Pre-Analysis plan for 'Encouraging Cooperation with the State – A Field Experiment on Household Connections to the Police'

*Anna Wilke**

9/10/2018

Contents

*amw2229@columbia.edu Columbia University

This document summarizes key features of a Randomized Controlled Trial implemented in collaboration with MeMeZa Shout Crime Prevention in Hebron, South Africa. This document predates endline data collection, and is therefore blind to outcomes. Any contingency not accounted for in this PAP will be dealt with according to the Standard Operating Procedures for Don Green's lab at Columbia as of June 7, 2016. This study has received approval from the Columbia University Institutional Review Board (IRB), protocol AAAR6346.

1 Motivation

Throughout the developing world, individuals are often reluctant to cooperate with the state's security apparatus. Not only are crimes reported to the police at systematically lower rates in developing compared to developed countries (Soares, 2004), but individuals regularly bypass the state and rely on other actors for protection. Such informal security providers range from organized street committees to vigilant community members and spontaneously formed mobs (Baker, 2007). More than a quarter of South Africans, for example, would first contact not the police, but friends, relatives or a community-level organization if they were under threat of becoming a victim of crime in their home.[1]

Reluctance to cooperate with the state creates a number of problems for governance. First, the police rely on the cooperation of citizens, for example in the form of information provision, to effectively combat crime. A public unwilling to engage with the police may hamper the state's ability to establish order. Second, reliance on informal security providers frequently results in illegal and gruesome punishment of alleged criminals. Anecdotal accounts of suspects being severely beaten or killed are abundant (Adinkrah, 2005; Smith, 2004; Schuberth, 2013; Human Rights Watch, 2012), and official statistics suggest that, in South Africa, group punishment of this kind leads to more than two recorded murders per day (South African Police Service, 2008/2009). Finally, to the extent that it serves as a signal of the dysfunction of the formal justice system, reliance on informal actors may further undermine the legitimacy of the state (Dickson, Gordon and Huber, 2015). In South Africa, where access to the state is highly unequal, the perception that the state is incapable or unwilling to provide protection to parts of the population frequently results in violent protests and social unrest (Alexander, 2010). Interventions that encourage trust in and use of the formal system may thus reduce crime, improve the welfare of those involved in criminal accusations, and ultimately contribute to an environment conducive for economic growth.

This study seeks to understand whether improved access to police protection increases cooperation with the police and reduces demand for informal ways of security provision, especially in the form of group punishment or 'mob justice.' The study takes the form of a field experiment in South Africa that provides households with a direct connection to the police.

2 Context

South Africa exemplifies many of the problems described above. Not only does it have one of the highest murder rates in the world, but according to a recent public opinion survey, 44% of South Africans are

[1] Numbers are taken from the 2014/2015 Victims of Crime Survey conducted by Statistics South Africa.

dissatisfied with the police,[2] and reliance on actors other than the police for protection has resulted in high rates of mob killings. This experiment takes place in Hebron, a semi-urban police precinct with around 42,000 residents in the country's North West province.[3] According to the 2011 Census, the median income in Hebron is slightly below (80%) the median income across all police precincts. The precinct was selected in collaboration with MeMeZa Shout Crime Prevention and the South African Police Service (SAPS), because of its high crime rate and, specifically, the high frequency of house robberies. According to official crime statistics from 2017, the number of house robberies per 1000 people in Hebron was approximately four times as high as the average across all police precincts. Additionally, around half (47%) of the respondents in the baseline survey of this project ($N = 670$) believe it 'somewhat likely' or 'very likely' that an alleged criminal could be beaten up in their community. Support for this form of mob justice is also substantially higher in Hebron than in South Africa at large. Around 30% of respondents supported the beating of alleged wrong-doers. The same scenario elicited only around 12% support in a recent nationally representative opinion survey ($N = 1,176$). One of the most cited justifications for the reliance on mob vigilantism during qualitative interviews was the absence of a timely police response to emergency situations. In fact, the majority of respondents in Hebron (57%) believe that the police would never show up when called in an emergency. Qualitative evidence suggests that, apart from resource constraints, a number of infrastructural conditions contribute to prolonged response times. First, the vast majority of streets in the police precinct are unnamed. Second, there are two separate house numbering systems. Not only do many houses display only one of the two house numbers, but numbers also tend to be out of sequence even within the same numbering system. Many houses do not display any house numbers at all. As a result, specific households are extremely difficult to locate even in broad daylight. The general lack of street lighting exacerbates this problem during nighttime.

3 Intervention Description

The treatment in this study consists of a home-based alarm system that links a household to the police. The alarm is a small electronic device that is activated via a panic button or cell phone. When triggered, the alarm sends text messages directly to the police station, including the police station management. The text messages indicate the alarm owner's name, contact numbers, and landmarks close to the alarm owner's home. Apart from sending text messages to the police, the alarm system is linked to the mobile phones of the executive leaders of the Community Policing Forum (CPF). Each police station in South Africa has a CPF composed of volunteers that serve as liaisons between the police station and the community. In Hebron, the CPF leaders have an office at the police station and are closely involved in the day-to-day operations of the police station. Every alarm owner can also nominate neighbors who receive text messages when the alarm is triggered.[4] The alarm can be triggered silently or such that a bright light flashes outside the alarm owner's house and a siren sounds. The alarm system also has a motion sensor that can be activated when the alarm owner is not at home. Houses equipped with an alarm can be identified by potential criminals, since the siren is visible on the outside, which opens up the possibility of a deterrence effect.

The alarm system was designed by MeMeZa Shout Crime Prevention, a South-African non-profit organization,

[2] See 2014/2015 Victims of Crime Survey by StatsSA.

[3] See here for population estimates per police precinct based on the 2011 Census.

[4] Households without access to electricity received a solar panel to power the alarm. The solar panels are specifically designed to power the alarm system and, beyond that, do not provide general access to electricity.

in consultation with the South African Police Service (SAPS), to overcome several challenges involved in reporting and responding to crime. First among those are the above-mentioned problems with finding households in emergency situations. By sending location details, flashing a light and creating a loud noise, the alarm guides law enforcement to the household in question. Second, the alarm system ensures that household members can contact the police even if they have no credit or if their phone was stolen. Finally, by alerting the police station management and the CPF, the alarm system allows senior police and CPF officials to follow up with individual incidents, thereby increasing the incentives for patrol officers to quickly respond to emergencies.

As part of this project, MeMeZa Shout Crime Prevention agreed to install 100 new alarm systems in a set of randomly selected households in the Hebron police precinct. The first 29 alarms were installed between August 30th and September 3rd 2018. Due to technical issues, the roll-out of the rest of the alarms was delayed and took place in the middle of October 2018. In total, MeMeZa Shout Crime Prevention managed to install 96 out of 100 alarms. 3 households assigned to an alarm refused the alarm. In one household, the alarm could not be installed due to construction work.

4 Theoretical Framework

Broadly, this project seeks to understand the conditions under which individuals are willing to rely on the state rather than on informal actors, specifically community justice in the form of mob violence, for protection and the establishment of order. We distinguish between two main sets of outcomes that we expect to be affected by the treatment: The first contains the willingness of individuals to draw, respectively, on the state and on other community members when (at risk of) being victimized by crime. The second comprises the general support for the sanctioning of criminals through the state or through extra-judicial punishment by the community. While the first set of outcomes concerns the proclivity of individuals to take certain actions when faced with a criminal attack, the second set captures an individual's general demand for the involvement of state and informal institutions in the establishment of order, irrespective of individual victimization.

4.1 Proclivity to reach out to state and community

Consider an individual who is about to be or has been a a victim of crime in her home. For example, as is a common occurrence in the police precinct where this study takes place, a criminal may have entered or may be about to enter the house of the individual to steal her belongings. The individual now faces the choice of reaching out to the police, to other community members, to both or to no one for help. The setting can be thought of as a discrete consumer-choice setting where consumers can purchase more than one service (see e.g. Gentzkow, 2007).

The benefits of reaching out to these actors can take multiple forms. If the crime has not yet been completed, the police or the community may be able to prevent the criminal from harming the victim or from stealing her belongings. Alerting the police or community members may also lead to the apprehension and punishment of the criminal and the recovery of the stolen goods. At the same time, reaching out to either of those actors entails costs. Calling the police requires information about the relevant phone number and, in some cases,

phone credit.[5] The transaction costs involved in reaching other community members are likely negligible. Neighbors are typically so close by that they can be alerted by shouting. Moreover, most households in Hebron own a whistle that is used to alert the community in such emergency situations. Yet, alerting other community members entails the risk that the community may harm the criminal. While physical punishment of the attacker may give psychological benefits to some victims, it is likely to be morally unacceptable to others. Additionally, instigating a mob violence incident bears the risks of legal sanctions. Almost half the respondents in the baseline survey of this project believe it 'very likely' or 'somewhat likely' that the police would hear about a mob violence incident and arrest those involved. In qualitative interviews, both the police and community members narrated multiple cases in which those involved in mob violence faced charges. One respondent even claimed to have mistakenly served a prison sentence for having been involved despite the fact that he tried to calm down the mob and did not participate in the violence.

We expect that the installation of an alarm system in a person's house provides a shock to the expected benefits and costs of reaching out to the police. Most of the benefits outlined above rely on a short response time. In order to offer protection against theft or physical harm and to apprehend the criminal, help needs to arrive at the victim's household in time. As described in the intervention section, the alarm system is designed to reduce response times by making households easier to find. It also improves the incentives of patrol officers to respond in a timely manner by improving the capacity of the police and the CPF to follow up on specific incidents. Finally, by eliminating the need for knowledge of a phone number or phone credit, the alarm system also lowers the costs of reaching out to the police. Even though the alarm can also be linked to neighbors, we nonetheless expect effects on the costs and benefits of alerting the community to be minor. After all, neighbors, in contrast to the police, know already where the house is located and the whistle system provides an inexpensive way of alerting them in an emergency.

One prediction is therefore that the treatment will increase individuals' willingness to reach out to the police in an emergency.

Prediction 1: The treatment will increase individuals' willingness to reach out to the police in an emergency.

An important second question is whether it increases or decreases the willingness to reach out to the community in addition to the police. By sending text messages to both the police and the neighbors, the alarm system does not allow those in the treatment group to use the alarm to alert only the police and not the community. But it does, for example, allow the alarm owner to trigger the alarm silently as opposed to using the siren, which, presumably, limits the extent to which the surrounding community that is not linked to the alarm is aware of the emergency. It is thus still meaningful to ask whether having the alarm increases or decreases an individual's willingness to reach out to the community for help.

The answer to this question will depend on whether police and community response are perceived as complements or substitutes. If the expected net benefit of alerting the community increases with reliance on the police, the treatment should increase the proclivity to alert the community. If the expected net benefit of reaching out to the community decreases with police involvement, the treatment should decrease reliance on the community. Both seem possible. On the one hand, calling the police will increase the risk of legal sanctions if community involvement leads to violence. On the other, if the community knows that the police are on their way, this may decrease the risk of community members engaging in violence in the first place.

[5] There is some confusion in South Africa about which emergency numbers can be called for free and which cost money. The most well-known emergency number, 10111, can only be called for free from a land line, which most people do not have available. The alternative number 112 can be called for free but is less well known.

While the effect of the treatment on the willingness to call the community per se thus seems ambiguous, we expect that a positive shock to the utility of reliance on the police will reduce the willingness to explicitly instigate community violence, as the risk of legal sanctions should be higher when the police are also involved.

Prediction 2a: The treatment may increase or decrease individuals' willingness to reach out to community members in an emergency, depending on whether the police and the community are seen as substitutes or complements.

Prediction 2b: The treatment will decrease the willingness to instigate or participate in community violence against criminals.

4.2 Support for state versus extra-judicial punishment of criminals

Second, we consider how the intervention may affect people's broader demand for the relative involvement of the state and of community members in the sanctioning of criminals, outside of an emergency situation. We presume that, rather than immediate concerns about protection conditional on facing a crime, a key driver of this demand are perceptions of the risk of crime victimization and its relationship to the dominant sanctioning institution. To fix ideas, consider the following simple formalization. Let us denote by e the share of crimes that are dealt with through community sanctions and by $1-e$ the share of crimes that are reported to the police. In other words, as e increases, community members become more involved in punishing crimes and individuals rely less on state institutions such as the police and courts. We abstract here from the possibility that individuals could rely on both. Suppose further that individual i's utility is given by

$$u_i = -q(e)p_i - c_i(e) \tag{1}$$

The first term represents individual i's risk of becoming a victim of crime. We think about this risk as the joined probability of a crime occurring, which happens with probability $q(e)$, and of individual i being targeted, which happens with probability p_i.[6] We assume that the overall probability of a crime occurring, $q(e)$, is a function of the relative involvement of communities and the state in criminal prosecution.[7] Whether the crime rate is increasing or decreasing with community involvement depends on the relative effectiveness of state and community justice in terms of deterrence and incapacitation of criminals. Qualitative evidence suggests that a non-trivial share of residents in Hebron presently believes that $q'(e) < 0$, i.e. that an increase in reliance on extra-judicial punishment of criminals relative to reliance on the state decreases the crime rate. Respondents in qualitative interviews frequently justified their support for mob violence with reference to its deterrence effect and mentioned the inability of the formal justice system to apprehend criminals and deter crime. $c_i(e)$ represents the costs to individual i of increasing reliance on community justice. This cost term can be thought of as incorporating a variety of concerns. For example, individuals may care to varying degrees about due process and thus be opposed to community justice on ethical grounds. Reliance on community justice, due to its capricious nature, also creates the risk of being falsely accused and punished. This constitutes a significant risk, especially for men. Assuming $-\frac{\partial^2 q}{\partial e^2}p_i - \frac{\partial^2 c_i}{\partial e^2} < 0$, individual i's preferred level of community involved in the sanctioning of criminals is implicitly defined by

[6] These two events may not be independent, which we abstract from here.
[7] p_i may also depend on e, which we plan to further explore in the future.

$$-\frac{\partial q}{\partial e}(e^*)p_i = \frac{\partial c_i}{\partial e}(e^*) \tag{2}$$

Given this framework, there are two ways in which we expect the treatment to affect the demand for community involvement in the sanctioning of criminals. First, while it seems unlikely that the installation of a small number of alarms in a large police precinct would result in a big shift in the overall crime rate $q(e)$, we expect a decrease in the risk of crime victimization for the households in which the alarm was installed, i.e. a decrease in p_i. The change in individual i's preferred level of community involvement in the sanctioning of criminals e^* is given by

$$\frac{de^*}{dp_i} = \frac{\frac{\partial q}{\partial e}(e^*)}{-\frac{\partial^2 q}{\partial e^2}p_i - \frac{\partial^2 c_i}{\partial e^2}} \tag{3}$$

The denominator is negative by our assumption about concavity of the objective function and we have argued above that there is evidence that many residents in Hebron believe the numerator to be negative as well. The overall expression is thus positive and we expect support for community involvement in the sanctioning of criminals to decrease as individuals feel more safe. Intuitively, we predict that individuals are less willing to tolerate the risks involved in community justice when they are less concerned about experiencing a crime.

Prediction 3: The treatment will have a negative effect on support for community involvement in criminal sanctioning by reducing individuals' perceived risk of crime victimization.

Second, we expect that having an alarm and thus easier access and potentially more exposure to the police may affect views about the relative effectiveness of the state in combating crime, i.e. perceptions about $\frac{\partial q}{\partial e}$. The change in individual i's preferred level of community involvement in $\frac{\partial q}{\partial e}$ is given by

$$\frac{de^*}{d\frac{\partial q}{\partial e}} = \frac{p_i}{-\frac{\partial^2 q}{\partial e^2}p_i - \frac{\partial^2 c_i}{\partial e^2}} \tag{4}$$

and is thus negative. On the one hand, the fact that the police is involved in the alarm project and potentially more responsive to alarm owners may lead to alarm owners becoming more trusting in the capabilities and motivations of the police. Relative to community justice, they may thus perceive the state to be more effective at reducing crime, i.e. an increase in $\frac{\partial q}{\partial e}$. (Since we assume $\frac{\partial q}{\partial e} < 0$, this involves the effect of increased community involvement relative to police involvement to be 'less negative'). We would expect this to lead demand for community involvement in the sanctioning of criminals to decrease. On the other hand, if treated households have more exposure to the police but the service offered by the police is worse than expected, alarm owners may become even more disillusioned with the police, in which case we would expect demand for community involvement in the sanctioning of criminals to increase. The direction of the effect may depend on individuals' prior satisfaction with the police.

Prediction 4: The treatment will increase or decrease support for community involvement depending on whether it improves or worsens perceptions of police effectiveness. The direction of this effect may be

conditional on prior expectations about the police.

5 Experimental Subject Pool

The unit of randomization in this study is the household. Households were sampled as part of a baseline survey that took place between May 25th and July 27th 2018. The target sample size was 300 households that would be eligible to receive an alarm.

5.1 Sampling strategy for households.

Sampling of households during the baseline survey was based on two different sampling strategies.

5.1.1 Police list

150 households were sampled from a list of households provided by the Hebron police station. The households on the list were selected from the police station's records of households that recently experienced a crime. This is how MeMeZa Shout Crime Prevention normally operates to select alarm recipients in their other projects. In total, the police provided a pool of 390 names and contact details. The survey team managed to geo-locate 336 of these 390 households. 54 households could either not be found or turned out to be duplicates of houses that had already been located. Among the geo-located households, a random walk algorithm was used to identify sets of households that satisfy the constraint that every household is located at least 150m apart from all other households in the set. This distance constraint was imposed to limit spillover effects. Among the sets identified through a large number (at least 1M) of iterations of the random walk algorithm, the set with the overall maximum distance between units and the largest share of units with a distance of at least 200m from all other units in the set was chosen. Due to practical and time constraints, the selection of households started prior to the complete pool of names having been received from the police and before all households had been geo-located. The above procedure was therefore applied repeatedly to subsets of households. For example, in the first round of sample selection, the random walk algorithm was run on the first 92 households that had been geo-located before the start of the survey. From the 92 households, a set of 67 units satisfying the 150m distance constraint was selected. Subsequently, as additional households were geo-located, the random walk algorithm was run again to select sets of additional households subject to the constraint that these households would be at least 150m apart from each other and from the 67 households in the initially selected set.

5.1.2 Listing of households in high crime areas

To the extent that crime victimization results in lower trust in state institutions (Carreras, 2013), the households nominated by the police are of particular interest to this project. On the other hand, the police may mainly know about crimes that happened to households who reported these crimes. Households nominated by the police may thus be among those most likely to collaborate with the police. The second half of the 300 eligible households was therefore selected independent of reported crime victimization from areas in Hebron that are perceived, by the police, as areas with a particularly high crime rate and low trust in the

police. In those areas, a listing exercise was conducted during which enumerators walked through the areas and geo-located the addresses of every tenth house. In total, 946 households were geo-located in 11 areas. The same sequential random-walk based procedure described above was used to identify sets of households that satisfy the constraint that every household is located at least 150m apart from all other households in the set and at least 150m from the set of households sampled from the police list.

Both sampling strategies cover a similar geographic area, since the crime victims identified by the police tend to reside in the same high-crime area that the rest of the sample was drawn from.

5.2 Household replacement criteria

Households identified through the process described above were excluded from the sample if one of the following conditions was true:

- *No adult woman available.* The baseline data collection focused on adult female respondents. As such, all-male households were excluded from the sample.
- *No South African citizen available.* Non-South-Africans are generally believed to be quite mobile and likely to move from one place to the other. They were thus excluded from the sample in order to decrease the risk of attrition.
- *No permanent resident available.* The sample was limited to households that stay permanently in Hebron in order to reduce the risk of attrition.
- *Relevant respondent could not be found after 3 visits.* See below for the within-household sampling strategy used to select respondents from household members. Enumerators were instructed to conduct three visits during different times of the day.
- *Relevant respondent refused to be interviewed.* See below for within-household sampling strategy used to select respondents from household members.

5.3 Within-household sampling strategy during baseline survey

During the baseline survey, one female respondent was interviewed in each eligible household. Among the female members of a given household, we sought to interview the woman who is most involved in household decision-making.

5.4 Practical challenges during baseline survey

In addition to these respondents, the baseline sample also includes 313 respondents who are neighbors of the eligible households. Neighbors were originally included in the design to allow for the assessment of spillovers. Due to practical constraints, this component of the design has been eliminated from the study. Neighbors were selected by asking the corresponding main respondent which neighbor she would most likely reach out to if she was about to become a victim of crime.

Additional challenges arose during the implementation of the above described sampling strategy for households:

- Towards the end of the baseline survey, surveying became impossible in two areas of the Hebron police precinct due to opposition from community members who felt suspicious towards the survey team. Households in these areas were replaced with more households from the other high crime areas in the sample.
- The existence of two separate house numbering systems and double-numbering (more than one house with the same house number) led to inaccuracies in the geo-coordinates captured during the listing exercise.
- In some cases, the list of names provided by the police was outdated and people had moved. In these cases, the new residents of the address listed on the list received from the police were interviewed.
- Due to the general difficulty of locating households, enumerators relied on geo-locations rather than addresses when approaching households selected through the listing of houses in high crime areas. This led to 34 interviews that were conducted far from the originally sampled location.

As a result of these problems, not all households in the sample satisfy the 150m distance constraint. Additional households were included in the sample to alleviate this problem.

The total number of respondents from eligible households included in the baseline survey is 358. 15 respondents refused to be interviewed and 16 respondents could not be found after 3 visits.

		N
Eligible Households	Police List	171
	Listing in High Crime Areas	187
	Total	*358*
Neighbors	Police List	162
	Listing in High Crime Areas	151
	Total	*313*
Total		**671**

5.5 Final selection of households through back-checks

From the 358 households sampled during the baseline survey, a set of 250 households was chosen as the pool of experimental units. The selection was made so as to minimize problems of non-compliance and attrition.

- *Exclusion of households not interested in alarm.* In order to avoid non-compliance, respondents were asked during the baseline survey whether they would be interested in receiving an alarm system.[8] 15 households (7 from the police and 8 form the listing in high crime areas) indicated that they would not want to be considered for an alarm.

[8]Specifically, respondents were asked: 'Before I leave, I would like to know whether you may be interested in receiving a home alarm system to protect your household. The system is an electronic device to be installed in your home that allows you to quickly alert the police or your neighbors. The alarm can also make a loud noise to indicate that there is a problem in your home. We are cooperating with a non-profit organization called MeMeZa that gives out these alarms for free. If you are interested in receiving an alarm, we will pass your contact details on to MeMeZa. However it is not certain that you will receive an alarm, since MeMeZa uses a lottery system to give out the alarms. Depending on the availability of alarms, some of those who are interested will receive an alarm relatively soon, some will receive one later and others may not receive one at all. Would you be interested in receiving an alarm system?'

- *Exclusion of CPF leaders.* The sample contained 4 of the 10 executive members of the CPF. The implementing partner decided to non-randomly allocate 10 alarm units these executive members in order to ensure buy-in for the project.
- *Exclusion of households that could not be reached during back-checks.* Given the challenges of locating households and due to concerns about data quality, it was decided to conduct telephonic back-checks of all remaining households that would be considered for randomization. Households that could not be reached via phone were visited in person. Only households that could be reached during the back-check phase and still showed interest in the alarm system were included in the final experimental subject pool.

	N
Police List	135
Listing in High Crime Areas	115
Total	**250**

5.6 Within-household sampling strategy during endline survey

Outcomes will be measured during an endline survey that will begin on November 6th 2018. Surveying will begin in blocks in which all units assigned to treatment have been treated in the first round of alarm installations at the end of August 2018 (see below for the blocking structure). We aim to ensure a minimum time of roughly four weeks between the earliest date on which an alarm was installed in a given block of households and the endline interview. Among the 250 households in the experimental subject pool, the same women respondent who has been interviewed during the baseline survey will be re-interviewed during the endline survey. Additionally, we will interview one male member of each of these households. Male respondents will be sampled at random from all adult men (older than 18 years) in the household. Where no men reside in a household, an additional female respondent will be sampled at random from all adult women (older than 18 years) in the household. In households that only have one adult member, we will interview only one respondent. According to baseline data, there are 18 single-member households in the sample. We will update the information on household size during the endline survey. Based on the baseline information, we expect to measure outcomes among 482 respondents that will form the main sample.

For safety reasons, enumerators will always work in pairs - even in situations in which one respondent in a given household has already been interviewed and only one outstanding interview is left. We therefore anticipate that there will be situations in which a second enumerator has access time to conduct an additional interview should another adult respondent of the relevant gender be available.[9] In these cases, we may interview more than two respondents per household. In addition to estimates among all respondents, we will report estimates among only the first two respondents per household sampled according to the above mentioned strategy as a robustness check.

[9]Female enumerators are allowed to interview male and female respondents. Male enumerators can only interview male respondents.

6 Estimands

Given that alarms could not be installed in 4 households that were assigned to treatment, we seek to estimate both intent-to-treat effects (ITT) and complier average causal effects (CACE). We focus on effects among the individuals in the experimental subject pool. For some outcomes, we seek to estimate these treatment effects on the household level, i.e. among the households in the sample. Finally, some of the estimands described below pertain to conditional treatment effects and differences between conditional treatment effects.

7 Randomization

7.1 Blocking

Prior to random assignment, the 250 units in the experimental sample were organized into 50 blocks of 5 units using the `blockTools` package. The following code was used for creating blocks.

```
# Subset by sampling method
sample_police <- dplyr::filter(block_data, sample == "police")

sample_random <- dplyr::filter(block_data, sample == "random")

# Check we still have the right sample size
stopifnot(nrow(sample_police) + nrow(sample_random) == 250)


block_vars <- c("latitude_corrected",
                "longitude_corrected",
                "beat_truck_driver",
                "join_mob")

# Make police blocks ------------------------------------------------------

block_object_police <- block(data = sample_police,
                      n.tr = 5,
                      id.vars = "id_resp",
                      block.vars = block_vars)
sample_police$block_id <- createBlockIDs(obj = block_object_police,
                                         data = sample_police,
                                         id.var = "id_resp")


# Make random sample blocks -----------------------------------------------
```

```
block_object_random <- block(data = sample_random,
                       n.tr = 5,
                       id.vars = "id_resp",
                       block.vars = block_vars)
sample_random$block_id <- createBlockIDs(obj = block_object_random,
                                         data = sample_random,
                                         id.var = "id_resp")

# Make sure that block ids are different for the two sample
sample_random$block_id <- sample_random$block_id + 1000

block_data <- rbind(sample_police, sample_random)
```

First, the experimental subject pool was divided into two sets of households according to the way in which households had been sampled, i.e. blocks were formed separately within the group of households that had been sampled from the police list and within the group that had been sampled by listing households in high crime areas. Second, blocks of five households were formed in order to minimize the within-block multivariate Mahalanobis distance of four variables, the latitude and longitude of the household's location and two baseline measures of support for mob violence:

1) Imagine the following situation: A truck driver drove [at random: drunk] through your neighborhood and knocked over a small girl and the girl died. A group of men from your community got hold of the truck driver. Which of the following do you believe they should do?

- The group of men should beat the truck driver to teach him a lesson (1)
- The group should leave it to the police to investigate (0)
- Don't know (NA)
- Refuse to answer (NA)

2) Imagine the following situation: A member of your community is shouting for help [at random: because he has been robbed/ and claims to have been robbed]. The [at random: accused] thief has been surrounded by a group of people who want to beat him up [at random: because they believe the police won't do anything]. Which action are you most likely to take?

- I would join the group in beating up the thief. (2)
- I would avoid the situation and mind my own business. (1)
- I would try to calm the group down and tell them we should wait for the police. (0)
- Don't know (NA)
- Refuse to answer (NA)

Missing values in these measures were imputed via chained equations as implemented in the `mice` package using other measures of attitudes towards the police.

7.2 Random assignment

Within each block of 5, two households were assigned to treatment using the following code that draws on the `randomizr` package:

```
set.seed(1234567)

block_data$treatment <- block_ra(blocks = block_data$block_id,
                                 block_m = rep(2, length(unique(block_data$block_id))))
```

8 Outcomes

8.1 Outcome measurement

Outcomes will be measured during the endline survey. We will measure main outcomes and intermediate outcomes that are related to the described theoretical framework. In addition, we will measure secondary outcomes. While there are reasons to believe that secondary outcomes will also be affected by the treatment, we are less certain about these treatment effects and they are less central to this study.

- Main Outcomes
  - Proclivity to reach out to state and community
  - Support for state versus extra-judicial punishment of criminals
- Intermediate Outcomes
  - Safety and risk of crime victimization
  - Views on the police
- Secondary Outcomes
  - Conceptions of justice
  - Views of government and the state
  - Community Cohesion
  - Views on the CPF

All outcomes have been flagged as such in the attached endline questionnaire. In addition, main outcomes have been marked in green, intermediate outcomes in yellow and secondary outcomes in blue. The attached endline questionnaire also shows how outcomes will be coded. Unless otherwise indicated, the responses 'don't know' and 'refuse to answer' will be coded as missing. Indices will be constructed in the following manner: All variables that form part of the index will be standardized to range from 0 to 1 by dividing by their maximum possible value. Subsequently, they will be summed and divided by the number of variables that form part of the index. Additional information on coding and index creation is provided in the section on outcome-specific hypotheses below.

8.2 Item-level missingness

Non-response to outcome questions will be dealt with through imputation methods. Specifically, we will use multivariate imputation via chained equations (MICE) as implemented in the `mice` package for `R`. Individual

outcomes will be imputed prior to the creation of indices. With one exception, imputations will be performed within outcome subcategories listed above. For example, all outcomes in the category 'Conceptions of justice' or all those in the category 'Views on the police' will be imputed together. In the category 'Proclivity to reach out to state and community', outcomes that ask, respectively, about reaching out to the police and about reaching out to the community will be imputed separately from each other. To assess the robustness of the results, we will also report results based on listwise deletion.

9 Outcome-specific Hypotheses

9.1 Main hypotheses

9.1.1 Proclivity to reach out to state and community

9.1.1.1 Willingness to reach out to police in emergency

Hypothesis: The treatment increases the willingness to reach out to the police in an emergency.

Direction: One-tailed (positive)

Outcome: Additive index of `alert_police` and `first_action_police`, where `first_action_police` is coded as 1 if the respondent mentioned any form of reaching out to the police, including sounding the MeMeZa alarm and as 0 otherwise.

9.1.1.2 Willingness to cooperate with police

Hypothesis: The treatment increases the proclivity to report crimes to the police.

Direction: One-tailed (positive)

Outcome: Index of `report_police`, `share_information` and `report_gbv`

9.1.1.3 Willingness to reach out to community in emergency

Hypothesis: The treatment increases or decreases the willingness to reach out to the community in an emergency.

Direction: Two-tailed

Outcome: Additive index constructed of `alert_community` and `alert_neighbors`

9.1.1.4 Willingness to rely on mob violence when victimized

Hypothesis: The treatment decreases the willingness to rely on mob violence in case of personal victimization.

Direction: One-tailed (lower)

Outcome: `join_beating`

9.1.2 Support for state versus extra-judicial punishment of criminals

9.1.2.1 Support for community involvement in crime prevention

Hypothesis: The treatment increases or decreases support for community involvement in crime prevention.

Direction: Two-tailed

Outcome: `community_deal _crime`

9.1.2.2 Support for mob violence

Hypothesis: The treatment increases or decreases support for mob violence.

Direction: Two-tailed

Outcome: Additive index of `beat _driver`, `beat_known_thief`, `beat_petty_thief`

9.1.2.3 Support for punishment of mob violence perpetrators

Hypothesis: The treatment increases or decreases support for the sanctioning of mob violence perpetrators through the criminal justice system.

Direction: Two-tailed

Outcome: `arrest_mob`

9.2 Intermediate Outcomes

9.2.1 Safety and risk of crime victimization

9.2.1.1 Feelings of safety

Hypothesis: The treatment makes subjects feel safer in their homes.

Direction: One-tailed (upper)

Outcome: Additive index of `feel_safe_night` and `feel_safe_day`

9.2.1.2 Perceived risk of victimization

Hypothesis: The treatment causes subjects to perceive it less likely that they or their families will be victimized by crime.

Direction: One-tailed (lower)

Outcome: `perceived_crime_risk`

9.2.1.3 Crime incidents

Hypothesis: The treatment causes subjects to experience less crime.

Direction: One-tailed (lower)

Outcome: `crime_victimization`

9.2.2 Views on the police

9.2.2.1 Knowledge of ways to reach the police

Hypothesis: The treatment increases the share of subjects who know a way and mention the alarm as a way to reach the police.

Direction: One-tailed (upper)

Outcome: `way_reach_police`, 0 = respondent does not know a way to reach the police, 1= respondent knows a way but does not mention the alarm, 2 = respondent knows a way and mentions the alarm

9.2.2.2 Contact with police

Hypothesis: The treatment increases the share of subjects who have recently spoken to someone from the police.

Direction: One-tailed (upper)

Outcome: `speak_to_police`

9.2.2.3 Perceived response time

Hypothesis: The treatment decreases the perceived response time of the police to an emergency call.

Direction: One-tailed (lower)

Outcome: `response_time`

9.2.2.4 Quality of service

Hypothesis: The treatment improves or worsens the quality of service that subjects anticipate to receive when they take a problem to the police.

Direction: two-tailed

Outcome: Additive index of `take_problem_seriously` and `appear_competent`

9.2.2.5 Collusion with criminals

Hypothesis: The treatment increases or decreases the perception that the police are colluding with criminals.

Direction: two-tailed

Outcome: Additive index of `pay_to_escape` and `collusion_with_criminals`

9.2.2.6 Effective sanctioning of criminals

Hypothesis: The treatment increases or decreases the overall perception that the state is effective at convicting criminals.

Direction: two-tailed

Outcome: `people_go_free`

9.2.2.7 Capacity versus lack of effort

Hypothesis: The treatment increases or decreases the perception that under performance by the police is due to a lack of effort rather than capacity constraints.

Direction: two-tailed

Outcome: `lack_of_effort`

9.3 Secondary hypotheses

9.3.1 Conceptions of justice

9.3.1.1 Punitiveness

Hypothesis: The treatment causes subjects to favor weaker punishment of criminals.

Direction: one-tailed (negative)

Explanation: If the alarm system makes subjects feel safer and less afraid of crime, the deterrence effect produced by harsh sanctions may become less important to them.

Outcome: Additive index of `punishment_preferences`, `courts_punish_not_enough`, `no_mercy`

9.3.1.2 Support for due process

Hypothesis: The treatment causes subjects to show greater concern for the rights of the accused.

Direction: one-tailed (positive)

Explanation: If the alarm system makes subjects feel safer and less afraid of crime, subjects may be more willing to trade off a decrease in the rate of correct convictions for a decrease in the rate of wrongful convictions.

Outcome: Additive index of `support_due_process` and `support_due_process_2`

9.3.2 Views of government and the state

9.3.2.1 Satisfaction with government efforts to keep citizens safe

Hypothesis: The treatment increases or decreases subjects' satisfaction with government efforts to keep citizens safe.

Direction: two-tailed

Explanation: The alarm system makes it easier for treated subjects to reach out to the police. On the one hand, treated subjects may thus be more satisfied with the police and with government efforts to establish order in general. On the other hand, the treatment may decrease satisfaction with the police and government efforts more broadly if treated subjects interact more with the police but are disappointed by the service that the police provides.

Outcome: `government_does_enough`

9.3.2.2 Demand for government spending on policing

Hypothesis: The treatment increases or decreases subjects' demand for government spending on the police

Direction: two-tailed

Explanation: The alarm system may cause subjects to be more or less satisfied with the police, which in turn may change demands for a redistribution of government resources towards or away from the police.

Outcome: Additive index of `spend_police_1` and `spend_police_2`

9.3.2.3 General satisfaction with government

Hypothesis: The treatment increases or decreases subjects' general satisfaction with the government.

Direction: two-tailed

Explanation: The alarm system may cause subjects to be more or less satisfied with the police, which in turn may change subjects' views of the government more broadly.

Outcome: Additive index of `government_corrupt` and `government_unresponsive`

9.3.2.4 Willingness to pay for government services

Hypothesis: The treatment increases or decreases subjects' willingness to pay for government services.

Direction: two-tailed

Explanation: Depending on the effect that the alarm system has on the views on treated subjects on the police and the state more generally, it may make them more or less inclined to rely on the state for services other than security provision.

Outcome: Additive index of `spend_electricity` and `spend_education`

9.3.2.5 Political participation

Hypothesis: The treatment increases or decreases subjects' willingness to engage in political participation.

Direction: two-tailed

Explanation: Depending on the effect that the alarm system has on the views on treated subjects on the police and the state more generally, it may make them more or less inclined to participate politically.

Outcome: Additive index of `voice_heard_1`, `voice_heard_2` and `vote_intention`

9.3.3 Community Cohesion

9.3.3.1 Frequency of interaction with community

Hypothesis: The treatment increases the frequency at which subjects interact with their neighbors and community members who are living close-by.

Direction: one-tailed

Explanation: Treated subjects can also link their neighbors to their alarms. Knowing that they may want to rely on their neighbors in case of an emergency may cause subjects to invest in the relationships with their neighbors.

Outcome: Additive index of `discuss_neighbors` and `neighbor_children`

9.3.3.2 Trust in neighbors

Hypothesis: The treatment increases the proportion of subjects that trusts their neighbor.

Direction: one-tailed (upper)

Explanation: Treated subjects can also link their neighbors to their alarms. Knowing that they may want to rely on their neighbors in case of an emergency may cause subjects to invest in the relationships with their neighbors.

Outcome: `trust_neighbor`

9.3.4 Views on the CPF

9.3.4.1 Interaction with the CPF

Hypothesis: The treatment increases contact between the subject and the CPF.

Direction: one-tailed (positive)

Explanation: Since the alarm system is also linked to the executive members of the CPF, we expect the treatment to increase knowledge about and interactions with the CPF.

Outcome: `interaction_cpf` coded as 0 = does not know about CPF, 1 = knows about CPF but is not a CPF member and has not spoken to anyone from the CPF, 2 = knows about CPF and has spoken to a CPF member over past month but is not a member of the CPF, 3 = knows about CPF and is a member of the CPF.

9.3.4.2 Willingness to reach out to CPF

Hypothesis: The treatment increases the willingness to reach out to the CPF in an emergency.

Direction: one-tailed (positive)

Explanation: Since the alarm system is also linked to the executive members of the CPF, we expect the treatment to make it easier for subjects to reach out to the CPF in an emergency.

Outcome: `reach_out_cpf`

9.3.4.3 Perceived effectiveness of the CPF

Hypothesis: The treatment increases or decreases the perception that the CPF helps to prevent crime.

Direction: two-tailed

Explanation: As with the police, the treatment makes it easier for treated subjects to reach out to the CPF, but could improve or worsen perceptions of the effectiveness of the CPF depending on expectations.

Outcome: `cpf_helps`

9.4 Sub-Group Analyses and heterogeneous effects

9.4.1 Gender

One main set of sub-groups of interest to this study are those defined by gender. When it comes to security concerns, men and women differ in important ways. First, prior research suggests that women tend to be more afraid of crime than men (Stanko, 1995). Women are also more supportive of mob violence across a range of contexts including Uganda, Tanzania and Papua New Guinea. Apart from the fear of crime, such higher levels of support may also be related to differential exposure to one of the key risks involved in mob justice, namely the risk of being falsely accused and assaulted for a crime one did not commit. In qualitative interviews, respondents overwhelmingly agreed that it is much less likely for women to be attacked by mobs than for men. Interestingly, women seemed also convinced that the risk of false accusations through the community is small, whereas especially younger men perceived this risk as substantial. Given these differences, we will separately analyse treatment effects among men and women for the following groups of outcomes:

- Safety and risk of crime victimization
- Proclivity to reach out to state and community
- Support for state versus extra-judicial punishment of criminals

The direction of hypothesis tests will remain the same as specified above. We will also estimate the difference in conditional treatment effects. Since the sign of this difference is a priori ambiguous, we will use a two-tailed test. Finally, we will perform the same analyses for the group that faces the highest risk of wrong accusations, which appear to be men who are younger than 35 years old.

9.4.2 Prior assessment of police

We expect that a closer connection to the police is more likely to improve perceptions of the police among subjects who were critical of the police at baseline. Conversely, we expect the potential for subjects to be disappointed with service quality when brought into closer contact with the police to be greater among those who thought highly of the police to begin with. We will use baseline data to construct an

index `police_evaluation_bl` using the following three variables measured at baseline: `response_time_bl`, `police_quality_bl`, `people_go_free_bl` (see baseline questionnaire for details on coding). Prior to the creation of the index, the variables `response_time_bl` and `people_go_free_bl` will be reversed such that higher values indicate a more positive assessment of the police. Since we have only interviewed the female respondent from each household at baseline, we will treat this measurement as a household-level covariate, assuming that views on the police correlate at the household level. Using the same procedure, we will create an index `police_evaluation` using the following outcomes measured during the endline survey:

- `response_time`
- Quality of service (additive index of `take_problem_seriously` and `appear_competent`)
- Collusion with criminals (additive index of `pay_to_escape` and `collusion_with_criminals`)
- `people_go_free`

We will separately estimate treatment effects among subjects from households whose baseline scores on `police_evaluation_bl` are lower or equal to the median of this index and subjects from households whose baseline scores fall above the median. We will conduct one-tailed hypothesis tests, since we expect the treatment to improve perceptions of the police among those who were critical of the police at baseline and worsen them among those who had favorable views of the police at baseline. Additionally, we will estimate the difference in conditional treatment effects and conduct a one-tailed hypothesis test, since we expect treatment effects to be smaller (more negative) among subjects with a high opinion of the police at baseline. If we are able to reject the null hypothesis of this test on the 10% significance level, we will assess the same conditional treatment effects and the difference between them with regard to the main outcomes of this study (outcomes in the categories 'Proclivity to reach out to state and community' and 'Support for state versus extra-judicial punishment of criminals').

9.4.3 Mob violence prime and prior beliefs about police's attitude towards mob violence

The treatment should increase the willingness to reach out to the community if community and police response are seen as complements. It should decrease the willingness to reach out to the community if they are seen as substitutes. As mentioned above, a key factor that may drive perceptions of complementarity and substitutability is the risk that community involvement can lead to mob violence: On the one hand, calling the police will increase the risk of legal sanctions if community involvement leads to violence. On the other, if the community knows that the police are on their way, this may decrease the risk of community members engaging in violence in the first place. We therefore expect that effects of the treatment on the willingness to reach out to the community to be greater (in absolute value) if respondents are primed to think that reaching out to the community could lead to mob violence. As can be seen in the attached questionnaire, we randomly assign respondents to receive such a prime as part of the question wording of the outcome denoted by `alert_community`. We will therefore estimate the difference in conditional treatment effects among subjects that did and did not receive this prime. We will conduct a one-tailed hypothesis test, where the direction of the hypothesis test will depend on the sign of the estimated main effect of the treatment on the willingness to reach out to the community.

By the same logic, we also expect that perceptions of substitutability and complementarity to be affected by the belief that the police would actually take action to sanction those who participate in mob violence. This belief was measured (for one member per household) during the baseline

survey (see `mob_violence_police_reaction_bl` in the attached questionnaire). We will again treat `mob_violence_police_reaction_bl` as a measurement on the household level, assuming that views on the police are correlated within households. We will compare the conditional treatment effects among subjects in households that believe it 'Not likely at all' or 'Not very likely' that the police would arrest those involved in mob violence to the conditional treatment effect among subjects in households that believe it 'Somewhat likely' or 'Very likely' that the police would arrest those involved in mob violence, using the following two outcomes: 'Willingness to reach out to community in emergency' and 'Willingness to rely on mob violence when victimized.' We expect stronger treatment effects (in absolute value) among those who find it 'Somewhat likely' or 'Very likely' that the police would arrest participants and will therefore conduct a one-tailed test, where the direction of the hypothesis test will again depend on the sign of the estimated main effect of the treatment on these outcomes.

9.4.4 Community relations

As outlined above, we expect the alarm to make subjects' feel safer primarily through its effect on access to the police and the capacity of the police to intervene in emergency situations. Neighbors can easily be alerted even without the police and don't face the same issues with finding a household. Nonetheless, we cannot rule out that the alarm will also have impacts that stem from the fact that neighbors have been linked to the alarm and receive text messages when the alarm is triggered. If such effects exist, we would expect them to be strongest among individuals that have neighbors who would be willing to help them in case of an emergency. To explore this possibility, we will assess whether the alarm is more effective among households that are in good standing with their neighbors and the people in their community. Specifically, we will use baseline data to create an index `community_ties_bl` using the following two variables measured at baseline: `able_to_name_bl` and `trust_neighbor_bl` (see baseline questionnaire for details on coding). Since we have only interviewed one female respondent from each household at baseline, we will again treat this measurement as a household-level covariate, assuming that views on the community correlate at the household level.

We will separately estimate treatment effects on outcomes in the category 'Safety and risk of crime victimization' among subjects from households whose baseline scores on `community_ties_bl` are lower than the median of this index and subjects from households whose baseline scores are equal or larger than the median. Additionally, we will estimate the difference in conditional treatment effects and conduct a one-tailed hypothesis test, since we expect the alarm to be, if anything, more effective among subjects that are in a good standing with their surrounding community at baseline. If we are able to reject the null hypothesis of this test on the 10% significance level, we will assess the same conditional treatment effects and the difference between them with regard to the main outcomes of this study (outcomes in the categories 'Proclivity to reach out to state and community' and 'Support for state versus extra-judicial punishment of criminals').

10 Estimation

10.1 Main specifications

The unit of observation for almost all analyses will be the respondent. Intent-to-treat effects will be estimated using OLS regression and the following specification:

$$\boldsymbol{Y} = \alpha + \tau \boldsymbol{z} + \delta \boldsymbol{n} + \boldsymbol{B}\boldsymbol{\gamma} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{5}$$

where $\boldsymbol{Y}$ is a vector of outcomes; α is an intercept; τ is the intent-to-treat effect (ITT) among the respondents in the subject pool; $\boldsymbol{z}$ is a vector of treatment assignments; $\boldsymbol{n}$ is a vector of cluster sizes (number of respondents interviewed in household j) and δ is the associated coefficient; $\boldsymbol{B}$ is a matrix of block indicators and $\boldsymbol{\gamma}$ a vector of associated coefficients; $\boldsymbol{X}$ is a matrix of other covariates and $\boldsymbol{\beta}$, again, a vector of associated coefficients; $\boldsymbol{\epsilon}$ is a vector of error terms that allows for clustering at the household level.

We condition on cluster size, because differential household sizes may lead to heterogeneity of cluster sizes in the sample. In all households that have only one member - 18 households according to the baseline survey - we will only be able to interview one respondent. In all other households, two or more respondents will be interviewed. If potential outcomes are correlated with cluster size, this may bias effect estimates.

The covariates in $\boldsymbol{X}$ will be selected using lasso regression (see details below). For transparency, we will report results with and without the inclusion of the covariates selected through the lasso procedure. The bare-bones version of the specification will still condition on block fixed effects and household size.

For one outcome, `crime_victimization`, the unit of observation will be the household. To estimate the effect on the occurrence of crime, we will collapse the data set to the household level, averaging covariates across respondents within a given household. Households will be coded as having experienced a crime if at least one of the household members reports that a household member has experienced a crime in the relevant time period. We will use the same specification to estimate ITTs among households, except for that error terms will not be adjusted to allow for clustering on the household level.

10.2 Covariate selection

We will use lasso regression to select the minimal number of covariates that best predict each outcome, and include only these in our estimation. The pool of covariates will be described in more detail below. We will perform lasso on covariate data that has item-level missingness removed through multiple chained equations (see description of the procedure for outcome measures above). Imputations will be performed using the entire set of covariates.

The lasso procedure that we plan to use features a generalized linear model with lasso penalization, and is implemented in the `glmnet` package for R. The loss function requires selecting a regularization parameter, λ, that determines the severity of the penalty for including extra covariates. Since this regularization parameter cannot be optimally chosen in advance, we will select it using 10-fold cross-validation. Specifically, for each outcome, we will choose the λ that minimizes the 10-fold cross-validation error averaged over 10 runs (since the folds are chosen at random). Only the covariates retained by the lasso will be included in the specification. In other words, for each outcome, we may include a different number of covariates on the right-hand side.

We will perform this lasso variable selection method using the entire pool of covariates that are listed in the appendix. We will also include, in the lasso procedure, block indicators and the cluster-size covariate, even though these variables will be included on the right-hand side of the regression irrespective of whether they are chosen by the lasso procedure.

The pool of covariates includes three sets of covariates:

- *Covariates from endline survey.* Data quality is expected to be better in the endline survey than in the baseline survey. A limited set of covariates that are plausibly unaffected by the treatment will therefore be taken from the endline survey. Where questions are asked to only one respondent in the household, these measurements will be interpreted as household level measurements and, as such, the same values will be imputed for the other respondent from the same household.
- *Covariates from baseline survey.* Most covariates are taken from the baseline survey. These include pre-treatment measurements of some of the outcomes. Only one measurement exists per households for covariates that are taken from the baseline survey. Again, this measurement will be interpreted as a household level measurement and the same values will be imputed for the other respondent from the same household.
- *Covariates from baseline and endline survey.* These covariates are plausibly unaffected by the treatment and will be constructed from baseline and endline measurements. Respondents who were asked these questions during the baseline will not be asked again during the endline. For each respondent, we will use whichever measurement is available, either the baseline or the endline measurement.

All covariates have been marked in the attached baseline and endline questionnaires. For questions with many unordered answer categories (e.g. `religion` or `education`), we will create indicator variables for every category that has been chosen by at least 10% of the respondents. Smaller categories will be aggregated to a 'minority' category (e.g. `minority_religion`). These variables have been marked with the word 'indicators' in parentheses after the variable name in the attached questionnaires as well as in the variable lists in the appendix. More information on how covariates will be coded can be found in the questionnaires and in the list of covariates in the appendix.

10.3 Characterization of uncertainty

Randomization inference using the random assignment function described above will be used to calculate p-values on quantities of interest. These p-values will be considered the final arbiters on the inference drawn. Standard errors will serve primarily as a heuristic and will not form the basis for inference. Cluster-robust standard errors will be calculated using the sandwich package for `R` for all least squares specifications on the individual level. Given the large number of clusters in this study, these methods for estimating standard errors are expected to generate reliable results.

10.4 Estimation strategy for heterogeneous effects

Conditional treatment effects and differences between them will be calculated by including an interaction with the treatment indicator on the right hand side of the regression.

11 Threats to Inference

11.1 Attrition

We will assess the rate of attrition based on the assumption that two respondents should have been sampled in each household, except for the 18 households for which information from the baseline survey suggests that only 1 adult resides in the household. For the latter group of households, we assume that only one respondent should have been interviewed. We will update these assumptions in light of new information about household size that will be acquired during the endline survey.

We will conduct two tests to assess whether attrition is related to treatment and whether the relationship between baseline covariates and attrition varies across treatment groups:

First, we will conduct two-tailed unequal-variances t-test of the hypothesis that treatment does not affect the attrition rate. We will conduct this test using randomization inference, i.e. we will compare the observed t-statistic to the distribution of t-statistics under random assignment of treatment using the random assignment function specified above.

Second, we will regress an attrition indicator on treatment, a set of baseline covariates, and treatment-covariate interactions. We will select the set of covariates through the same lasso procedure detailed above, where we use the attrition indicator as an outcome and include all covariates that have been collected during the baseline survey (see appendix for a list of these covariates) on the right-hand side. The regression of the attrition indicator on baseline covariates and treatment-covariate interactions will contain only the covariates that retain non-zero coefficients in the lasso procedure. We will then perform an F-test of the hypothesis that all the treatment-by-covariate interaction coefficients are zero. Again, we will rely on randomization inference to conduct this test.

If none of the tests produces a p-value smaller than 0.05 or if the rate of attrition is smaller than 25%, we will report naive estimates among the respondents for whom we have obtained outcome data. Additionally, we will assess the robustness of our results by reporting extreme value bounds.

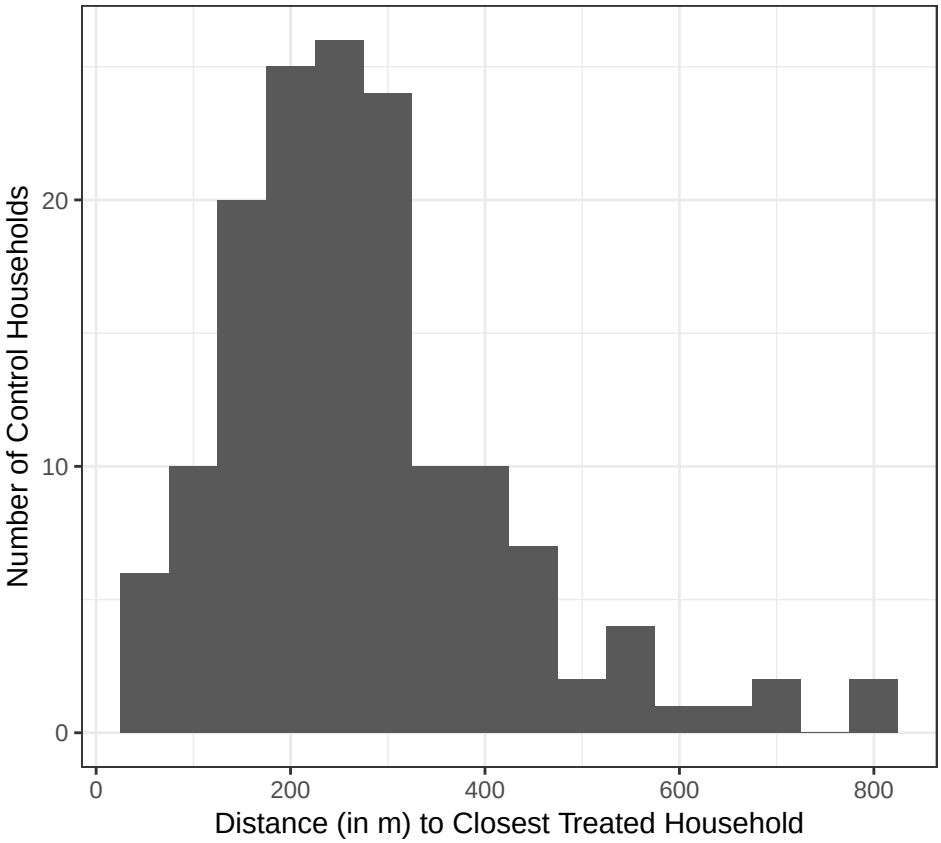
Otherwise, we will rely on a second round of sampling of attrited respondents in combination with an extreme value bounds approach. In this case, we will randomly sample 20 attrited respondents, 10 in the control group and 10 in the treatment group.

11.2 Spillovers

The efforts to sample households that are located at least 150 m apart in order to limit spillovers were only partially successful. The figure below shows the distribution of the minimum distance between a household in the control group and the closest household in the treatment group.

There are several ways in which the effects of an alarm system may spill over to households located nearby. Qualitative evidence collected during exploratory work suggests that individuals living close to those who own an alarm also have increased access to the police, since they tend to ask alarm owners to trigger the alarm on their behalf in emergencies. Moreover, criminals may not be able to discern which houses in an area

are protected by an alarm and may thus decide to avoid an entire group of houses, once it becomes known that one house is protected by an alarm.



While the main analysis assumes no spillover effects, we will explore the existence of spillover effects in a separate analysis. As part of this analysis, we will assume that subjects have not two but three potential outcomes, denoted by Y_{00}, Y_{10} and Y_{01}. Y_{00} represents the untreated potential outcome. We assume that subjects reveal their untreated potential outcome if no alarm system is installed in their house and no household within a 200 m radius of this household is treated. Y_{10} represents the treated potential outcome, which is revealed if an alarm system has been installed in a subject's home. We assume that subjects reveal Y_{01} if at least one household within a 200 m radius of the household has been treated but no alarm system has been installed in their own house. The treatment status of subject i can then take on three different values, $d_i \in \{\text{control, indirectly treated, directly treated}\}$. Based on this model of spillovers, we define two estimands of interest:

- Effect of direct treatment: $E[Y_{10} - Y_{00}]$
- Effect of indirect treatment: $E[Y_{01} - Y_{00}]$

We will obtain estimates of these quantities by estimating two different OLS regression models using specifications that are almost identical to the one given in equation 5. To estimate the direct treatment effect, we will subset the data to subjects living in households that were assigned to the control condition or the direct treatment condition. The vector of treatment assignment indicators $\mathbf{z}$ will reflect whether a subject lives in a household that has been assigned to direct treatment or to the control group. To estimate the indirect treatment effect, we will subset the data to subjects living in households that were assigned to

the control condition or the indirect treatment condition. The vector of treatment assignment indicators **z** will reflect whether a subject lives in a household that has been assigned to indirect treatment or to the control group. The probability of being assigned to each of these conditions varies across households, since it depends on the spatial arrangement of units as well as the blocking structure. Moreover, some households have zero probability of being assigned to the indirect treatment. While there are 52 indirectly treated households in the sample, 82 households are further than 200m apart from any other household in the sample and could thus not have been indirectly treated. When estimating direct and indirect treatment effects, we will therefore weight each unit by the inverse of its probability of being assigned to the condition that it was assigned to and exclude units that had zero probability of being assigned to the respective condition. Treatment assignment probabilities will be calculated by performing a large number of treatment assignments using the random assignment function described above and averaging across assignments.

11.3 Non-compliance

As mentioned above, four households that were assigned to receive an alarm did not comply with their treatment assignment. The implementing partner kept the four alarm units that could not be installed, limiting the potential for two-sided non-compliance. Prior to endline data collection, five randomly selected households from the set of 96 households that were reported to have received an alarm by the implementing partner were called to verify compliance. All five households reported that they indeed received an alarm system and that the alarm system is operational.

In order to avoid having to disclose the relationship between alarm installations and the endline survey to enumerators, enumerators will not ask any questions about compliance during the endline survey. Instead, the following two strategies will be used to further assess compliance: First, supervisors will be instructed to note down whether they can spot an alarm siren at the respondents' premises. Second, we will assess compliance through a separate phone call that will take place after the endline survey.

We will contact one respondent in every household in the treatment group. If more than one respondent has been interviewed from a household, we will randomly select one respondent to be contacted first. If this respondent does not pick up the call, we will attempt to contact the second respondent. Respondents who answer the call will be asked whether a MeMeZa alarm system has been installed in their household. If the first respondent who picks up the phone does not know about the MeMeZa alarm system, the second respondent will be contacted. A household will be defined as treated if at least one of the respondents from a given household reports that a MeMeZa alarm system has been installed in that household.

If this procedure does not reveal any non-compliance in the treatment group beyond the four alarms that could not be installed, non-compliance is unlikely to have occurred in the control group as well, since the implementing partner, MeMeZa, had a fixed number of 100 alarms available to install in the Hebron police precinct.[10]

In general, the risk of two-sided non-compliance is limited by the fact that the contact details of households in the control group were not available to MeMeZa. Given the size of the police precinct, it appears unlikely that, were MeMeZa to accidentally install an alarm system in the wrong household, this household would be

[10] 10 additional alarms were non-randomly allocated to the leadership of the CPF. Yet, all members of the leadership of the CPF have been removed from the subject pool prior to randomization.

part of the control group. That said, local community members were involved in the installation of alarm systems and may have had incentives to divert alarms towards households in the control group.

If the calls to the treatment group do not reveal any or very limited non-compliance, we will contact a random sample of 10 households from the control group. If the rate of non-compliance in the treatment group is greater than 10%, the same phone call procedure will be used to assess non-compliance for all households in the control group. Should two-sided non-compliance occur, it seems likely that the assumption of monotonicity would hold. If community members were to divert alarm systems to favored households in the control group, these households would have likely also received an alarm system if they had been treated.

We will report both, estimates of the intent-to-treat (ITT) and estimates of the complier average causal effect (CACE). Estimates of the CACE will be obtained through instrumental variable regression using the same specifications as described above. Hypothesis tests that involve the instrumental variable estimates will be based on the estimated ITT and on p-values obtained through randomization inference as described above. We will also report cluster-robust standard errors.

During the compliance phone call, we will also ask respondents whether their MeMeZa alarm system is operational. In principle, respondents with a non-functioning MeMeZa alarm system could be defined as partially treated and those with a functioning alarm system could be defined as 'fully treated'. Yet, it seems unlikely that the exclusion restriction required to assess the effect of 'full treatment' in this sense through an instrumental variables approach would be satisfied. After all, the effect of receiving a non-functioning alarm system may not be zero. Yet, assuming that receipt of a non-functioning alarm system produces an effect that has the same sign as the effect of a fully functioning alarm system,[11] we can find an upper bound for the CACE by classifying those whose alarm system is not functioning as 'untreated.' We will provide these upper bounds as robustness checks for our main results.

12 Balance on Covariates

We examine balance on observable pre-treatment covariates measured during the baseline survey ($N = 250$). For each covariate, we test for a significant relationship to the treatment using randomization inference to conduct a likelihood ratio test. The first column of the table below shows the name of the covariate and the following columns show means of the covariate in the treatment and control condition. The last column shows the p-value from the likelihood ratio test. The 'full' model regresses the covariate on a treatment indicator and block fixed effects. The restricted model regresses the covariate on block fixed effects only. The observed likelihood ratio is compared to 2000 likelihood ratios simulated under the null of no effect of treatment on the covariate for all units by re-permuting the treatment assignment and re-estimating the likelihood. The p-value is equal to the proportion of such simulations at least as great as the observed likelihood ratio. Note that under independence, in expectation x% of the covariates should exhibit imbalance that is significant at the x% level. Table 3 reports balance of 73 covariates among all respondents in the baseline: 6/73 (8%) tests exhibit a p-value equal to or less than .05.

[11] This assumption may be satisfied, for example, if even a dysfunctional alarm system increases feelings of safety due to the visible siren at the outside of the house that may have a deterrence effect. Yet, it may not be satisfied, for example, if receipt a fully functioning alarm system increases satisfaction with the police while receipt of a dysfunctional alarm system results in additional disappointment with the police.

	Control	Treatment	P-value
electric_stove_bl	0.80	0.91	0.01
pensions_bl	0.07	0.15	0.03
prepaid_electricity	0.82	0.90	0.03
number_births_bl	2.77	2.39	0.05
criminals_from_area_bl	0.43	0.30	0.05
approached_police_bl	0.65	0.52	0.05
criminals_from_oth_area_bl	0.21	0.31	0.07
discuss_neighbors_bl	1.71	1.48	0.08
own_refuse_dump_bl	0.80	0.89	0.08
pray_private_bl	7.59	7.74	0.09
microwave_bl	0.59	0.69	0.11
pipe_in_yard_bl	0.60	0.70	0.13
dishwasher_bl	0.05	0.01	0.15
prisoners_guilty_bl	0.47	0.56	0.18
spend_police_1_bl	0.25	0.18	0.20
spend_electricity_bl	0.56	0.48	0.21
interest_pulic_affairs_bl	2.21	2.06	0.22
member_organization_bl	0.78	0.84	0.23
government_does_enough_bl	0.53	0.60	0.25
police_ask_for_bribe_bl	0.88	0.76	0.29
pit_latrine_bl	0.72	0.66	0.30
experienced_violent_crime_bl	0.11	0.15	0.32
pay_tv_bl	0.45	0.51	0.33
other_organizations_bl	0.07	0.04	0.35
flush_toilet_tank_bl	0.15	0.18	0.41
sotho_bl	0.14	0.10	0.42
washing_machine_bl	0.48	0.53	0.49
trust_neighbor_bl	0.77	0.73	0.50
discuss_government_bl	2.19	2.31	0.51
satisfaction_services_bl	1.16	1.10	0.51
mob_violence_plausibility_bl	1.65	1.74	0.53
religious_service_bl	1.38	1.31	0.56
adequate_force_bl	0.56	0.59	0.59
observed_conditions_bl	2.69	2.61	0.62
motor_vehicle_bl	0.23	0.20	0.63
guard_dogs_bl	0.22	0.25	0.63
mob_violence_police_reaction_bl	1.83	1.76	0.64
shout_to_alert_bl	0.75	0.73	0.70
flush_toilet_public_bl	0.13	0.15	0.70
government_corrupt_bl	0.60	0.62	0.71
voice_heard_bl	0.92	0.88	0.72

number_children_bl	1.81	1.89	0.73
state_official_bl	0.12	0.13	0.73
punishment_preferences_bl	0.71	0.73	0.73
call_police_bl	2.27	2.31	0.74
length_stay_bl	4.15	4.19	0.74
tswana_bl	0.68	0.66	0.75
attend_meetings_cpf_bl	1.53	1.48	0.78
join_mob_bl	0.45	0.43	0.78
police_quality_bl	1.59	1.55	0.79
able_to_name_bl	1.81	1.84	0.80
blow_whistle_bl	0.14	0.15	0.82
street_committee_connection_bl	0.42	0.39	0.82
beat_truck_driver_bl	0.28	0.28	0.82
discussed_crime_bl	0.90	0.91	0.83
attend_meetings_street_committee_bl	0.43	0.46	0.83
report_informal_provider_bl	0.77	0.76	0.84
spend_police_2_bl	0.53	0.54	0.84
feel_safe_bl	0.27	0.26	0.86
number_school_children_bl	1.38	1.41	0.87
government_unresponsive_bl	0.77	0.79	0.87
number_incidents_bl	0.94	0.91	0.88
response_time_bl	3.13	3.16	0.88
salary_bl	0.31	0.32	0.88
know_state_official_bl	0.40	0.39	0.89
cpf_connection_bl	1.35	1.33	0.89
social_grants_bl	0.43	0.43	0.92
perceived_crime_risk_bl	1.87	1.88	0.94
pipe_in_house_bl	0.10	0.10	0.94
know_number_bl	0.81	0.82	0.95
due_process_bl	0.87	0.87	0.96
hh_size_bl	4.82	4.81	0.97
spend_education_bl	0.61	0.62	0.99

Table 3: Balance on covariates among respondents in baseline.

13 Appendix

13.1 List of Covariates

13.1.1 Covariates from endline survey

- kind_day

- age
- female
- non_black
- hh_position (indicators)
- marital_status (indicators)
- education (indicators)
- religion (indicators)
- employment_status (indicators)
- earn_salary
- floor_material (indicators)
- others_present
- interview_language (indicators)
- cooperative
- traditional_background (indicators)

13.1.2 Covariates from baseline survey

- observed_conditions_bl
- hh_size_bl
- number_children_bl
- number_school_children_bl
- number_births_bl
- discuss_neighbors_bl
- religious_service_bl
- pray_private_bl
- state_official_bl
- know_state_official_bl
- able_to_name_bl
- trust_neighbor_bl
- crime_perpetrators_bl (indicators)
- member_organization_bl
- interest_public_affairs_bl
- discuss_government_bl
- discussed_crime_bl
- government_corrupt_bl
- government_does_enough_bl
- government_unresponsive_bl
- voice_heard_bl
- satisfaction_services_bl (additive index out of 2 of the following items that respondent was randomly assigned to answer: satisfaction_schools_bl, satisfaction_electricity_bl, satisfaction_roads_bl, satisfaction_water_bl)
- due_process_bl
- punishment_preferences_bl

- courts_punish_not_enough_bl
- feel_safe_bl
- perceived_crime_risk_bl
- first_action_bl (indicators)
- call_police_bl
- report_informal_provider_bl
- number_incidents_bl
- experienced_violent_crime_bl
- approached_police_bl
- police_quality_bl
- response_time_bl
- know_number_bl
- police_ask_for_bribe_bl
- adequate_force_bl
- prisoners_guilty_bl
- cpf_connection_bl
- attend_meetings_cpf_bl
- street_committee_connection_bl
- attend_meetings_street_committee_bl
- other_organizations_bl
- mob_violence_plausibility_bl
- mob_violence_police_reaction_bl
- beat_truck_driver_bl
- join_mob_bl
- spend_police_1_bl
- spend_electricity_bl
- spend_police_2_bl
- spend_education_bl
- income_source_bl
- water_source_bl (indicators)
- electricity_source_bl (indicators)
- toilet_facility_bl (indicators)
- rubbish_collection_bl (indicators)
- pay_tv_bl
- electric_stove_bl
- microwave_bl
- washing_machine_bl
- dishwasher_bl
- motor_vehicle_bl
- guard_dogs_bl
- private_security_bl

13.1.3 Covariates from baseline and endline survey

- `length_stay`, `length_stay_bl`
- `main_language`, `main_language_bl` (indicators)

References

Adinkrah, Mensah. 2005. "Vigilante homicides in contemporary Ghana." *Journal of Criminal Justice* 33(5):413–427.

Alexander, Peter. 2010. "Rebellion of the poor: South Africa's service delivery protests–a preliminary analysis." *Review of African political economy* 37(123):25–40.

Baker, Bruce. 2007. *Multi-choice policing in Africa*. Nordiska Afrikainstitutet.

Carreras, Miguel. 2013. "The impact of criminal violence on regime legitimacy in Latin America." *Latin American Research Review* pp. 85–107.

Dickson, Eric S, Sanford C Gordon and Gregory A Huber. 2015. "Institutional sources of legitimate authority: An experimental investigation." *American Journal of Political Science* 59(1):109–127.

Gentzkow, Matthew. 2007. "Valuing new goods in a model with complementarity: Online newspapers." *American Economic Review* 97(3):713–744.

Human Rights Watch. 2012. Mob Justice in Burundi. Official Complicity and Impunity. Technical report.
URL: *https://www.hrw.org/sites/default/files/reports/burundi0310webwcover.pdf*

Schuberth, Moritz. 2013. "Challenging the weak states hypothesis: Vigilantism in South Africa and Brazil." *Journal of Peace, Conflict & Development* 20:38–51.

Smith, Daniel Jordan. 2004. "The Bakassi boys: vigilantism, violence, and political imagination in Nigeria." *Cultural Anthropology* 19(3):429–455.

Soares, Rodrigo R. 2004. "Development, crime and punishment: accounting for the international differences in crime rates." *Journal of development Economics* 73(1):155–184.

South African Police Service. 2008/2009. Annual Report 2008/2009. Technical report.
URL: *http://www.khayelitshacommission.org.za/bundles/category/14-file-c.html?download=307:File%20C*

Stanko, Elizabeth A. 1995. "Women, crime, and fear." *The Annals of the American Academy of Political and Social Science* 539(1):46–58.