the research team was also present at every POP meeting. Finally, we designed the intervention so that LGU officials took the lead in implementing the public safety initiatives decided on during POP in their own communities, with the PNP playing a supporting role. We were prepared to immediately end our involvement in the project if we learned that the intervention created opportunities for individual PNP officers to abuse or intimidate citizens. No instances of abuse were reported.

Several ethical considerations were especially salient in our context. First, there were important ethical challenges that arose from working with the PNP in the midst of President Duterte's War on Drugs. During the period from late 2016 through September 2021, the Drug War resulted in 12,000-30,000 deaths, including at least 6,200 at the hands of the police ("How many people have been killed in Rodrigo Duterte's war on drugs?" *The Economist*, November 2021, https://www.economist.com/graphic-detail/2021/11/22/how-many-people-have-been-killed-in-rodrigo-dutertes-war-on-drugs. We were wary of increasing contact between police officers and citizens due to the risk of police abuse. Our scoping work shed important light on how this risk manifested in Sorsogon. First, the PNP is a very decentralized organization, with provincial chiefs having a great amount of discretion over policing practices. The vast majority of Drug War violence occurred in major urban areas, especially Manila. Before approaching the PNP, we learned from our contacts in the area that the Sorsogon Provincial Police Chief was a well-known reformer who opposed the violent tactics of the Drug War.

A survey of citizens we conducted before approaching the PNP also suggested low rates of police abuse in Sorsogon: 99% of respondents had not heard of any instances of police abuse in their community in the last six months, and 95% of respondents said that the PNP's methods for fighting drugs had not become more violent in their area since Duterte's election. To reduce response bias to these potentially sensitive questions, respondents privately recorded their responses on an iPad that locked after responses were submitted. ACLED's special project on the Philippines Drug War recorded only five drug-related deaths in Sorsogon Province between Duterte's election and the initiation of our project, and none in the nine months preceding the intervention. Still, we were prepared to walk away from the project in response to indications of police abuse. We employed field staff who attended each community policing event, so they could report any instances of coercion or abuse. We received no reports of concerning behavior.

The second main ethical concern stemmed from the threat of NPA presence in Sorsogon, including the risk that the intervention might increase incentives for insurgents to target civilians or the police with violence. To understand the nature of this risk, we engaged in extensive conversations with local experts on the insurgency, including individuals with current and former ties to the NPA in Sorsogon. Among these experts, there was unanimous agreement that the community policing intervention would not increase the risk of violence if implemented in our proposed sample areas with low NPA activity. We relied heavily on our local network in each municipality to identify areas where NPA presence was strong enough to make the intervention unsafe. We also consulted the local AFP command and each municipal PNP station, who wanted to exclude barangays where there was a risk of ambushes. Finally, we required approval from each barangay captain in the already pared-down sample. All captains in our sample consented to the intervention and the study.

A broader concern is whether the goal of improving trust and cooperation with the PNP as an institution was actually a good thing during a time when the institution was the face of abusive human rights violations. We wrestled significantly with this concern, but in the end several factors led us to proceed. First, the majority of citizens in Sorsogon seem to recognize the degree of decentralization in the PNP and the vastly different approaches to policing in major urban areas like Manila (where the Drug War was primarily taking place) and rural provinces

like Sorsogon. In other words, citizens can generally separate between the PNP "institution" at these two levels. Second, we devoted significant attention to mitigating the risk of the Sorsogon PNP capitalizing on increased citizen trust to enable abusive practices over the long term. Based on discussions with key informants, we came to believe that One Sorsogon would empower the community-focused municipal Police Community Relations (PCR) officers to remain influential, even if a changeover in provincial police leadership occurred. This proved correct when the provincial chief was promoted to the regional office early in the POP intervention. Although his successor wanted to implement his own initiatives, the momentum of One Sorsogon convinced him to continue with a community-focused approach. In terms of our own commitment as researchers, we viewed our involvement in One Sorsogon to be the beginning of longer term efforts to help implement "best practices" of policing across the province and the region. Each of us is still involved in ongoing research on policing in the region and at no time did we consider this to be a one-off project.

Finally, the participation of local researchers and stakeholders proved to be a key factor in our ability to accurately assess ethical risks and benefits. As we mentioned in the manuscript and in the preceding paragraphs, we relied on an extensive network of local stakeholders to brainstorm and assess the situations that might arise as a result of the intervention. One of the study authors was born and raised in the region, as was the full-time field research manager. The field research manager has a long-running relationship with the PIs that allows her to freely express ethical or other concerns about things she observed in the field. These researchers' extensive local knowledge and experience proved critical in our ability to make these determinations and carry out the research in a safe, ethical manner.

A.7 Study Timeline

- January 2016: PIs begin conducting observational research on policing in Sorsogon Province, including background interviews and a citizen survey.
- September 2016: PIs initiate discussions with PNP Provincial Director about evaluating his planned community policing initiative.
- February–April 2017: Police hotline rollout and advertisement (evaluated in other research)
- August 2017: Baseline survey of PNP officers
- December 2017–May 2018: Problem Oriented Policing program implementation
- June 2018: Endline survey of PNP officers

A.8 Pre-Analysis Plans

There are two pre-analysis plans relevant to this study, which are available online at https://osf.io/tfpx7/ and https://osf.io/zxefj/. The first is a multi-country PAP for a harmonized study on community policing in six different country sites. The second is a country-specific PAP for the Philippines component of the larger study. [For the purpose of journal review, we append a blinded version of both PAPs below.] The final versions of both PAPs were officially filed in January 2020. The gap between our endline survey and the filing date was driven by the fact that ours was the first country intervention implemented in the larger study referenced in the meta-PAP. The steering committee for the larger study needed to incorporate additional details relevant to the interventions that were rolled out later, and asked that all teams hold off on uploading PAPs until this process was completed. In accordance with this plan, we were restricted from downloading any data until the PAPs were filed.

Because these PAPs include pre-registered analyses for multiple papers, we include information below on 1) where the relevant measures and analyses for this paper are located in the PAPs and 2) how the analyses in the PAP were implemented, including some minor deviations.

A.8.1 Hypotheses and Measurement

Guide to PAP

- The main hypotheses for the *Empathy*, *Accountability* and *Corruption* outcomes (as well as the umbrella hypothesis on *Officer Attitudes*) are pre-registered as Hypothesis 3a under the header "Primary Outcome Family 3" on page 12 of the meta-PAP.
- The tests of these hypotheses (associated with the outcome indices) are specified in Table 4 on page 14 of the meta-PAP and in the table on page 17 of the country-specific PAP.
- The survey items that compose the officer-facing indices specified in Table 4 are found in the table on 59-62 of the meta-PAP. Here you can also find the list of survey items included in each outcome index.
- The hypothesis corresponding with the "Shared Concerns" outcome is pre-registered as hypothesis 3c on page 19 of the country-specific PAP. The construction of the measure is labeled *mip_officer_match* on page 27 of the country-specific PAP.
- The pre-registered method for index construction (including details on standardization and imputation of missing items) is located on page 39-40 of the meta-PAP.
- The remaining hypotheses in both PAPs are citizen-facing hypotheses, and many aspects of the research design, measurement, and estimation are only relevant to these citizen-facing aspects of the study.

Implementation and Deviations

- We did not ask a set of questions relating to a hypothetical scenario about officer abuse of citizens (hypothetical scenario 5 in the meta-PAP). After consultation with PNP leadership, it was determined that these questions risked introducing substantial preference falsification among officers, including on remaining survey items, given the political climate at the time. Three of the questions about this hypothetical scenario (regarding whether officers would report the incident and whether leaders would punish them for it) were included in the pre-registered *Accountability* index, but were not included in our measure. Another subset of these questions were pre-registered as the *Abuse* index in the meta-PAP, which we exclude.
- We inadvertently forgot to include the *Trust* index in our country-specific PAP, although we always intended for it to be a part of this design. This intention is clear in the way that we grouped these questions in our officer survey by asking them sequentially in the section labeled "trust," and these three survey items were labeled with the prefix "*trust_*" in the officer questionnaire included in our country-specific PAP. Nevertheless, we recognize this oversight.
- The pre-registered method for index construction involves imputing missing values for survey items prior to combining the items into the index. We present the results for the indices constructed in this way (which we label the alternative indices) in SI Table A.12, as well as in the other tables throughout the SI. The results are all consistent with simple additive versions of the indices (constructed without imputation), which we present in the main body of the paper for ease of interpretation.
- The pre-registered *Shared Concerns* outcome (*mip_officer_match* in the country-specific PAP) matches officer concerns with the *three* most commonly mentioned citizen issues. These results are presented in Table A.13. In the main body of the paper, we use an alternative version of the measure that matches officer concerns with the *four* most commonly mentioned citizen issues. The results are consistent using these two measures.

A.8.2 Research Design and Estimation

Guide to PAP

- The officer-specific estimation equation was not included in the PAP, which was an oversight. However, the estimation strategy for this component of the study follows the same structure as the other pre-registered specifications relating to officer-facing outcomes in the meta-PAP and the other citizen-facing specifications

in the country-specific PAP. The structure of Equation 1 (panel specification) is pre-registered in the country-specific PAP as the third equation in the page header above Hypothesis 3c.

- The clustering strategy is pre-registered on page 15 of the country-specific PAP.
- The process for dealing with missing data in baseline outcomes by replacing missing values with zeros and adding an indicator variable for missingness is pre-registered on page 40 of the meta-PAP.

Implementation and Deviations

- Equation 2 (cross-sectional specification) is not pre-registered. We include it as a robustness check to mitigate concerns regarding attrition.
- Because the officer randomization occurred within the subset of barangays assigned to POP treatment, the alternate treatment arms included in the estimating equations for the citizen-facing outcomes are not relevant to the officer-facing outcomes.
- On page 15 of the country-specific PAP, we note an issue with the barangay-level blocking code that led us to exclude block fixed effects in our models. However, this issue is specific to the barangay-level randomization and not the officer-level randomization. The officer randomization was blocked based on station and rank, as noted in section 7.3 of the country-specific PAP. This was implemented as planned, and so we retain the fixed effects for these variables, consistent with the estimation strategy for officer-facing outcomes in the meta-PAP.
- None of the heterogeneity analyses were pre-registered, and we do not consider these analyses to constitute hypothesis tests. Rather, the heterogeneity analyses were conducted in an exploratory fashion to show whether patterns in the data were consistent with our inductive theory developed using qualitative evidence.