# How Does the State Replace the Community? Experimental Evidence on Crime Control from South Africa*

Anna M. Wilke†

April 6, 2023

**Abstract**

Mob vigilantism - the punishment of criminal suspects by groups of citizens - is widespread throughout the developing world. This paper sheds light on the relationship between state capacity and citizens' choice between reliance on the state and vigilantism. I implemented a field experiment in South Africa that randomly varies the capacity of police to locate households. Findings from surveys conducted several months later suggest households that have become legible to police are more willing to rely on police and less willing to resort to vigilantism. An additional information experiment points towards increased fear of state punishment for vigilantism rather than improved police service quality as the likely mechanism. The broader implication is that citizens' willingness to cooperate with capable state institutions need not reflect satisfaction with state services. Such cooperation can also be due to the state's ability to limit citizens' choices by ruling out informal alternatives like vigilantism.

*With sincere thanks to MeMeZa Shout Crime Prevention, especially Thuli Mthethwa, Elmarie Pereira and Herman de Jager, who have been fantastic implementing partners. Thank you also to the Abdul Latif Jameel Poverty Action Lab (J-PAL) that provided funding for exploratory fieldwork. I am especially grateful to Donald P. Green for his support of this project, both intellectually and financially. A huge thank you to all the hard-working enumerators, and especially to the talented Itumeleng Motshegoa for her excellent work as field manager. Special thanks go to Isadora Amaral for excellent support during the endline survey. Thank you also to Macartan Humphreys for his advice and feedback. For invaluable comments on earlier versions of this paper, my thanks also go to Thomas Leavitt, Georgiy Syunyaev, Tara Slough, David Stasavage, Kate Baldwin, Alexandra Hartman, Laura Paler, Ben Morse, Nicholas Rush Smith, John Huber, John Marshall, Alexandra Scacco, Morgan Wack as well as participants of the 2020 session of the Contemporary African Political Economy Research Seminar, the 14th session of the Northeast Workshop in Empirical Political Science, the Harvard Workshop on the Political Science of Lynching in Global Comparative Perspective, and the WGAPE 2020 Annual Meeting. Finally, I would like to thank Citizen Surveys for including some of my survey questions on vigilantism in the questionnaire of their 2018 nationally representative opinion survey in South Africa. This project received IRB approval from Columbia University (protocol AAAR6346). Pre-analysis plans and addenda can be found at https://osf.io/87u4f.

†Washington University in St. Louis, wanna@wustl.edu

Maintaining order is a core task of governments. Yet, citizens often use mechanisms other than the state's justice system to deal with crime. In many contexts, spontaneously formed groups of ordinary citizens physically "punish" criminal suspects. Such mob vigilantism has been documented across almost all regions of the world (Jung and Cohen, 2020). This tendency to bypass state institutions has severe downsides. First, vigilante mobs commit gruesome assaults, often in response to minor offenses. Second, police and courts rely on information provided by citizens to function effectively (Tyler and Huo, 2002). Reluctance to cooperate with these institutions may undermine their effectiveness.

Many attribute the popularity of informal alternatives to the state's justice system to state weakness (e.g., Baker, 2002; Tankebe, 2009). As institutions like police increase in capacity, they are expected to supersede informal alternatives like vigilantism. This perspective is widely embraced by policy makers and has informed major state-building initiatives (ICG, 2022; Lake, 2010). Yet, up until recently (Acemoglu et al., 2020; Blair, Karim, and Morse, 2019; Nussio and Clayton, 2023), there has been little evidence to support these claims. This paper investigates the relationship between police capacity and vigilantism, both theoretically and empirically, through a field experiment in South Africa.

South Africa has one of the highest rates of vigilantism worldwide (Jung and Cohen, 2020). This situation is often attributed to poor police performance (Smith, 2019, chap. 1), even though South Africa falls into the upper half of the distribution of police capacity across Sub-Saharan Africa (UNODC, 2015; SAPS, 2022). Others cite a history of violence and strained citizen-police relations under the Apartheid regime (e.g., Super, 2022). These legacies make South Africa a hard case for the hypothesis that vigilantism can be addressed by strengthening contemporaneous state institutions. If doing so discourages vigilantism in a context where community punishments and distrust in the state are historically entrenched, the same may hold where vigilantism arises from more recent concerns.

Throughout history, state capacity was shaped by technological innovations such as cadastral maps (Scott, 1998), the telegraph (Martland, 2014), facial recognition software (Xu,

2021) and biometric identification systems (Muralidharan, Niehaus, and Sukhtankar, 2016). These technologies have expanded the state's reach by helping to identify and locate citizens or, as Scott (1998) puts it, by making them "legible" to state agents. This study leverages a similar shift, though on a much smaller scale. Together with a South African non-profit organization, I randomly assigned 100 of 250 sampled households to receive a police alarm system. The alarm is installed in the home and can be triggered using a panic button or cellphone. When activated, the alarm sends text messages with owners' names, contact details and location to the police. This information is also on file at the local police station.

The alarm was designed to address challenges that are common to South African townships and slums in other parts of the world. In such contexts, police tend to face a confusing street layout and a lack of street names and lights. Many citizens believe police would never arrive when called to a crime scene and are hesitant to rely on them. The alarm seeks to improve police's familiarity with and ability to locate households. Involvement in the alarm project may also be interpreted as a signal of the police's general willingness to perform.

The main finding of this article is that the police alarm encouraged cooperation with police and discouraged vigilantism. I measure outcomes through mid- ($N = 483$) and endline ($N = 448$) surveys conducted, respectively, one and eight months after treatment roll-out. Respondents in the treatment group appear more inclined to reach out to police and less willing to resort to vigilantism, especially among pre-registered subgroups who were a priori pessimistic about police. Hence, my results support the widespread intuition that expanding the reach of state institutions can reduce the popularity of informal alternatives and increase cooperation with the state.

The second goal of this article is to explore the mechanisms behind the alarm's effects. Doing so is particularly important because the nascent literature on state and non-state justice institutions has already produced mixed results. A small number of experimental studies including this one find strengthening the state discourages reliance on informal alternatives (Acemoglu et al., 2020; Blair, Karim, and Morse, 2019), while others suggest expanding the

state's reach can strengthen non-state actors (Cooper, 2019). Without an understanding of mechanisms, it is difficult to disentangle what drives such divergent findings.

I argue there are two main links between police capacity and vigilantism. The first stems from the logic of competition between service providers. Vigilante mobs apprehend and punish law-breakers. Hence, vigilantism produces services that resemble those provided by the state. More capable police may provide higher quality services. For example, police who can rely on information provided through the alarm may be able to quickly find a household which increases the likelihood of a successful arrest. Citizens who expect high quality police services may voluntarily substitute reliance on police for vigilantism.

A second link arises from the logic of regulation. Vigilantism is also a crime and more capable police may be better able to ensure that perpetrators of vigilantism go to prison. For example, police may use their familiarity with alarm-protected households to identify household members in a mob situation. The police logo on the alarm console provides a daily reminder to owners that their contact information is on file at the police station. Coupled with the knowledge that police are motivated enough to participate in the alarm project, this change may make alarm owners weary about resorting to vigilantism.

I use three empirical strategies to disentangle these mechanisms. First, I provide evidence that respondents assigned to an alarm developed both a more positive view of police service quality and a greater expectation of state punishment for vigilantism. Second, I report results from an additional information experiment that helps elucidate the relative importance of these changes. Finally, I leverage theoretical predictions about how the alarm and information treatments should interact if either mechanism is at play. Taken together, the results point to increased fear of state punishment as the likely link between the alarm and vigilantism. I provide evidence against several alternative explanations including the worry that results may be driven by experimenter demand.

These findings shed new light on the mechanisms that link state capacity to informal alternatives like vigilantism. The existing literature argues that citizens, when faced with

multiple providers of enforcement services, rely on whichever provider yields the best outcome at lowest cost (Sandefur and Siddiqi, 2012). In these accounts, state capacity allows the state to expand to unserved areas (e.g., Cooper, 2019) or to make state services compare more favorably with informal ones (e.g., Sandefur and Siddiqi, 2012). Hence, the state is treated as one of many entities competing for citizens' demand in a market for enforcement services.

The evidence presented here demonstrates the relationship between the state and informal actors is not limited to competition. There is a second mechanism through which capable state institutions may weaken informal alternatives. Where the state has declared an informal alternative illegal, an increase in police presence may induce citizens to abandon this alternative out of concern about state punishment. Instead of the demand, increased state presence can dampen the supply of informal responses to crime.

This insight also helps us understand the conditions under which state capacity may be most effective at weakening informal alternatives. Such alternatives do not always fall outside the law. Rural traditional courts and chiefs, for example, are often recognized by the state (Cooper, 2019; Baldwin, 2016). If state capacity affects the choice between state and non-state alternatives primarily by increasing the risk of state punishment, it seems intuitive that a stronger state does not discourage reliance on actors that the state recognizes as legitimate. Henn (2022) indeed finds state capacity weakens the role of chiefs only if they are not integrated into a country's legal framework.

The broader implication is that state capacity can have downsides for citizens. Recent work on state capacity, including studies of legibility initiatives like biometric identification systems (Muralidharan, Niehaus, and Sukhtankar, 2016; Bossuroy, Delavallade, and Pons, 2019), highlights the benefits of more effective service delivery. Yet, even democratic states use their capacity not only for service delivery but also to regulate behavior. Vigilantism is one example of an illegal practice that enjoys widespread support. Others include electricity theft, unlicensed street vending, and tax evasion. Where such activities are common, citizens may perceive state capacity as a double-edged sword.

# 1 Police Capacity and Mob Vigilantism

Sanctioning law breakers is a core task of states. Yet, formal justice institutions often co-exist with informal enforcement mechanisms. These sometimes consist of non-state actors like traditional authorities or gangs (Baker, 2008). I focus here on settings where informal punishments are meted out by ordinary community members. A report from South Africa describes an example:

> We heard a woman screaming (...) My bag! My Bag! Here's a thief!. In no time, (...) everybody was coming out (...). Then they descended upon this man – they came with all sorts of weapons to assault him. Rocks on the street were thrown at him. In no time, the man was gone (...) in a matter of a few minutes, perhaps seconds, a man is dead, killed by a group of people in my community for snatching a woman's handbag on her way to work. (Khayelitsha Commission, 2014, p.342)

Such accounts are abundant throughout the developing world (Smith, 2004; Adinkrah, 2005; Schuberth, 2013; Kirsch and Grätz, 2010). When asked why they support mob vigilantism, respondents in qualitative interviews often point to the police. One South African respondent said: "It is not a good thing to take the law into your own hands, but since the police is [*sic*] not doing a good job, people have no other option." How would citizens' views change if police became more capable?

Classic theories of crime control presume citizens wish for some transgressions to be punished, be it due to a desire for deterrence or vengeance (Becker, 2000). Imagine a citizen who has information about an offense and decides between reporting to the police and rallying her family, friends, and neighbors. Qualitative evidence from South Africa suggests four considerations that may affect her choice.

*Nature of offense.* State and community punishment may not both be viable options. The demand for punishment of witches is high in many contexts but the state does not typically punish witchcraft (Smith, 2019; Miguel, 2005). Conversely, vigilante mobs rarely attack perpetrators whose coercive capacity far outweighs that of communities. Individuals

linked to the drug trade, for example, tend to be heavily armed. Nonetheless, many offenses are addressed through both mechanisms, including minor crimes such as petty theft and burglary as well as violent crimes like murder and rape.

*Expected punishment of suspect.* Vigilante punishments tend to be harsher than state sentences (García-Ponce, Young, and Zeitzoff, 2022). The contrast is particularly stark for petty crimes that can end in grave injury or death when addressed by vigilante mobs. Community punishments also attract crowds of spectators and are more public than state sanctions. Many South Africans cite its harsh and public nature as an advantage of vigilantism but others as a drawback.

*Probability that suspect will be punished.* A precondition for punishment is that perpetrators are apprehended and evidence is collected to meet whatever standard is used to determine guilt. Community members are well positioned to apprehend suspects, because they are spatially proximate to the crime scene and well informed. Communities also often mete out punishment even if evidence of guilt is tenuous. On the state's side, apprehending perpetrators and investigating crimes are classic tasks of police. As police increase in capacity, they may be better able to provide these services.

*Risk of punishment for vigilantism.* Mob vigilantism amounts to serious crimes like assault or murder. Victims thus have access to legal recourse. A victim in the study precinct opened an assault case against his attackers, and a group of men received lengthy prison sentences for killing two suspected thieves. As police forces increase in capacity they may become more effective at investigating vigilantism. Such an increase in the risk of state punishment may be most relevant to the decision to actively inflict vigilante violence. The woman who screamed "Here's a thief!" in the anecdote above, for example, may not be held accountable if she did not participate in the assault. Yet, perpetrators – often but not exclusively men – tend to be the instigator's immediate community. Even the decision to encourage vigilantism may thus be affected by an increase in the risk that one's husbands, sons, and brothers could be arrested.

This discussion suggests a stronger police force may affect the choice between reliance on the state and vigilantism through two mechanisms depicted in Figure 1:

1. *Improved Police Service Quality:* An increase in police capacity may encourage cooperation with police and discourage vigilantism by increasing the probability that perpetrators of crime who are reported to police are sanctioned by the state.

2. *Increased Risk of State Punishment:* An increase in police capacity may encourage cooperation with police and discourage vigilantism by increasing the probability that participation in mob vigilantism leads to state punishment.

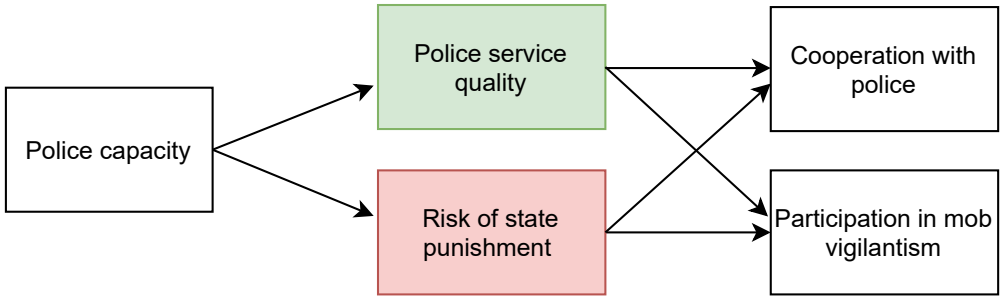


**Figure 1:** Police capacity and the choice between the state and mob vigilantism.

While not mutually exclusive, these mechanisms are qualitatively distinct. The first centers on the state becoming more attractive and the second on vigilantism becoming more costly. Citizens who oppose vigilantism may of course perceive police efforts to counter it as part of police service delivery. From the perspective of these citizens, the distinction between vigilantism and other kinds of crimes is arbitrary. Yet, the focus here is on citizens who see vigilantism as a legitimate option. Such citizens tend to perceive a stark difference between vigilantism and other crimes. Almost half of control group respondents in this study oppose prison sentences even for vigilantes who killed a criminal suspect.

The two mechanisms have drastically different implications for the relationship between such citizens and the state. The first mechanism suggests state capacity helps the state *outcompete* vigilantism. According to this logic, citizens choose to rely on a strong state, because

they perceive it as the best option. Hence, citizens should welcome a strong state, even if they currently rely on vigilantism. This perspective dominates the literature on state and non-state enforcement (see Jaffrey, 2023, for an exception), and relies on the assumption that citizens perceive state and vigilante justice as substitutes. The second mechanism implies state capacity helps the state *rule out* informal alternatives that it deems illegal. Rather than voluntarily, citizens may cooperate with the state because it has limited their options. In fact, if citizens prefer gruesome public punishments without regard for due process protections, the state may not be able to out-compete vigilante mobs simply by becoming more effective.

## 2 Experimental Design

### 2.1 Context

South Africa has one of the highest crime rates worldwide. Crime is particularly prevalent in townships, which are racially segregated areas at the outskirts of cities that were created under the Apartheid regime. This study takes place in a semi-urban, predominantly black, low-income township in the Northwest Province. The implementing partner and police selected this precinct because of its high rate of burglaries and house robberies, which the intervention may help address. At baseline, roughly 44% of households report a member experienced a crime over the past year.

Scott (1998) famously argued communities have to be "legible" to state agents for state institutions to function effectively. Townships were once designed to be easily policed, but tend to be difficult to "read" today. Many have grown substantially through the expansion of informal settlements. The result is a confusing layout and address system. Street names are rare and houses are numbered within sections comprising thousands of houses. The study precinct uses three different numbering systems and numbers can be out of sequence even within one system. Street lighting is sparse. These conditions complicate the work of police, especially in terms of emergency response.

Dissatisfaction with police is widespread. Figure 2 shows slow response times are by

far the most prevalent grievance. Other complaints include that one's area is not served at all and that police are corrupt. Opinions in the study precinct mirror these countrywide concerns. For example, 55% of control group respondents believe police would never come or take longer than two hours when called to an emergency.

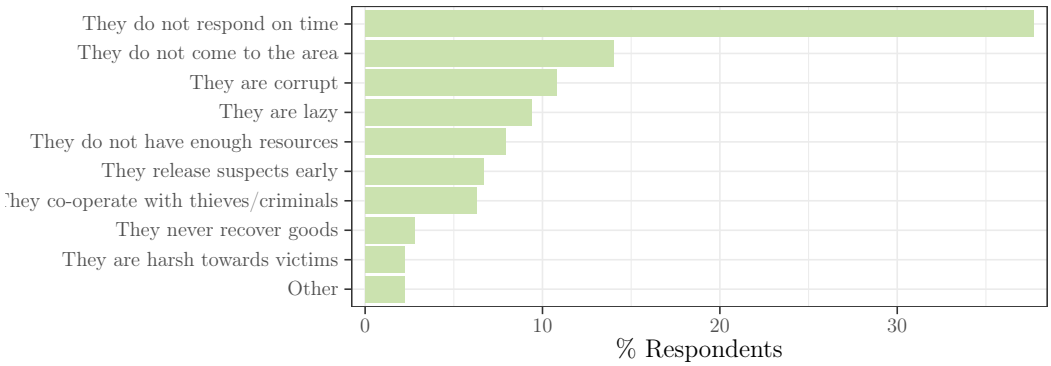


**Figure 2:** Slow response is main reason for dissatisfaction with police ($N = 8,906$)
Calculated among 43% of respondents who are dissatisfied with police. Question asked about the main reason for respondent's dissatisfaction. Source: StatsSA (2016/2017).

Mob vigilantism is the primary alternative to the state's justice system in the study precinct. Many households own a whistle to summon the community in emergencies. Figure 4 in the appendix plots data from the endline survey which asked respondents how many vigilante incidents happened in their area between May and July 2018. At least a quarter of respondents in most areas recalled one incident or more. In qualitative interviews, most respondents could describe at least one case. Anecdotes involved accusations of burglary, theft, or sexual violence that led to severe injuries or even the death of the accused.

The approach of police to vigilantism is ambiguous. There is no shortage of anecdotes about police turning a blind eye, but arrests occur. Table 1 displays the joint distribution of baseline perceptions of police service quality and the risk of state punishment for vigilantism. Around 16% of respondents have above median service quality expectations but do not think vigilantism perpetrators would be arrested. Conversely, almost one third of respondents consider legal repercussions for vigilantism likely but have little hope that police provide

high quality services. In qualitative interviews, respondents in the latter group complained that police do little for communities but intervene when communities protect themselves.

| | Police service quality | | |
|---|---|---|---|
| **Risk of punishment MV** | **Low** | **High** | **Total** |
| **High** | 27.2% | 26.0% | 53.2% |
| **Low** | 31.2% | 15.6% | 46.8% |
| **Total** | 58.4% | 41.6% | $N =$ 250 |

**Table 1:** Baseline perceptions of police outputs

Sample includes one woman per household. Percentages indicate shares of respondents. Respondents who perceive a low (high) punishment risk thought it "not very likely" ("somewhat likely") or "not likely at all" ("very likely") that perpetrators of vigilantism would be arrested. Service quality perceptions are measured through an index of *Customer Service*, *Arrive quickly* and *Send guilty to prison*. Respondents with high (low) service quality perceptions fall above (below) the sample median of the index. See appendix section D.5 for question wording.

### 2.2 Intervention

The police alarm system was developed by a South African non-profit organization together with South African police. The alarm is an electronic device installed in the house that can be triggered via a panic button, motion sensor or cell phone.[1] The alarm sends text messages to personnel at the closest police station including the station management, officers on duty and the Community Policing Forum (CPF) – volunteers that liaise between police and community.[2] Text messages indicate alarm owners' names, phone numbers, and landmarks close to the respective home. This information is also on file at the police station. The alarm can be triggered silently or such that a light flashes and a siren sounds outside the house.

[1]Households without electricity received a solar panel that only powers the alarm and the alarm is equipped with a 24-hour battery.

[2]Alarm owners can nominate two neighbors to receive text messages. Surveys with study households and neighbors provide no evidence that the alarm changed community relations.

The non-profit organization registers all alarm panics through a back-end system.

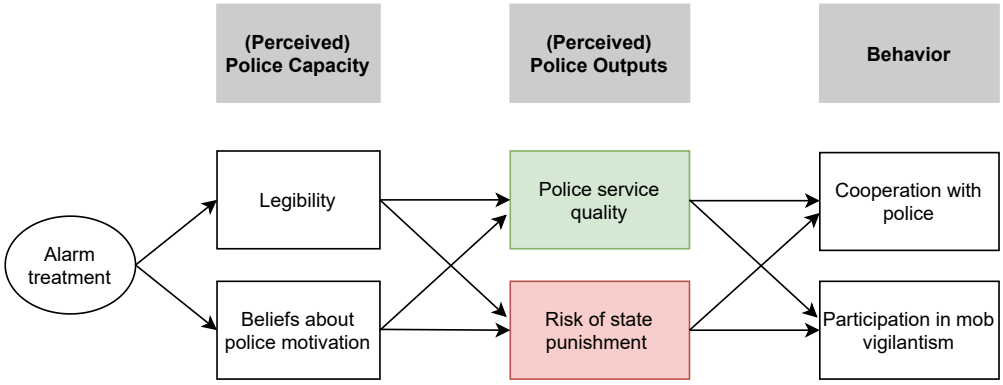


**Figure 3:** Hypothesized effects of police alarm.

Figure 3 depicts how the alarm may alter the de facto or perceived capacity of police to intervene in a household. First, the alarm aims to make households more "legible" to police, i.e., easier to locate and identify. This change applies most obviously to households who trigger their alarm. Alerting police without an alarm requires calling a centralized emergency hotline or police station.[3] Without reliable addresses, it can be difficult to explain one's location to a call center agent unfamiliar with the specific township. The alarm sends location details directly to local police, flashes a light and sounds a siren. Moreover, with only 100 alarms in a precinct with more than 42,000 residents, many officers were able to find protected households from memory alone. Alarm recipients may thus expect better police services in the form of a faster response, which may increase the chance that perpetrators are apprehended and convicted.

Being known to police is beneficial if one reaches out for help, but may cause worry if one intends to break the law. Many study participants reside in informal settlements and are thus not registered with local authorities. Names and contact details of alarm owners, however, are on file at the police station. In fact, the alarm console shows the police logo, providing a daily reminder that police have a record of the household. The effect of this change may to some degree be psychological. In practical terms, police often described the

[3]Calls to the hotline from a landline are free.

fact that vigilantism draws an "anomymous crowd" that is unwilling to testify as a major difficulty for police investigations. Being known to police may make it easier for police to identify household members in a mob situation. Hence, alarm recipients may worry about the legal consequences of vigilantism.

Second, alarm owners may update their more general views about police. The alarm project likely has differential costs and benefits for different "types" of police. The alarm facilitates access to police and provides information about incidents to an outside party. Unmotivated police may perceive these changes as a nuisance,[4] but highly motivated police may welcome them as a way to improve police performance. Police involvement in the project may thus signal to alarm owners that police are motivated to perform.[5] Learning may also occur through interactions with police that result from the alarm treatment. Alarm recipients who believe police are highly motivated may again become both optimistic about the services police provide and worried about the legal consequences of vigilantism.

Hence, the alarm may affect the choice between formal and informal crime control through the two logics outlined above. The belief that police are able to respond fast and are highly motivated may make reliance on the state more attractive. Importantly, the alarm is well set up to address burglaries, one of the most common triggers of vigilantism in the study precinct. On the other hand, being "known" to police who take their job seriously may also increase the fear of state punishment for mob vigilantism.

Finally, the alarm may cause surprises among citizens with low expectations, while reaffirming beliefs or disappointing among those who already expect a lot from police. Thus, I pre-registered sub-group analyses by prior beliefs about the two relevant mediators, police service quality and the risk of state punishment for mob vigilantism.

[4]Police sometimes voiced concern about the alarm creating too much "demand."

[5]Other residents who learn about the alarm project could draw the same inference. Yet, surveys with neighbors showed no evidence of effects on surrounding households.

### 2.3 Household sampling and baseline survey

The 250 study households were sampled during a baseline survey between May and July 2018. 135 were chosen from a list of vulnerable homes provided by the police, which is how the implementing partner usually selects beneficiaries. In practice, the CPF was heavily involved in creating the list and added regular attendees of crime-related community meetings. The remaining 115 households were chosen from a pool of households created by geo-locating every tenth house in the precinct's eleven most high crime areas.

Households were selected from these pools in non-random ways to limit non-compliance, attrition and interference. To limit interference, I used a stochastic algorithm to select the largest sample such that each household is located no closer than 150 m to all other sampled households. Due to location inaccuracies, only 67% of the sample satisfies this constraint. To limit non-compliance, the sample excludes 27 households who indicated at baseline that they are not interested in an alarm.[6] To limit attrition, the sample excludes 77 households that were interviewed at baseline but could not be reached during subsequent back-checks. Appendix section A.4 provides more details on sampling.

Figure 5 in the appendix shows the sample's geographic layout. There is spatial overlap across households sampled through, respectively, the police and the geo-location exercise. The police sample is more dispersed, because the geo-location exercise was concentrated in the most crime-ridden areas. In each household, the baseline survey interviewed the woman most involved in household decisions.[7] Appendix section A.6 shows respondents sampled

[6]Uninterested households were particularly pessimistic about police. Since I find treatment effects are concentrated among respondents who are a priori pessimistic, the exclusion of uninterested respondents should, if anything, bias effect estimates downwards.

[7]This rule ensured respondents were able to confirm their household's interest in the alarm. The implementing partner was most interested in the views of women and budget constraints prevented me from surveying two household members at baseline. Mid- and endline surveys interview a woman and a man in each household where available.

through the police expected more from police at baseline and had a greater willingness to turn to police than those sampled through the geo-location exercise.

### 2.4 Random assignment

Households were organized into 50 blocks of 5. I first divided the sample into two sets by how households were sampled. Within each set, blocks were formed to minimize the within-block multivariate Mahalanobis distance of four variables: baseline support for and the willingness to participate in vigilantism and the household's latitude and longitude.[8] 100 households, two in each block, were assigned to the alarm treatment. Alarm installations took place in September and October 2018.

### 2.5 Treatment take-up and compliance

Only 27 of 358 baseline respondents were not interested in the alarm and hence excluded from the study prior to random assignment. At midline, 93 of 100 households in the treatment group and none in the control group had received an alarm. Among the seven households that did not comply with their assigned treatment, four refused the alarm, one dismantled it after installation and two remained unprotected due to administrative errors. Of the latter two, one received an alarm before the endline interview.

The widespread interest in the alarm and the high compliance rate may seem at odds with the argument that the alarm increases the fear of state punishment for vigilantism. If many respondents support vigilantism, why would they agree to a treatment that supplies their personal details to the police? One explanation is that respondents trade off the downsides of increased police supervision against the promise of improved service delivery. Vigilantism is not a panacea for all crime. Gun violence is common in the study precinct. Many may be willing to limit their freedom to resort to vigilantism in exchange for improved police

[8]Blocking on the sampling strategy and geo-locations ensured buy-in from the police and the CPF, who wanted alarms to be spread across the precinct and for enough alarms to go to police list households.

protection against heavily armed perpetrators that the community is unwilling to confront.

How citizens resolve this trade-off may depend on their beliefs about police services and taste for vigilantism. Figure 7 in the appendix plots responses from all baseline respondents, including those who were excluded from the study sample. The figure compares respondents interested in the alarm to those who refused it at baseline or a later stage.[9] Leaving aside that few households refused the alarm ($N = 31$), the plot suggests these groups differ in intuitive ways. For example, those not interested in an alarm are 45% more likely to participate in vigilantism and less optimistic about the police's emergency response.

Figure 8 in the appendix shows how treated households used the alarm. The implementing partner registered 159 alarm panics between 1 November 2018 and the endline in mid-June 2019.[10] 72 of the 94 protected households triggered their alarm at least once. Interestingly, only 15 households in the treatment group reported having experienced a crime since the previous Christmas at endline. Some panics may thus reflect false alarms, or emergencies other than crime. For example, one household triggered an alarm because the neighboring house was on fire. Panics can also result from maintenance procedures. Importantly, even panics unrelated to crime can yield a police response. For example, one respondent was surprised to find police outside her door after her child accidentally triggered the alarm.

### 2.6 Outcome measurement

I measure outcomes using two waves of household surveys that took place, respectively, one and eight months after treatment roll-out. The same two respondents were interviewed in each household at midline and endline: the woman sampled at baseline and one randomly selected adult man. In all-women households, a second woman was selected at random. Since 23 of 250 households have only one member, the target sample size was $N = 477$.

[9]This comparison should be interpreted with care, because it is unknown whether households assigned to the control group may have refused the alarm had they been assigned to treatment.

[10]Multiple panics in the same household on the same day are collapsed into one incident.

Response rates were 92% ($N = 438$) at midline and 85% ($N = 407$) at endline. Appendix section B.2 shows rates and patterns of attrition seem unaffected by treatment.

Additional respondents were interviewed if other household members were available during the interview. 45 respondents were added across 39 households at midline and 39 across 38 households at endline. Appendix section B.3 shows there is no statistically significant relationship between treatment and the number of additional respondents. As pre-registered, my analyses include all respondents, but all main results are robust to the exclusion of additional respondents. See appendix section B.1 for evidence of covariate balance.

Missing values due to non-response to outcome questions are imputed using multivariate imputation via chained equations. Outcomes are imputed within pre-specified families (e.g., "vigilantism related outcomes"). The procedure does not condition on treatment status or covariates. Outcome measures range from zero to one. Indices are created by averaging constituent items after imputation. Appendix section D provides the question wording.

### 2.7 Estimation and hypothesis tests

I estimate sample intent-to-treat (ITT) effects using the following regression specification:

$$\boldsymbol{Y} = \alpha + \tau \boldsymbol{z} + \delta \boldsymbol{n} + \boldsymbol{\epsilon}.$$

$\boldsymbol{Y}$ here is a vector of outcomes, $\alpha$ an intercept, $\tau$ the sample ITT, $\boldsymbol{z}$ a vector of treatment assignments, $\boldsymbol{n}$ a vector storing the number of respondents per household with associated coefficient $\delta$, and $\boldsymbol{\epsilon}$ a vector of error terms that allow for clustering at the household level. I control for the number of respondents per household, since estimates of the sample ITT may be biased if cluster size correlates with potential outcomes.

To estimate conditional ITTs and differences between them, I add to this regression an indicator for high prior beliefs about either police service quality or the risk of state punishment for vigilantism at baseline and the interaction with the treatment assignment indicator. One respondent was interviewed per household at baseline and their response is

interpreted as a household-level measure of prior beliefs. See appendix section D.5 for the question wording and Table 1 for the distribution of prior belief measures.

I also pre-registered a specification that controls for covariates selected through lasso regression. Robustness of main results to this specification is shown in appendix section C.2. $p$-values are calculated via randomization inference by permuting treatment assignment 2000 times to simulate the sampling distribution under the sharp null hypothesis of no (positive or negative) treatment effect for any unit. Appendix section A.1 summarizes divergences from the pre-analysis plan.

### 2.8 Ethics

I discuss several ethical considerations and my efforts to address them. A first question is whether the alarm could produce adverse effects. For example, one may worry the siren could be used to instigate mob vigilantism. However, the alarm always sends messages to police. Hence, the implementing partner who had already installed almost 2,000 alarms throughout South Africa considered this scenario unlikely. To guard against adverse consequences for recipients, households were given detailed information about and ample opportunity to refuse the alarm. Moreover, this study neither increased nor decreased the number of alarms. The implementing partner had funding for exactly 100 alarms that would have been installed irrespective of this study.

Turning to data collection, one risk was the re-traumatization of respondents. The questionnaire was vetted extensively through pretesting and discussions with the local research team. Questions about crime victimization focused on household-level rather than individual experiences and did not ask details about crimes. Women respondents were matched to women enumerators and interviews were conducted in private. To avoid implicating respondents in illegal behavior, I did not ask about actual participation in vigilantism. Instead, questions focused on hypothetical scenarios or respondents' recollection of incidents.

Finally, several measures were taken to protect study staff. Enumerators were residents of the study community but worked in sections other than their own. Approvals were obtained

from community leaders and police were aware of all survey activities. Where communities seemed hostile, enumeration was stopped until problems were solved with local authorities. Enumerators worked in pairs and were picked up from households by car. Walking was kept to a minimum and enumerators received bags to avoid carrying equipment like tablets in the open. Wherever possible, enumeration stopped before nightfall. If enumerators worked after dark, a car was kept close and they were brought home afterwards.

## 3 Main Results

Table 2 shows the alarm treatment seems to have increased citizens' willingness to rely on police and decreased their willingnes to resort to mob vigilantism, especially among respondents who expected little from police at baseline.

I measure respondents' willingness to rely on police through an index with two components that capture, respectively, respondents' inclination to alert police if someone is trying to enter their home and their general proclivity to share crime-related information with police. Column 1 suggests the alarm increased the willingness to rely on police at midline by roughly one third of a control group standard deviation ($p < 0.01$). The estimated effect at endline is slightly smaller but highly statistically significant ($p < 0.01$).

Analyses in columns 5 and 7 allow effects to vary across prior beliefs. Both columns indicate effects on the willingness to rely on police are concentrated among households with low baseline expectations. Among those with low prior beliefs about, respectively, the risk of being arrested for vigilantism and police service delivery, the alarm seems to have increased the willingness to rely on police at endline by roughly one third of a control group standard deviation. The interaction terms suggest effects among high prior groups are statistically significantly smaller and close to zero ($p < 0.05$ and $p < 0.1$).

I measure the willingness to resort to mob vigilantism using one item at midline and an index of the same and another item at endline. Both measures ask respondents what they would do if the community apprehended a suspect. One item involves three ordered

options: advocate for handing the suspect over to police, let others beat the suspect but do not participate, or personally participate in beating the suspect. The other item only asks whether respondents would personally inflict harm.

Column 3 suggests the alarm decreased the willingness to participate in vigilantism at midline by roughly one fifth of a control group standard deviation ($p < 0.01$). The endline estimate is substantially smaller and falls short of statistical significance.[11] Once I allow for effect heterogeneity across prior beliefs, however, I find evidence of a negative effect even at endline. Column 6 suggests the alarm decreased the willingness to participate in vigilantism among those with low prior beliefs about the risk of being punished for doing so by about one third of a control group standard deviation ($p < 0.05$). This decrease is of similar magnitude as the increase in the willingness to rely on police among this subgroup. The interaction term indicates the effect is statistically significantly less negative among the corresponding high prior group ($p < 0.01$). Prior beliefs about service delivery appear to condition effects in similar ways, but the patterns are less pronounced (see column 8).

In the appendix, I investigate whether these results generalize to other, related outcomes. Table 13 shows the alarm does not appear to affect whether respondents recently spoke to police. This finding is surprising given the estimated increase in respondents' willingness to rely on police. A possible explanation is that the alarm acts as a deterrent of household-level crime, thereby reducing the need to reach out to police. Respondents in the treatment group indeed feel safer, but the evidence of a reduction in victimization is limited (see Table 25).

[11] Table 18 in the appendix shows the same holds for both index components.

| | Rely police | | Join MV | | Rely police | Join MV | Rely police | Join MV |
|---|---|---|---|---|---|---|---|---|
| | Midline | Endline | Midline | Endline | Endline | Endline | Endline | Endline |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Alarm | 0.097*** | 0.075*** | −0.078*** | −0.012 | 0.132*** | −0.100** | 0.126*** | −0.042 |
| | (0.028) | (0.031) | (0.032) | (0.028) | (0.044) | (0.044) | (0.042) | (0.037) |
| Alarm × High Prior Punishment | | | | | −0.108** | 0.158*** | | |
| | | | | | (0.062) | (0.057) | | |
| Alarm × High Prior Service | | | | | | | −0.104* | 0.066 |
| | | | | | | | (0.061) | (0.057) |
| Control Mean | 0.6 | 0.64 | 0.24 | 0.17 | 0.64 | 0.17 | 0.64 | 0.17 |
| Control SD | 0.31 | 0.31 | 0.37 | 0.29 | 0.31 | 0.29 | 0.31 | 0.29 |
| RI p-value Main | 0 | 0.003 | 0.006 | 0.344 | 0.002 | 0.011 | 0.001 | 0.148 |
| Hypothesis Main | upr | upr | lwr | lwr | upr | lwr | upr | lwr |
| RI $p$-value Diff. | - | - | - | - | 0.034 | 0.002 | 0.061 | 0.206 |
| Hypothesis Diff | - | - | - | - | lwr | upr | lwr | upr |
| Number HHs | 245 | 237 | 245 | 237 | 237 | 237 | 237 | 237 |
| Observations | 483 | 448 | 483 | 448 | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 2:** Effects of alarm on willingness to rely on police and resort to mob vigilantism.

Outcomes range from 0 to 1. Appendix section D.5 contains details on measures of prior beliefs and table 1 shows their distribution. Appendix section A.2 provides details on model specification and testing, and appendix section D.1 on outcome question wording.

Table 15 in the appendix suggests the alarm reduced not only the willingness to participate in vigilantism but also support for the participation of others – at least at midline and among those with low prior beliefs about the risk of state punishment. That said, alarm owners do not appear less inclined to reach out to neighbors in case of an attack. This pattern provides a first clue regarding mechanisms. Fear of state punishment should matter most for outcomes related to violence. Merely alerting one's neighbors is not a crime. Hence, if the alarm's effects result from increased concerns about state punishment, the apparent absence of an effect on this last outcome seems intuitive. If the alarm worked by convincing citizens that police provide high quality services, one would expect reduced demand for *all* kinds of community involvement.

## 4 Mechanisms

I now explore further *how* the alarm may have encouraged reliance on police and discouraged vigilantism. Questions about mediation are notoriously difficult to answer. I provide evidence using three empirical strategies each of which has its own advantages and limitations.

### 4.1 Effects of the alarm on perceptions of police capacity and outputs

First, I analyze effects of the alarm on potential mediators. The experiment allows me to identify these effects without additional assumptions. The main text presents estimates of effects at endline. Because the main treatment effects appear to be concentrated among households who were a priori pessimistic about police, I estimate conditional effects on mediators as well. Estimates of unconditional effects at mid- and endline are reported in online appendix section C.4.

Table 3 presents evidence on the first step in the causal chain in Figure 3, the alarm's effects on perceptions of legibility and police motivation. This analysis can be thought of as a manipulation check. Did the alarm indeed affect perceptions of police capacity?

Columns 1 and 3 suggest the alarm increased respondents' sense of being legible to police. The outcome combines two items that ask whether local police know, respectively,

the respondent's house, and the name of a household member. Column 1 shows an upward shift in this outcome by around one fourth of a control group standard deviation ($p < 0.05$). This effect does not seem to vary with prior beliefs about the risk of state punishment for vigilantism. The estimated effect among respondents with low prior beliefs about service delivery corresponds to a little less than one fifth of a control group standard deviation ($p < 0.1$). The corresponding interaction term, while estimated imprecisely, suggests a larger effect in the high prior group.

| | Police... | | | |
|---|---|---|---|---|
| | know HH | are motivated | know HH | are motivated |
| | (1) | (2) | (3) | (4) |
| Alarm | 0.114** | 0.216*** | 0.077* | 0.186*** |
| | (0.068) | (0.068) | (0.061) | (0.063) |
| Alarm × High Prior Punishment | −0.00003 | −0.179** | | |
| | (0.091) | (0.095) | | |
| Alarm × High Prior Service | | | 0.093 | −0.109 |
| | | | (0.090) | (0.094) |
| Control Mean | 0.44 | 0.47 | 0.44 | 0.47 |
| Control SD | 0.45 | 0.5 | 0.45 | 0.5 |
| RI p-value Main | 0.05 | 0.001 | 0.076 | 0.001 |
| Hypothesis Main | upr | upr | upr | upr |
| RI $p$-value Diff. | 0.476 | 0.029 | 0.744 | 0.146 |
| Hypothesis Diff | lwr | lwr | lwr | lwr |
| Number HHs | 237 | 237 | 237 | 237 |
| Observations | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 3:** Effects of alarm on perceptions of legibility and police motivation.
Outcomes range from 0 to 1. Appendix section D.5 contains details on measures of prior beliefs and table 1 shows their distribution. Appendix section A.2 provides details on model specification and testing, and appendix section D.2 on outcome question wording.

Columns 2 and 4 indicate a substantial improvement in perceptions of police motivation. 47% percent of control group respondents think it unlikely that a slow police response results from a lack of motivation. The alarm seems to increase this share by roughly 20 percentage

points among both low prior groups ($p < 0.01$). The interaction terms suggest effects among those with high prior beliefs are substantially smaller ($p < 0.05$ and $p < 0.2$).

In short, respondents in the treatment group are more likely to believe they are known to police and that police are motivated. The effect on legibility perceptions does not appear to vary consistently with prior beliefs, which is intuitive since police acquired information about all protected households. Positive shifts in beliefs about police motivation seem concentrated among those who expected little from police at baseline. Next, I turn to the potential downstream effects of these changes.

Did the alarm change perceptions of the risk of being punished for vigilantism and police service quality? Table 4 suggests the answer is yes. Columns 1 and 4 provide little evidence that the alarm changed perceptions of the speed with which police respond to vigilantism. Columns 2 and 5, however, indicate the alarm increased the share of respondents who think police send perpetrators of vigilantism to prison by roughly ten percentage points among both low prior groups ($p < 0.1$ and $p < 0.05$). The interaction terms suggest treatment effects on this outcome are close to zero among those with high prior expectations.[12]

Columns 3 and 6 show similar patterns for an index of police service quality perceptions. The alarm seems to increase this outcome by around one third of a control group standard deviation among low prior groups ($p < 0.05$), while estimated effects among high prior groups are statistically significantly smaller ($p < 0.1$). Table 20 in the appendix section shows these estimates reflect improvements in perceptions of both the speed with which police would respond when called to a respondent's home and the police's general inclination to send perpetrators of crime to prison.

[12]I also pre-specified an analysis of an index of perceptions of the risk that police would find out about illegal behaviors other than vigilantism. I find little evidence of an effect on these perceptions among low prior groups, but estimate a small upward shift in the sample as a whole (appendix sections C.1 and C.4).

| | *Risk of state punishment* | | *Service quality* | *Risk of state punishment* | | *Service quality* |
|---|---|---|---|---|---|---|
| | Respond MV | Imprison MV | Service index | Respond MV | Imprison MV | Service index |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Alarm | 0.043 | 0.091* | 0.085** | 0.023 | 0.110** | 0.085** |
| | (0.051) | (0.066) | (0.044) | (0.045) | (0.055) | (0.037) |
| Alarm × High Prior Punishment | −0.074 | −0.075 | −0.094* | | | |
| | (0.070) | (0.088) | (0.059) | | | |
| Alarm × High Prior Service | | | | −0.040 | −0.145** | −0.091* |
| | | | | (0.073) | (0.089) | (0.060) |
| Control Mean | 0.67 | 0.71 | 0.55 | 0.67 | 0.71 | 0.55 |
| Control SD | 0.36 | 0.46 | 0.3 | 0.36 | 0.46 | 0.3 |
| RI $p$-value Main | 0.235 | 0.07 | 0.034 | 0.262 | 0.032 | 0.021 |
| Hypothesis Main | upr | upr | upr | upr | upr | upr |
| RI $p$-value Diff. | 0.166 | 0.136 | 0.06 | 0.217 | 0.047 | 0.075 |
| Hypothesis Diff | lwr | lwr | lwr | lwr | lwr | lwr |
| Number HHs | 237 | 237 | 237 | 237 | 237 | 237 |
| Observations | 448 | 448 | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 4:** Effects of alarm on perceptions of risk of state punishment for mob vigilantism and of police service quality.
Outcomes range from 0 to 1. Appendix section D.5 contains details on measures of prior beliefs and Table 1 shows their distribution. Appendix section A.2 provides details on model specification and testing, and appendix section D.3 on outcome question wording.

Respondents assigned to an alarm thus expect more from police, in terms of both service quality and the risk of state punishment for vigilantism. Both effects appear concentrated among the low prior groups that also saw the greatest shifts in the willingness to rely on police and resort to vigilantism. Interestingly, prior beliefs about one police output seem to condition effects on perceptions of another. For example, respondents with low prior beliefs about police service quality see larger effects on perceptions of *both* service quality and the risk of punishment for vigilantism. A likely reason is that prior beliefs are not independent. Low expectations of one police output may reflect the view that police capacity is low and updating beliefs about capacity may increase expectations of various police outputs.

### 4.2 Effects of information treatments that vary perceptions of police outputs

So far, results are consistent with both hypothesized mechanisms. Note however that it is possible for the alarm to have an average effect on a potential mediator even if it does not mediate the relationship between the alarm and the willingness to rely on police or resort to vigilantism (Green, Ha, and Bullock, 2010). Moreover, both mechanisms may be at play but may not be equally important. To provide additional evidence on how the alarm produces its effects, I use two information treatments, each designed to change one mediator but not the other. These information treatments allow me to understand how effectively shifts in perceptions of service delivery and punishment risks discourage vigilantism. Provided that the alarm and information treatments produce changes in the mediators that are similar, the results shed light on how the alarm intervention is likely to work (Ludwig, Kling, and Mullainathan, 2011).

Both treatments consist of local news articles that were read to respondents during their endline interview (see appendix section A.7). To minimize the potential for simultaneous effects on both mediators, I chose articles that focus on specific police efforts rather than broad capabilities. The "Police fight MV" treatment describes police efforts to convict perpetrators of vigilantism. The "Police fight crime" treatment depicts police efforts to convict perpetrators of crimes against women and children, a service that is in high demand. More-

over, women and children are rarely attacked by vigilante mobs, which made it unlikely that this treatment would affect perceptions of the police's approach to vigilantism.

Respondents were randomly assigned to one, both or none of the treatments, creating a full factorial design. Treatments were randomized across the entire endline sample ($N = 815$) which includes two members of a neighboring household for each of the 250 study households.[13] As pre-registered, my analyses include respondents from main and neighboring households. Enumerators were unaware of the goal to understand effects on subsequent responses and thought the aim was to elicit opinions about the articles using several open-ended questions. Outcomes were measured later during the interview. Due to time constraints, I only measure perceptions of police efforts and the willingness to resort to vigilantism. Table 8 in the appendix shows the distribution of respondents across information treatments. I estimate the effects of each treatment by regressing outcomes on the corresponding assignment indicator, marginalizing across the other information and the alarm treatment.

I pre-specified that the information treatments would be analyzed within two low prior subgroups should they appear ineffective in the sample as a whole. Table 22 in the appendix suggests neither information treatment shifts respondents' expectations about police effort among all respondents. Analyses in Table 5 subset to respondents with low prior beliefs about the police's inclination to arrest vigilantism perpetrators.[14]

Column 1 of Table 5 shows respondents assigned to the "Police fight crime" treatment are almost twelve percentage points more likely to believe "police do everything they can to ensure that criminals receive the punishment that they deserve" ($p < 0.05$), an increase of around 50% from a control group share of 23%. Column 3 provides little evidence that this treatment affects respondents' expectations about whether police send perpetrators of vigilantism to prison. Columns 2 and 4 show a similar pattern for the "Police fight MV"

[13]See appendix section A.5.

[14]Analyses that do not involve the alarm treatment make use of endline measurements of priors that were asked before the administration of the information treatments.

treatment. While this treatment appears to increase the perception that police would send vigilantism perpetrators to prison by around 24% ($p < 0.1$), there is no evidence of an effect on expectations about service delivery efforts.

| | Believes police fight crime | | Believes police fight MV | | Would participate MV | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Police fight Crime | 0.116** | | −0.001 | | −0.0005 | |
| | (0.058) | | (0.052) | | (0.049) | |
| Police fight MV | | 0.035 | | 0.087* | | −0.083** |
| | | (0.059) | | (0.052) | | (0.048) |
| Control Mean | 0.23 | 0.28 | 0.42 | 0.37 | 0.39 | 0.43 |
| Control SD | 0.43 | 0.45 | 0.41 | 0.4 | 0.39 | 0.37 |
| RI p-value | 0.026 | 0.287 | 0.49 | 0.059 | 0.494 | 0.044 |
| Hypothesis | upr | upr | upr | upr | lwr | lwr |
| Observations | 244 | 244 | 244 | 244 | 244 | 244 |

*p<0.1; **p<0.05; ***p<0.01

**Table 5:** Effect of information treatments among respondents with low priors about risk of state punishment for mob vigilantism

Outcomes range from 0 to 1. The sample includes respondents from main and neighboring households with low priors about the risk of state punishment for vigilantism as measured during the endline survey. Appendix section A.2 provides information on model specification and testing, section D.5 on prior belief measures, and section D.4 on outcome question wording.

Hence, it seems the information treatments successfully created independent shifts in the two mediators. The question of interest is whether these shifts translate into changes in the willingness to participate in vigilantism. Column 5 contains little evidence that the "Police fight crime" treatment affected this outcome. The "Police fight MV," on the other hand, seems to have decreased the willingness to participate in vigilantism by almost 20% ($p < 0.05$). These findings are mirrored in Table 23 in the appendix which presents estimates among respondents with low prior beliefs about service delivery. Here, too, the "Police fight crime" treatment appears to have improved service delivery expectations; yet, this shift does not seem to translate into a reduction in the willingness to resort to vigilantism. The "Police fight MV" treatment appears to shift neither beliefs about police effort nor the willingness

to participate in vigilantism among this subgroup.

To summarize, even though the "Police fight crime" treatment appears to have improved service delivery expectations in both subgroups, there is little evidence that these shifts discouraged vigilantism. The "Police fight MV" treatment seems to have increased the belief that police send vigilantism perpetrators to prison only among those who, a priori, did not expect police to do so. Among this subgroup, the "Police fight MV" treatment also appears to have discouraged participation in vigilantism. Hence, the evidence suggests increasing the perceived risk of state punishment is more effective at discouraging vigilantism than improvements in service delivery perceptions.

### 4.3 Interactions between alarm and information treatments

The degree to which the results in the previous section elucidate the mechanisms through which the alarm produces its effects depends on whether the alarm and information treatments affect the mediators in similar ways. Another strategy is to lean into the fact that the alarm and information treatments are conceptually distinct, but leverage theoretical expectations about how these treatments should interact if a given mechanism is at play.

I argue the alarm and information treatments are complements. The alarm makes households more legible to police. This change may facilitate a speedier police response, but only if police attempt to find the household. Similarly, members of protected households may think police could use information about the household to identify members in a mob situation. This change should worry respondents only if they expect police to investigate vigilantism.

Hence, if service quality perceptions are an important mediator, the alarm should be particularly effective at discouraging vigilantism if combined with a treatment that convinces respondents that police make effort to deliver high quality services. Likewise, if punishment risk perceptions are an important mediator, the alarm should be particularly effective at discouraging vigilantism if combined with a treatment that convinces respondents of police efforts to convict vigilantes.

To test these predictions, I regress my endline measure of respondents' willingness to re-

sort to vigilantism on assignment indicators for the alarm and a given information treatment as well as the interaction. Estimates in columns 1 and 3 in Table 6 are based on all respondents from the 250 main households. Analyses in columns 2 and 4 exclude respondents who were assigned to the respective other information treatment.

| | Would Participate MV | | | |
|---|---|---|---|---|
| | All | Police fight MV = 0 | All | Police fight crime = 0 |
| | (1) | (2) | (3) | (4) |
| Alarm | −0.045 | 0.036 | 0.034 | 0.036 |
| | (0.050) | (0.072) | (0.050) | (0.072) |
| Alarm × Police fight Crime | 0.061 | 0.0002 | | |
| | (0.067) | (0.093) | | |
| Alarm × Police fight MV | | | −0.097* | −0.169** |
| | | | (0.068) | (0.097) |
| Control Mean | 0.28 | 0.24 | 0.27 | 0.24 |
| Control SD | 0.36 | 0.33 | 0.35 | 0.33 |
| RI p-value Alarm | 0.164 | 0.672 | 0.755 | 0.672 |
| Hypothesis Alarm | lwr | lwr | lwr | lwr |
| RI p-value Diff. | 0.816 | 0.47 | 0.076 | 0.043 |
| Hypothesis Diff | lwr | lwr | lwr | lwr |
| Number HHs | 237 | 174 | 237 | 161 |
| Observations | 448 | 228 | 448 | 211 |

*p<0.1; **p<0.05; ***p<0.01

**Table 6:** Interactive effects of alarm and information treatment on willingness to participate in mob vigilantism

Outcomes range from 0 to 1. Columns 1 and 3 include all respondents from main households. Columns 2 and 4 exclude respondents assigned to the respective other information treatment. "Control Mean" is the average outcome among respondents assigned to neither the alarm nor the respective information treatment. Appendix section D.4 provides details on outcome question wording and section A.2 on model specification.

Columns 1 and 2 show the alarm's estimated effect on the willingness to participate in vigilantism among respondents not assigned to the "Police fight crime" treatment is small and statistically insignificant. The interaction terms provide little evidence that the "Police fight crime" treatment made the alarm more effective at discouraging vigilantism. Columns 3 and 4 similarly suggest the alarm did not affect the willingness to resort to vigilantism among

respondents not assigned to the "Police fight MV" treatment. Yet, here, both interaction terms are negative and statistically significant, suggesting the "Police fight MV" treatment increased the extent to which the alarm discourages vigilantism ($p < 0.1$ and $p < 0.05$).

Priming respondents to believe police seek to convict vigilantism perpetrators appears to make the alarm more effective at discouraging vigilantism. Priming citizens to think police are committed to sanctioning perpetrators whom most citizens would like to see convicted does not seem to have the same effect. Again, the results point towards perceptions of punishment risks rather than service quality as a link between police capacity and vigilantism.

## 5 Alternative Explanations

Next, I consider whether my findings could be driven by factors other than the hypothesized mechanisms. One concern is that results may be driven by experimenter demand. While difficult to discard completely, several observations speak against this interpretation. First, respondents were asked about interest in the alarm at baseline, but mid- and endline interviews did not mention the alarm and enumerators were unaware of the study's purpose. Second, experimenter demand cannot explain the apparent concentration of effects among certain outcomes and subgroups. It its not obvious why respondents would censor only some of their opinions or why respondents with low expectations would be most inclined to do so.

One may be most worried about experimenter demand driving effects on views about vigilantism. To assess this possibility, respondents were asked at endline how many vigilante incidents they recall and had witnessed between May and July 2018. The alarms could not have affected these outcomes, because they were installed in September and October 2018. The treatment group remembering or witnessing fewer incidents than the control would thus suggest a treatment-induced reluctance to be linked to vigilantism. I find no evidence of such an effect (see Table 24 in the appendix.)

Another concern is that effects reflect changes among the control rather than the treatment group. Control group respondents who did not receive an alarm may have become

frustrated with police. Alternatively, police may have focused efforts on alarm owners, neglecting other households. Yet, Figure 9 in the appendix shows control group respondents became more positive about police and less supportive of vigilantism over time. Perhaps knowledge of the alarm project caused the control group to change in similar ways as the treatment group. If so, I would underestimate the alarm's effects.

Finally, the alarm's effects may be due to respondents feeling safer, which may decrease their demand for deterrence through harsh and immediate vigilante punishments. However, even though alarm owners seem to feel safer in their homes, I find little evidence of a reduction in crime victimization or in respondents' demand for harsh and immediate punishments per se (see Table 25 in the appendix).

## 6 Discussion

Many have suggested the prevalence of informal ways to deal with crime reflects the inability of state institutions to provide citizens with high quality law enforcement services. If institutions like the police were more effective, citizens would choose to rely on the state. Drawing on experimental variation in the police's ability to intervene in certain households but not others, I find an increase in police capacity indeed encouraged reliance on police and discouraged vigilantism. However, the treatment group developed both more sanguine views of police service quality and a greater sense that the state will punish vigilantes. Results from an information experiment indicate the risk of state punishment may play a bigger role in the decision to resort to vigilantism than police service quality.

Why may improvements in police service quality have limited effects on vigilantism? One possibility is that effects of service delivery improvements take longer to materialize, especially in South Africa where the Apartheid legacy strains citizen-state relations. Another possibility is that state justice is seen as an imperfect substitute for community justice, even if administered effectively. Smith (2019) argues South Africans resort to vigilantism because they dislike due process protections and wish for punishments to be harsher than those

provided by the state. Indeed more than 50% of South Africans are dissatisfied with the courts, the most common complaint being sentences are too lenient (StatsSA, 2016/2017). If citizens fundamentally oppose how the state sanctions law-breakers, it may not be possible to out-compete community punishment through improved police service delivery.

This article shows police capacity can nonetheless discourage vigilantism, because a capable police may help ensure that vigilantism perpetrators go to prison. That vigilantism is less prevalent in high capacity contexts hence need not reflect citizens' voluntary cooperation with the state. Instead, this pattern may result from the state's ability to limit citizens' choices by sanctioning those who take the law into their own hands.

Figure 6 in the appendix shows discontent with the state's punishment regime is widespread in Sub-Saharan Africa. Hence, South Africa is unlikely to be the only context where state and community punishments are perceived as imperfect substitutes. Similar logics may also apply to other informal practices. Traditional healers, for example, often supply controversial remedies or procedures like abortions that have been criminalized by the state. Where citizens see formal health care as an imperfect substitute, informal providers may remain popular even as government service quality improves. Another example are unlicensed moneylenders who often prevail despite increased availability of formal credit (Tsai, 2004). Notwithstanding high interest rates, borrowers may prefer loan sharks, for example because they do not require formal contracts. A lower prevalence of these informal services in high capacity contexts may in part reflect the state's ability to regulate suppliers.

A broader implication is that state capacity can have downsides from the perspective of citizens. Those who favor illegal practices like vigilantism may be wary of increased state presence, even if it improves government service quality. Study participants seemed cognizant of this trade-off. Support for vigilantism was particularly widespread among respondents who refused the alarm, and alarm owners sometimes asked to have its siren installed out of sight. This request seems counterintuitive if the alarm's only effect is better protection against intruders. If the alarm also deters vigilantism, however, households may want to hide the

alarm to enjoy improved police service delivery while upholding the threqat of community punishment. In line with this interpretation, I find no evidence that the alarm's effects spilled over to neighboring households.

An obvious caveat is that the alarm intervention may not be as effective at a larger scale. If every household received an alarm, it seems unlikely that police would be as familiar with or able to attend to alarm protected households. Fortunately, the theoretical interest here is not with the alarm intervention per se but with the downstream effects of the perception that police are capable to intervene in one's life on the choice between the state and vigilantism. The finding that receiving an alarm as part of a concentrated intervention in a small number of households discourages vigilantism provides evidence in favor of the popular intuition that state presence can help supersede informal alternatives, even if a larger alarm roll-out would do little to expand the reach of the state. Open questions remain, however, especially regarding the incentives of state officials. For example, given widespread support for vigilantism, why do politicians and bureaucrats enforce laws against it? An important next step will be to study the behavior not only of citizens but also of police and their political principals.

## References


Acemoglu, Daron, Ali Cheema, Asim I Khwaja, and James A Robinson. 2020. "Trust in State and Nonstate Actors: Evidence from Dispute Resolution in Pakistan." *Journal of Political Economy* 128 (8): 3090–3147.

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

Baker, Bruce. 2002. "Living with non-state policing in South Africa: the issues and dilemmas." *The Journal of Modern African Studies* 40 (1): 29–53.

Baker, Bruce. 2008. *Multi-choice policing in Africa*. Uppsala: Nordiska Afrikainstitutet.

Baldwin, Kate. 2016. *The Paradox of Traditional Chiefs in Democratic Africa*. New York: Cambridge University Press.

Becker, Gary S. 2000. "Crime and punishment: An economic approach." In *The economic dimensions of crime*, ed. Nigel G. Fielding, Alan Clarke, and Robert Witt. London: Palgrave Macmillan pp. 13–68.

Blair, Robert A, Sabrina M Karim, and Benjamin S Morse. 2019. "Establishing the rule of law in weak and war-torn states: Evidence from a field experiment with the Liberian National Police." *American Political Science Review* 113 (3): 641–657.

Bossuroy, Thomas, Clara Delavallade, and Vincent Pons. 2019. Biometric tracking, healthcare provision, and data quality: experimental evidence from tuberculosis control. NBER Working Paper 26388. Cambridge: National Bureau of Economic Research.
**URL:** *https://www.nber.org/papers/w26388*

Cooper, Jasper. 2019. State capacity and gender inequality: Experimental evidence from Papua new Guinea. Unpublished Manuscript.
**URL:** *https://jasper-cooper.com/papers/Cooper_ CAP.pdf*

García-Ponce, Omar, Lauren E Young, and Thomas Zeitzoff. 2022. "Anger and support for retribution in Mexico's drug war." *Journal of Peace Research* pp. 1–17.

Green, Donald P, Shang E Ha, and John G Bullock. 2010. "Enough Already about "Black Box" Experiments: Studying Mediation Is More Difficult than Most Scholars Suppose." *The Annals of the American Academy of Political and Social Science* 628 (1): 200–208.

Henn, Soeren J. 2022. "Complements or Substitutes? How Institutional Arrangements Bind Traditional Authorities and the State in Africa." *American Political Science Review* pp. 1–20.

ICG. 2022. Managing Vigilantism in Nigeria: A Near-term Necessity. Africa Report No 308. Brussels: International Crisis Group.
**URL:** *https://www.crisisgroup.org/africa/west-africa/nigeria/managing-vigilantism-nigeria-near-term-necessity*

Jaffrey, Sana. 2023. "Mechanics of Impunity: Vigilantism and State-Building in Indonesia." *Comparative Politics* 55 (2): 287–311.

Jung, Danielle F, and Dara Kay Cohen. 2020. *Lynching and Local Justice: Legitimacy and Accountability in Weak States*. Cambridge: Cambridge University Press.

Khayelitsha Commission. 2014. "Towards A Safer Khayelitsha. The Report of the Commission of Inquiry into Allegations of Police Inefficiency and a Breakdown in Relations between SAPS and the Community in Khayelitsha.".
**URL:** *http://www.saflii.org/khayelitshacommissionreport.pdf*

Kirsch, Thomas G, and Tilo Grätz. 2010. *Domesticating vigilantism in Africa*. Oxford: James Currey.

Lake, David. 2010. "The practice and theory of US statebuilding." *Journal of Intervention and Statebuilding* 4 (3): 257–284.

Ludwig, Jens, Jeffrey R Kling, and Sendhil Mullainathan. 2011. "Mechanism experiments and policy evaluations." *Journal of Economic Perspectives* 25 (3): 17–38.

Martland, Samuel J. 2014. "Standardizing the state while integrating the frontier: the Chilean telegraph system in the Araucanía, 1870–1900." *History and Technology* 30 (4): 283–308.

Miguel, Edward. 2005. "Poverty and witch killing." *The Review of Economic Studies* 72 (4): 1153–1172.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2016. "Building state capacity: Evidence from biometric smartcards in India." *American Economic Review* 106 (10): 2895–2929.

Nussio, Enzo, and Govinda Clayton. 2023. "A Wave of Lynching: Morality and Authority in Post-Tsunami Aceh." *Comparative Politics* 55 (2): 313–336.

Sandefur, Justin, and Bilal Siddiqi. 2012. Citizen or Subject? Forum Shopping and Legal Pluralism in Rural Liberia. IGC Working Paper S-3030-CCP-1. International Growth Centre.
**URL:** *https://www.theigc.org/sites/default/files/2012/03/Sandefur-Siddiqi-2012-Working-Paper.pdf*

SAPS. 2022. Annual Report 2021/2022. South African Police Service.
**URL:** *https://www.saps.gov.za/about/stratframework/annual_report/2021_2022/Annual-Report-2021-22.pdf*

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

Scott, James C. 1998. *Seeing like a state: How certain schemes to improve the human condition have failed.* New Haven: Yale University Press.

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

Smith, Nicholas Rush. 2019. *Contradictions of Democracy: vigilantism and rights in post-apartheid South Africa*. New York: Oxford University Press.

StatsSA. 2016/2017. Victims of Crime Survey. Statistics South Africa.
**URL:** *http://www.statssa.gov.za/?p=10521*

Super, Gail. 2022. "Cars, compounds and containers: Judicial and extrajudicial infrastructures of punishment in the 'old'and 'new'South Africa." *Punishment & Society* 24 (5): 824–842.

Tankebe, Justice. 2009. "Self-help, policing, and procedural justice: Ghanaian vigilantism and the rule of law." *Law & society review* 43 (2): 245–270.

Tsai, Kellee S. 2004. "Imperfect substitutes: the local political economy of informal finance and microfinance in rural China and India." *World Development* 32 (9): 1487–1507.

Tyler, Tom R, and Yuen Huo. 2002. *Trust in the law: Encouraging public cooperation with the police and courts*. New York: Russell Sage Foundation.

UNODC. 2015. Data on Police Personnel. United Nations Office on Drugs and Crime.
**URL:** *https://dataunodc.un.org/data/crime/Police%20personnel*

Xu, Xu. 2021. "To Repress or to Co-opt? Authoritarian Control in the Age of Digital Surveillance." *American Journal of Political Science* 65 (2): 309–325.

# Appendix

# A Additional Information

## A.1 Pre-registration

Pre-analysis plans (PAPs) can be found at https://osf.io/87u4f. Here, I describe divergences from the PAPs. Most divergences arise because of inconsistencies between the midline PAP (registered prior to the midline) and the endline PAP (registered in between the mid- and endline survey). Generally, I follow the endline PAP.

**Regression Specification.** The specification in the midline PAP includes block fixed effects, which are omitted in the endline PAP. The inclusion of fixed effects for a large number of small blocks (50 blocks with 5 units) substantially increases the number of parameters to be estimated. Moreover, in the presence of attrition, entire blocks may drop out of the analysis, especially when estimating conditional effects. Since fixed effects are not required for unbiasedness and in keeping with the endline PAP, I do not condition on block fixed effects. Both PAPs include a specification without and one with covariates selected through a LASSO procedure. I prioritize the barebones specification for transparency but show robustness to the other specification in appendix section C.2.

**Index construction.** I create indices as specified in the PAP for a given survey wave. Hence, indices do not always contain the same items across waves. More information is provided in appendix section D. Outcome construction diverges from the pre-specification as follows:

- *Alert Police.* The midline PAP specified this item would be combined with an indicator for whether the respondent "mentioned any form of reaching out to the police, including sounding the MeMeZa alarm" in response to an open-ended question about what she would do if attacked in her home. The estimated effect on this item is substantial. Yet, the measure conflates alarm availability with willingness to rely on police and was hence excluded form the endline survey. In keeping with the endline PAP, I do not analyze it here.
- *Support MV.* This index was only pre-registered at endline. The midline PAP pre-specified constituent items would be analyzed separately. Analyses of constituent items are in the appendix.
- *Service Index.* This index was only pre-registered at endline. The midline PAP pre-specified that some of the constituent items would be analyzed separately, while others would be combined into a sub-index. The item *Police are motivated* was meant to be part of the sub-index. I analyze this item separately, because it measures police motivation rather than service quality. Including the item *Police are motivated* in the *Service Index* does not materially change the results and analyses of all constituent items of the *Service Index* are shown in the appendix.
- *Rely Police.* Both PAPs pre-specified that constituent parts of this index (*Alert Police* and *Cooperate Police*) would be analyzed separately. These analyses are included in the appendix.

**Hypothesis tests.** All testing follows the endline PAP and diverges from the midline PAP as follows:

- *Support MV.* One-tailed test (lower). The midline PAP pre-specified a two-tailed test.

- *Call Com.* One-tailed test (lower). The midline PAP pre-specified a two-tailed test.
- *Service Index.* One-tailed test (upper). The midline PAP pre-specified a two-tailed test.
- *Join MV.* One-tailed test (upper) regarding the difference in conditional treatment effects across low and high prior groups. The midline PAP pre-specified the opposite one-tailed test (lower).

**Non-registered analyses.** The following analyses have not been pre-specified:

- *Table 4.* Both PAPs pre-specified analyses of treatment effect heterogeneity across prior beliefs about service quality for service quality outcomes (column 6) and across prior beliefs about punishment risks for perceptions of this risk (columns 1 and 2). While not pre-specified, analyses in the other columns arise from the same logic. The two dimensions of prior beliefs are not independent, and hence beliefs about one output may condition effects on beliefs about another.
- *Table 5.* Analyses in columns 2 and 3 have not been pre-specified. These analyses were added to demonstrate that the treatments did *not* affect beliefs they were *not* intended to affect.
- *Table 6.* Analyses in columns 2 and 4 have not been pre-specified, but provide a clean comparison if the information treatments interact to effect the willingness to participate in vigilantism.

**Registered analyses not reported in this paper:**

- *IV estimation.* The midline PAP pre-specified an IV estimator, which has been omitted from the endline PAP because of the high compliance rate.
- *Spillover analyses.* Both PAPs pre-specified analyses of spillover effects. The midline PAP specified a spatial spillover model. The endline PAP specified analyses of the sample of neighbors. The main text mentions I find no evidence of spillovers. Results are available upon request.
- *Omnibus tests.* The endline PAP specified two omnibus tests of the joint significance of two subsets of hypotheses to shed light on mechanisms. It has become clear that these tests are not well suited to discriminate between mechanisms. E.g., it is plausible that prior beliefs about one output may condition effects on believes about both outputs under both mechanisms.
- *Behavioral measure.* The endline PAP pre-specified a behavioral measure (respondents' choice between two kinds of t-shirts offered as a thank you gift). The results do little to strengthen or counter the results presented here and are available upon request.
- *Demand for policing.* The endline PAP pre-specified unrelated hypotheses about the the information treatments' effects on demand for policing that inform a follow-up project.

## A.2 Explanatory note for regression tables

**Alarm treatment.** Unless otherwise indicated, the unit of analysis is the respondent. Standard errors allow for clustering on the household level unless the dataset is collapsed to the household level. As pre-specified, I control for cluster size, i.e. the number of respondents interviewed per household. Unless indicated otherwise, no additional covariates are included. $p$-values are calculated, as pre-specified, using randomization inference by permuting treatment assignment 2,000 times to simulate

the sampling distribution under the sharp null hypothesis of no (positive/negative) treatment effect for any unit. The row labeled "hypothesis" in each table indicates the direction of hypotheses tests. Heterogeneous effects analyses make use of the pre-registered interaction specification, regressing the outcome on an indicator for treatment assignment, the moderator and the interaction between the two as well as the cluster size control. Randomization inference $p$-values for hypotheses involving conditional intent-to-treat effects (ITTs) are calculated by sub-setting to the respective group and using the same procedure of permuting treatment assignment 2,000 times. Tests of hypotheses involving the difference between conditional ITTs pertain to the sharp null hypothesis that the treatment effect for each unit is equal to the estimated ITT in the sample as a whole. The testing procedure is as follows: First, I adjust outcomes in the treatment group as if the estimated ITT for the sample as a whole were the true unit-level effect. Second, I permute treatment assignment 2,000 times. Third, I estimate the ITT in each subgroup and the difference between the two ITTs for every permutation. Finally, I compare the observed difference in conditional ITT estimates to the simulated sampling distribution to calculate a $p$-value. The resulting $p$-values can differ from parametric $p$-values based on clustered standard errors but the differences tend to be minor and can go in either direction (larger/smaller $p$- values).

**Information treatments.** Information treatments were randomized across the entire endline sample including 448 respondents from main and 376 from neighboring households. Analyses in Tables 5, 22 and 23 pertain to information treatments only, marginalize across the alarm treatment, include respondents from main and neighboring households, and estimate the effect of one factor of the full factorial design while marginalizing over the other. The pre-specified regression is

$$\boldsymbol{Y} = \alpha + \tau \boldsymbol{z}_{info} + \boldsymbol{\epsilon},$$

where $\boldsymbol{Y}$ is a vector of outcomes; $\alpha$ an intercept; $\tau$ the ITT of either the "Police fights crime" or the "Police fights mob vigilantism" prime; $\boldsymbol{z}_{info}$ a vector of indicators of assignments to the respective prime; and $\boldsymbol{\epsilon}$ a vector of error terms. The unit of analysis is the respondent. Standard errors are heteroskedasticity robust. Hypothesis tests are based on randomization inference drawing on the same simple random assignment procedure used to assign the information treatments in the first place.

**Alarm and information treatments.** Table 6 shows estimates from the following pre-registered regression specification:

$$\boldsymbol{Y} = \alpha + \tau_1 \boldsymbol{z}_{alarm} + \tau_2 \boldsymbol{z}_{info} + \tau_3 \boldsymbol{z}_{alarm} * \boldsymbol{z}_{info} + \delta \boldsymbol{n} + \boldsymbol{\epsilon},$$

where $\boldsymbol{Y}$ is a vector of outcomes; $\alpha$ an intercept; $\tau_1$ the ITT of the alarm treatment among respondents not assigned to the respective information treatment; $\boldsymbol{z}_{alarm}$ is a vector of indicators of assignment to the alarm; $\tau_2$ is the ITT of the information treatment among those who were not assigned to the alarm; $\boldsymbol{z}_{info}$ is a vector of indicators of assignments to the information treatment; $\tau_3$ is the difference in the effect of the alarm across those who were and were not assigned to the information treatment; $\boldsymbol{n}$ is a vector of cluster sizes and $\delta$ the associated coefficient; and $\boldsymbol{\epsilon}$ a vector of error terms that allows for clustering at the household level. $p$-values that pertain to hypotheses about $\tau_1$ and $\tau_2$ are calculated by sub-setting to the respective group and using randomization inference based on the random assignment function that was used, respectively, to assign households to the alarm or respondents to the information treatment. The $p$-value for $\tau_3$ pertains to the sharp null hypothesis that the treatment effect for each unit is equal to the estimated ITT in the sample as a

whole and is calculated using the procedure described above, this time permuting both assignment of the alarm and of the information treatments.

## A.3 Additional figures

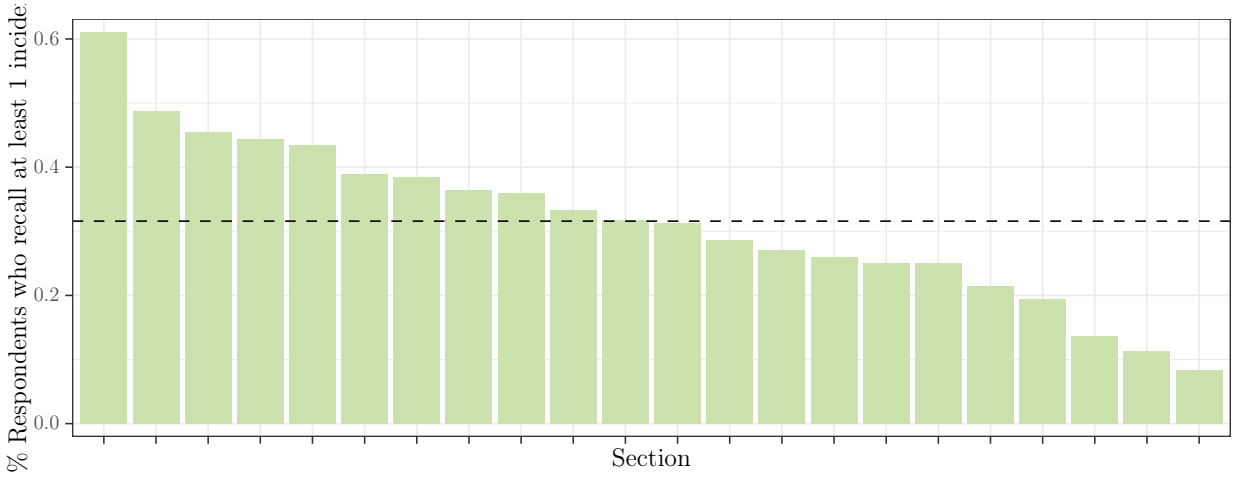


**Figure 4:** How many mob violence incidents can you recall that happened in your section?

Plot is based on all responses during the endline survey ($N = 815$). Respondents were asked "I would like you to think back to last year last winter, meaning May, June and July last year (2018). Can you recall any mob justice incidents that happened in your section during last winter?" If the respondent answered yes, they were asked "How many mob justice incidents can you recall from last winter?" Bars correspond to the share of respondents in each section that can re-call at least one incident.

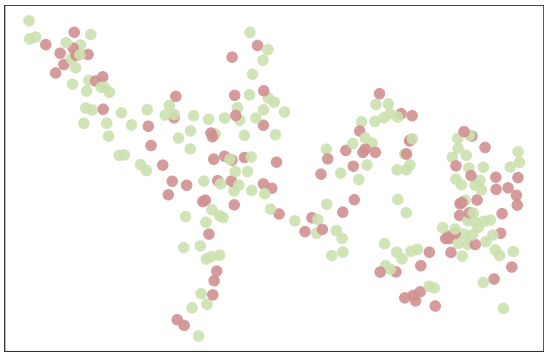


**Figure 5:** Households in study sample

Precise locations are not shown to protect respondent identities.

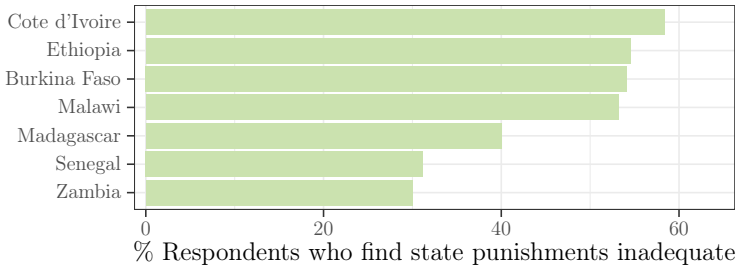


**Figure 6:** Views on state punishment in Sub-Saharan Africa
Based on data from a 2017 survey by the World Justice Project with around 1,000 respondents per country from urban centers. Question wording: "Please tell us how confident you are that the criminal justice system gives punishments which fit the crime?" The figure shows the % respondents who chose "Not very confident" or "Not at all confident."

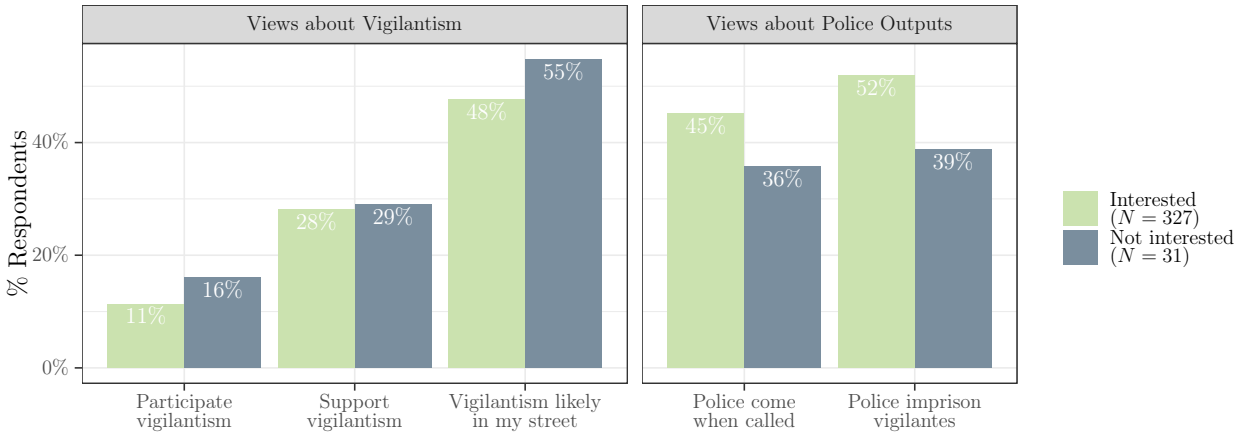


**Figure 7:** Baseline views by interest in police alarm.
Sample includes 358 baseline respondents. 250 are part of the experimental sample. Appendix sections 2.3 and A.4 provide details on selection criteria. 15 households are coded as "not interested," because they showed no interest during the baseline interview. 12 households were interested during the baseline, but had changed their mind when contacted during subsequent back-checks. Four households refused the alarm after they had been assigned to receive it.

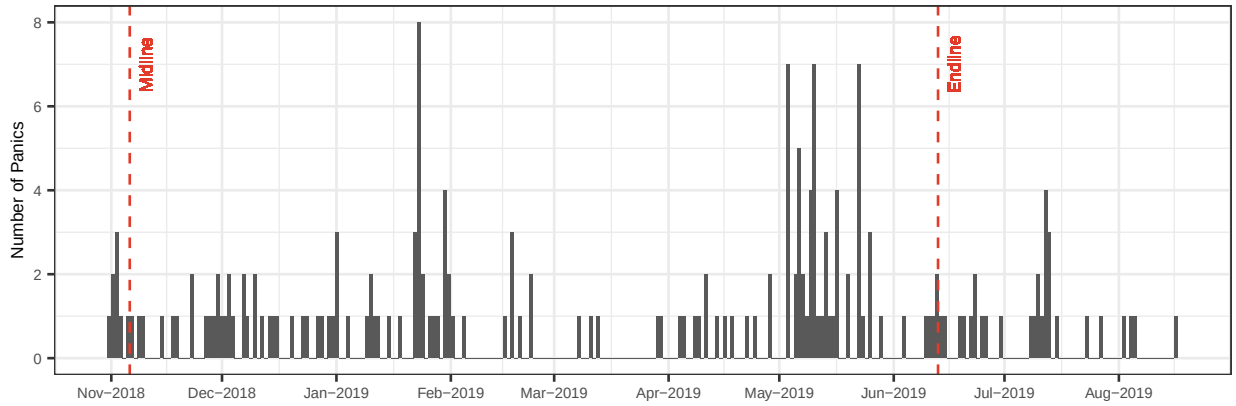


**Figure 8:** Alarm panics over time.

Data are taken from the back end system of the implementing partner.

## A.4 Sampling strategy for households

At baseline, I sampled households through two strategies. In line with the implementing partner's standard way of selecting alarm recipients, 150 households were sampled from a list provided by police. The police listed 390 households, of which 336 could be geo-located. To limit spillovers, a stochastic algorithm was used to identify sets of households such that every household is located at least 150 m from all other households in the set. Among the identified sets, the set with the overall maximum distance between units and the largest share of units with a distance of at least 200 m from all other units in the set was chosen. Second, 150 households were selected from areas with particularly high crime rates and low trust in the police as judged by local police. Enumerators walked through these areas and geo-located every tenth house. 946 households were geo-located in 11 areas. The same algorithm was used to identify sets of households that satisfy the distance constraint. Both sub-samples cover a similar area, since the police identified households in the same high-crime areas from which the second sub-sample was drawn. Sampled households were replaced if there was no adult woman residing there permanently, if the respondent could not be found after 3 visits, or if the respondent refused to be interviewed. 15 respondents refused and 16 could not be found. Towards the end of the survey, surveying became impossible in two areas due to opposition from the community who did not trust the surveyors. Thus, more households from the other high crime areas in the sample were added. Moreover, the existence of multiple houses with the same numbers led to inaccuracies in the geo-coordinates. Hence, not all households in the sample satisfy the 150 m distance constraint. Additional households were sampled to alleviate this problem. In total, 358 households were interviewed at baseline, 171 from the police list and 187 sampled from high crime areas. 250 were chosen as experimental units so as to minimize non-compliance and attrition:

- *Exclusion based on alarm interest.* Respondents were asked whether they are interested in an alarm. 15 households said no. 12 said yes, but changed their mind during back-checks.
- *Exclusion of CPF leaders.* The sample contained 5 of 10 executive members of the CPF, who received alarms non-randomly to ensure buy-in for the project.
- *Exclusion of households not reached during back-checks.* All remaining households were re-contacted via phone or in person during back-checks.

## A.5 Sampling strategy for neighbors

At endline, one neighboring household was sampled for each of the 250 main households. One randomly selected adult woman and man were interviewed in each household. In single gender households, I interviewed two men or two women. 84 neighboring households were one-member households. Hence, the target sample size was $N = 416$. The response rate was 84% ($N = 349$). Additional respondents were interviewed if available during the interview. 18 neighbors were sampled in this way, giving a total sample size of $N = 367$.

## A.6 Descriptive statistics

| | Police Sample (N = 135) | Listing Sample (N = 115) |
|---|---|---|
| Would participate mob vigilantism | 0.11 | 0.12 |
| Supports mob vigilantism | 0.28 | 0.28 |
| Would definitely call police | 0.51 | 0.33 |
| Perceives high risk of punishment for vigilantism | 0.62 | 0.43 |
| Believes police ensure the guilty go to prison | 0.34 | 0.24 |
| Feels safe in home | 0.24 | 0.30 |
| Age | 46.44 | 44.33 |
| Married | 0.44 | 0.36 |
| Household Size | 4.90 | 4.71 |
| Owns flushing toilet | 0.36 | 0.23 |
| Has tap water in house | 0.13 | 0.07 |
| Owns pay TV | 0.59 | 0.34 |
| Owns electric stove | 0.81 | 0.88 |
| Owns microwave | 0.62 | 0.63 |
| Owns washing machine | 0.54 | 0.45 |
| Owns motor vehicle | 0.23 | 0.20 |

**Table 7:** Averages of baseline covariates by sampling procedure

## A.7 Information treatments

**Police Fight Crime:** *Rapists sentenced to 13 life sentences and 240 years.* Two rapists were combinedly sentenced to 13 life sentences, as well as 240 years imprisonment after a rape and robbery spree in the Brits area in 2016. Obed Pilusa (31) and Sipho Nampa (31) were found guilty of numerous cases of rape and robbery between January and May 2016 and were sentenced by the Gauteng North High Court. Pilusa was sentenced to six life sentences for rape and 120 years imprisonment for eight counts of robbery. Nampa was sentenced to seven life sentences for rape and 120 years imprisonment for eight counts of robbery. The North West Provincial Police Commissioner, Lieutenant General Baile Motswenyane welcomed the hefty sentences. She congratulated the detectives of the Brits police's Family Violence, Child Protection and Sexual Offences Unit (FCS) for working tirelessly to ensure that the perpetrators were brought to book. "The sentences will serve as an indication that the police will not hesitate to deal harshly with those who commit crimes against women and children," she said.

**Police Fight Vigilantism:** *Acts of Vigilantism, A Concern to Northwest Police Commissioner.* The Provincial Commissioner Lieutenant General Baile Motswenyane is concerned about cases of vigilantism that are mushrooming in the province. According to police spokesperson in the North West, Colonel Sabata Mokwabone, the Provincial Commissioner's concerns stem from several acts of vigilantism where even some lives of people who were suspected of having committed crimes were lost. "Acts of vigilantism are condemned in the strongest terms they deserve. On the basis of the Constitution, I therefore make an appeal to communities not to commit acts of vigilantism, when you are found, the law will have to deal harshly with you." There are more than 40 cases

of vigilantism that have been reported in the province which the police are currently investigating and several suspects have been arrested so far. The Provincial Commissioner has warned that those responsible in perpetuating acts of vigilantism will soon feel the full might of the law.

| | Police fight crime = 1 | Police fight crime = 0 |
|---|---|---|
| Police fight mob vigilanitsm = 1 | 210 (113) | 189 (107) |
| Police fight mob vigilanitsm = 0 | 223 (124) | 193 (104) |

**Table 8:** Number of respondents across information treatment conditions.
First number in each cell pertains to respondents from all (main and neighboring) and number in parentheses to respondents from main households.

# B Identification

## B.1 Covariate balance

Table 9 examines balance among endline respondents. Most covariates are from the endline. Some measures plausibly unaffected by treatment (e.g. age) are from the endline. Columns show covariate means across conditions. To calculate the two-tailed $p$-value in the last column, I regress each covariate on a treatment assignment indicator and the cluster size control. I simulate the sampling distribution under the sharp null hypothesis of no effect of treatment on a covariate for any unit by permuting treatment assignment 2,000 times and re-estimating the same model. Then, I compare the observed coefficient of the treatment assignment indicator to the sampling distribution. If tests were independent, we would expect 5% of covariates to show imbalance significant at the 5% level. Here, 7/103 (7%) of tests yield a $p$-value equal to or less than .05.

| | Control | Treatment | $p$-value |
|---|---|---|---|
| prepaid_electricity_bl | 0.83 | 0.93 | 0.01 |
| electric_stove_bl | 0.80 | 0.94 | 0.01 |
| microwave_bl | 0.58 | 0.75 | 0.01 |
| approached_police_bl | 0.68 | 0.53 | 0.02 |
| spend_police_1_bl | 0.27 | 0.15 | 0.03 |
| floor_material_missing_el_fu | 0.29 | 0.20 | 0.04 |
| number_births_bl | 2.84 | 2.45 | 0.05 |
| earn_salary_el_fu | 0.55 | 0.48 | 0.09 |
| prisoners_guilty_bl | 0.46 | 0.58 | 0.09 |
| dishwasher_bl | 0.05 | 0.01 | 0.10 |
| tsonga_el_fu | 0.08 | 0.15 | 0.11 |
| interest_public_affairs_bl | 2.21 | 1.99 | 0.11 |
| own_refuse_dump_bl | 0.81 | 0.89 | 0.11 |
| discuss_neighbors_bl | 1.69 | 1.47 | 0.12 |
| pray_private_bl | 7.58 | 7.74 | 0.12 |
| criminals_from_area_bl | 0.40 | 0.29 | 0.13 |
| mob_violence_police_reaction_bl | 1.91 | 1.71 | 0.13 |
| sepedi_el_fu | 0.18 | 0.11 | 0.14 |
| experienced_violent_crime_bl | 0.09 | 0.17 | 0.16 |
| tap_water_in_yard_bl | 0.61 | 0.69 | 0.16 |
| in_a_relationship_el_fu | 0.19 | 0.15 | 0.19 |
| spend_electricity_bl | 0.56 | 0.49 | 0.20 |
| other_organizations_bl | 0.08 | 0.04 | 0.21 |
| retired_el_fu | 0.13 | 0.18 | 0.22 |
| interview_tswana_el_fu | 0.97 | 0.94 | 0.22 |
| work_full_time_el_fu | 0.19 | 0.14 | 0.23 |
| length_stay_el_fu | 4.02 | 4.16 | 0.23 |
| age_el_fu | 41.62 | 43.08 | 0.24 |
| pay_tv_bl | 0.45 | 0.53 | 0.26 |
| no_religion_el_fu | 0.10 | 0.13 | 0.29 |
| washing_machine_bl | 0.49 | 0.58 | 0.29 |
| home_language_sepedi_el_fu | 0.11 | 0.07 | 0.31 |
| hh_head_el_fu | 0.38 | 0.39 | 0.32 |
| government_does_enough_bl | 0.54 | 0.61 | 0.32 |
| main_income_salary_bl | 0.29 | 0.35 | 0.32 |
| pit_latrine_bl | 0.73 | 0.68 | 0.32 |
| private_security_bl | 0.02 | 0.01 | 0.33 |
| single_el_fu | 0.31 | 0.35 | 0.34 |
| flush_toilet_tank_bl | 0.13 | 0.16 | 0.37 |
| unemployed_el_fu | 0.41 | 0.38 | 0.38 |
| member_organization_bl | 0.80 | 0.86 | 0.38 |
| satisfaction_services_bl | 0.38 | 0.35 | 0.39 |

| | | | |
|---|---|---|---|
| kind_day_el_fu | 1.58 | 1.63 | 0.40 |
| state_official_bl | 0.10 | 0.14 | 0.40 |
| religious_service_bl | 1.38 | 1.30 | 0.41 |
| punishment_preferences_bl | 0.69 | 0.74 | 0.41 |
| know_number_bl | 0.80 | 0.86 | 0.41 |
| main_income_pensions_bl | 0.09 | 0.14 | 0.42 |
| observed_conditions_bl | 2.75 | 2.65 | 0.44 |
| join_mob_bl | 0.41 | 0.43 | 0.45 |
| motor_vehicle_bl | 0.25 | 0.23 | 0.45 |
| others_present_el_fu | 0.21 | 0.24 | 0.46 |
| police_ask_for_bribe_bl | 0.85 | 0.74 | 0.47 |
| completed_secondary_education_el_fu | 0.36 | 0.32 | 0.50 |
| tiled_floor_el_fu | 0.25 | 0.30 | 0.50 |
| trust_neighbor_bl | 0.79 | 0.76 | 0.50 |
| guard_dogs_bl | 0.22 | 0.28 | 0.50 |
| discuss_government_bl | 2.17 | 2.30 | 0.52 |
| child_hh_head_el_fu | 0.28 | 0.27 | 0.53 |
| lutheran_el_fu | 0.25 | 0.29 | 0.54 |
| concrete_floor_el_fu | 0.38 | 0.42 | 0.56 |
| mob_violence_plausibility_bl | 1.67 | 1.75 | 0.56 |
| beat_truck_driver_bl | 0.30 | 0.31 | 0.57 |
| due_process_bl | 0.85 | 0.88 | 0.59 |
| blow_whistle_bl | 0.15 | 0.13 | 0.60 |
| voice_heard_bl | 0.92 | 0.86 | 0.61 |
| spend_education_bl | 0.59 | 0.60 | 0.61 |
| number_incidents_bl | 0.93 | 0.86 | 0.63 |
| flush_toilet_public_bl | 0.13 | 0.16 | 0.64 |
| know_state_official_bl | 0.39 | 0.37 | 0.65 |
| spouse_hh_head_el_fu | 0.24 | 0.22 | 0.66 |
| work_part_time_el_fu | 0.15 | 0.14 | 0.67 |
| able_to_name_bl | 1.84 | 1.79 | 0.68 |
| discussed_crime_bl | 0.89 | 0.92 | 0.68 |
| secondary_education_incomplete_el_fu | 0.43 | 0.41 | 0.71 |
| report_informal_provider_bl | 0.78 | 0.76 | 0.71 |
| spend_police_2_bl | 0.52 | 0.55 | 0.71 |
| shout_community_bl | 0.73 | 0.76 | 0.74 |
| street_committee_connection_bl | 0.44 | 0.37 | 0.76 |
| police_quality_bl | 1.57 | 1.62 | 0.77 |
| home_language_tswana_el_fu | 0.69 | 0.68 | 0.81 |
| government_unresponsive_bl | 0.76 | 0.76 | 0.82 |
| call_police_bl | 2.30 | 2.33 | 0.84 |
| hh_size_bl | 4.98 | 5.09 | 0.85 |
| response_time_bl | 3.09 | 3.07 | 0.87 |
| tap_water_in_house_bl | 0.12 | 0.12 | 0.87 |
| zcc_el_fu | 0.18 | 0.19 | 0.88 |
| perceived_crime_risk_bl | 1.90 | 1.87 | 0.88 |
| number_school_children_bl | 1.46 | 1.50 | 0.89 |
| main_income_social_grants_bl | 0.44 | 0.43 | 0.90 |
| apostolic_el_fu | 0.21 | 0.21 | 0.91 |
| cpf_connection_bl | 1.38 | 1.39 | 0.91 |
| government_corrupt_bl | 0.61 | 0.62 | 0.94 |
| number_children_bl | 1.91 | 1.98 | 0.95 |
| feel_safe_bl | 0.28 | 0.29 | 0.97 |
| attend_meetings_street_committee_bl | 0.47 | 0.42 | 0.97 |
| female_el_fu | 0.64 | 0.63 | 0.98 |
| criminals_from_outside_bl | 0.38 | 0.38 | 0.98 |
| courts_punish_not_enough_bl | 0.75 | 0.73 | 0.98 |
| tswana_el_fu | 0.49 | 0.51 | 0.99 |
| adequate_force_bl | 0.57 | 0.57 | 0.99 |
| attend_meetings_cpf_bl | 1.54 | 1.63 | 0.99 |
| married_el_fu | 0.34 | 0.36 | 1.00 |

**Table 9:** Balance on covariates among all respondents in endline ($N = 448$)

## B.2 Attrition

| | Treatment | Control | $p$-value |
|---|---|---|---|
| Single Member Household | 10 (N = 100) | 13 (N = 150) | 0.836 |
| Respondent Not Interviewed Midline | 13 (N = 190) | 26 (N = 287) | 0.452 |
| Respondent Not Interviewed Endline | 21 (N = 190) | 49 (N = 287) | 0.121 |

**Table 10:** Reported household size and rates of attrition across experimental conditions
The outcome in row 1 is an indicator for whether a household has only one member. The unit of analysis is the household. The outcomes in rows 2 and 3 are indicators for whether a respondent attrited in the midline or endline survey, respectively. The unit of analysis is the respondent. Rows 2 and 3 assume that for the response rate to be 100%, 477 respondents should have been interviewed, two for each household other than the 23 single-member households. $p$-values stem from an unequal variance $t$-test conducted via randomization inference by permuting treatment assignment 2,000 times to generate the distribution of the test statistic under the sharp null hypothesis of no effect of treatment on reported household size or attrition for any unit.

| | $p$-value | N |
|---|---|---|
| 1 | 0.222 | 477 |
| 2 | 0.818 | 477 |

**Table 11:** $F$-test of treatment-by-covariate interactions in models of attrition
$P$-values come from an $F$-test that compares the following two models. The full model regresses an indicator for whether a respondent attrited on an indicator for treatment assignment and all treatment-by-covariate interactions using eight pre-registered baseline covariates. The nested model restricts all interaction terms to be zero. Row 1 pertains to the midline survey and row 2 pertains to the endline survey. The unit of analysis is the respondent and the analysis is based on two "completed" datasets which assume that, for the response rate to be 100%, 477 respondents should have been interviewed, two respondents per household other than the 23 households that have only one household member. $p$-values have been calculated using randomization inference by permuting treatment assignment 2,000 times.

## B.3 Additional respondents

| | Any Additional Resp. | | N Additional Resp. | |
|---|---|---|---|---|
| | Midline | Endline | Midline | Endline |
| | (1) | (2) | (3) | (4) |
| Alarm Treatment | 0.058 | 0.066 | 0.068 | 0.077 |
| | (0.048) | (0.049) | (0.059) | (0.051) |
| Control Mean | 0.136 | 0.134 | 0.156 | 0.134 |
| Control SD | 0.344 | 0.342 | 0.433 | 0.342 |
| RI $p$-value | 0.22 | 0.248 | 0.174 | 0.126 |
| Number HHs | 245 | 237 | 245 | 237 |
| Hypothesis | two | two | two | two |
| Observations | 245 | 237 | 245 | 237 |
| Adjusted $R^2$ | 0.002 | 0.004 | 0.001 | 0.005 |

$^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

**Table 12:** Additional respondents sampled across experimental conditions
The unit of analysis is the household. The sample contains all main households in which at least one respondent was interviewed at, respectively, midline and endline. The outcome in columns 1 and 3 is an indicator for whether an additional respondent was interviewed in a given household. The outcome in columns 2 and 4 is the number of additional respondents interviewed. Outcomes are regressed on an indicator for treatment assignment. $p$-values are calculated using randomization inference.

# C Additional Analyses

## C.1 Additional outcomes

| | Spoken to police | | | |
|---|---|---|---|---|
| | Midline | Endline | Endline | Endline |
| Alarm | −0.017 | 0.041 | 0.042 | 0.054 |
| | (0.035) | (0.050) | (0.071) | (0.065) |
| Alarm × High Prior Punishment | | | 0.014 | |
| | | | (0.098) | |
| Alarm × High Prior Service | | | | −0.007 |
| | | | | (0.096) |
| Control Mean | 0.18 | 0.44 | 0.44 | 0.44 |
| Control SD | 0.38 | 0.5 | 0.5 | 0.5 |
| RI p-value Main | 0.671 | 0.213 | 0.297 | 0.182 |
| Hypothesis Main | upr | upr | upr | upr |
| RI *p*-value Diff. | - | - | 0.552 | 0.405 |
| Hypothesis Diff | - | - | lwr | lwr |
| Number HHs | 245 | 237 | 237 | 237 |
| Observations | 483 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 13:** Effects of the alarm on whether respondents have recently spoken to police
Outcome is binary. Appendix section D.5 contains details on measures of prior beliefs and table 1 shows their distribution. See appendix section A.2 for model specification, and appendix section D.6 on outcome question wording.

| | Police would discover | |
|---|---|---|
| Alarm | 0.027 | 0.001 |
| | (0.038) | (0.036) |
| Alarm × High Prior Punishment | 0.024 | |
| | (0.050) | |
| Alarm × High Prior Service | | 0.098 |
| | | (0.047) |
| Control Mean | 0.78 | 0.78 |
| Control SD | 0.26 | 0.26 |
| RI $p$-value Main | 0.277 | 0.515 |
| Hypothesis Main | upr | upr |
| RI $p$-value Diff. | 0.721 | 0.978 |
| Hypothesis Diff | lwr | lwr |
| Number HHs | 237 | 237 |
| Observations | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 14:** Effects of alarm on perceptions of likelihood that police would find out about illegal behavior.
Outcome measures range from 0 to 1. Appendix section D.5 contains details on prior belief measures and table 1 shows their distribution. See appendix section A.2 for model specification, and appendix section D.8 on outcome wording.

| | Support MV | | Call Com. | | Support MV | Call Com. | Support MV | Call Com. |
|---|---|---|---|---|---|---|---|---|
| | Midline | Endline | Midline | Endline | Endline | Endline | Endline | Endline |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Alarm | −0.040* | −0.031 | −0.002 | 0.030 | −0.118** | 0.032 | −0.042 | 0.048 |
| | (0.032) | (0.042) | (0.024) | (0.025) | (0.066) | (0.036) | (0.056) | (0.029) |
| Alarm × High Prior Punishment | | | | | 0.156** | −0.007 | | |
| | | | | | (0.086) | (0.048) | | |
| Alarm × High Prior Service | | | | | | | 0.013 | −0.053 |
| | | | | | | | (0.087) | (0.050) |
| Control Mean | 0.3 | 0.37 | 0.78 | 0.76 | 0.37 | 0.76 | 0.37 | 0.76 |
| Control SD | 0.33 | 0.41 | 0.25 | 0.27 | 0.41 | 0.27 | 0.41 | 0.27 |
| RI p-value Main | 0.09 | 0.23 | 0.462 | 0.883 | 0.045 | 0.771 | 0.226 | 0.933 |
| Hypothesis Main | lwr | lwr | lwr | lwr | lwr | lwr | lwr | lwr |
| RI $p$-value Diff. | - | - | - | - | 0.043 | 0.507 | 0.426 | 0.83 |
| Hypothesis Diff | - | - | - | - | upr | upr | upr | upr |
| Number HHs | 245 | 237 | 245 | 237 | 237 | 237 | 237 | 237 |
| Observations | 483 | 448 | 483 | 448 | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 15:** Effects of the alarm treatment on respondents' support for mob vigilantism and willingness to call the community. Outcomes range from 0 to 1. Analyses in columns 5 to 8 regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction, and the cluster size control. One respondent was interviewed per household at baseline and their response is interpreted as a household-level measure of prior beliefs. The measure of priors about punishment (columns 5 and 6) asks whether it is likely (unlikely) that vigilantism perpetrators would be arrested. The measure of service quality priors (columns 7 and 8) indicates whether respondents fall above the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See appendix section D.5 for question wording and Table 1 for the distribution of prior beliefs. The table displays randomization inference $p$-values and directions of hypothesis tests. Appendix section A.2 provides details on model specification and testing, and appendix section D.7 on outcome question wording and coding.

## C.2 Adjusting for covariates

| | Rely police | | Join MV | | Rely police | Join MV | Rely police | Join MV |
|---|---|---|---|---|---|---|---|---|
| | Midline | Endline | Midline | Endline | Endline | Endline | Endline | Endline |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Alarm | 0.066*** | 0.049** | −0.070** | −0.003 | 0.071* | −0.069 | 0.096*** | −0.027 |
| | (0.026) | (0.029) | (0.029) | (0.026) | (0.042) | (0.038) | (0.037) | (0.036) |
| Alarm × High Prior Punishment | | | | | −0.047 | 0.127** | | |
| | | | | | (0.056) | (0.053) | | |
| Alarm × High Prior Service | | | | | | | −0.108** | 0.055 |
| | | | | | | | (0.055) | (0.054) |
| Control Mean | 0.6 | 0.64 | 0.24 | 0.17 | 0.64 | 0.17 | 0.64 | 0.17 |
| Control sd | 0.31 | 0.31 | 0.37 | 0.29 | 0.31 | 0.29 | 0.31 | 0.29 |
| RI p-value Main | 0.008 | 0.033 | 0.012 | 0.466 | 0.052 | 0.114 | 0.004 | 0.292 |
| Hypothesis Main | upr | upr | lwr | lwr | upr | lwr | upr | lwr |
| RI $p$-value Diff. | - | - | - | - | 0.208 | 0.018 | 0.034 | 0.118 |
| Hypothesis Diff | - | - | - | - | lwr | upr | lwr | upr |
| Number of LASSO Cov. | 30 | 31 | 6 | 22 | 31 | 22 | 30 | 22 |
| Number HHs | 245 | 237 | 245 | 237 | 237 | 237 | 237 | 237 |
| Observations | 483 | 448 | 483 | 448 | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 16:** Effects of alarm treatment on respondents' willingness to rely on police and participate in mob vigilantism estimated with covariate adjustment.

Outcomes range from 0 to 1. Analyses in columns 5 to 8 regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction, and the cluster size control. In addition, all specifications control for a set of covariates selected through a pre-specified LASSO regression procedure. One respondent was interviewed per household at baseline and their response is interpreted as a household-level measure of prior beliefs. The measure of priors about punishment (columns 5 and 6) asks whether it is likely (unlikely) that vigilantism perpetrators would be arrested. The measure of service quality priors (columns 7 and 8) indicates whether respondents fall above the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See appendix section D.5 for question wording and Table 1 for the distribution of prior beliefs. The table displays randomization inference $p$-values and directions of hypothesis tests. Appendix section A.2 provides details on model specification and testing, and appendix section D.1 on outcome question wording and coding.

## C.3 Disaggregating indices

| | Alert police | | Coop. police | | Alert police | Coop. police | Alert police | Coop. police |
|---|---|---|---|---|---|---|---|---|
| | Midline | Endline | Midline | Endline | Endline | Endline | Endline | Endline |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Alarm | 0.103*** | 0.079** | 0.091*** | 0.070** | 0.140*** | 0.124** | 0.132*** | 0.120*** |
| | (0.038) | (0.037) | (0.030) | (0.035) | (0.051) | (0.054) | (0.051) | (0.047) |
| Alarm × High Prior Punishment | | | | | −0.114* | −0.103* | | |
| | | | | | (0.074) | (0.071) | | |
| Alarm × High Prior Service | | | | | | | −0.110* | −0.098 |
| | | | | | | | (0.072) | (0.071) |
| Control Mean | 0.65 | 0.7 | 0.56 | 0.58 | 0.7 | 0.58 | 0.7 | 0.58 |
| Control SD | 0.4 | 0.38 | 0.34 | 0.35 | 0.38 | 0.35 | 0.38 | 0.35 |
| RI p-value Main | 0.003 | 0.015 | 0 | 0.019 | 0.002 | 0.017 | 0.004 | 0.006 |
| Hypothesis Main | upr | upr | upr | upr | upr | upr | upr | upr |
| RI *p*-value Diff. | - | - | - | - | 0.06 | 0.084 | 0.07 | 0.107 |
| Hypothesis Diff | - | - | - | - | lwr | lwr | lwr | lwr |
| Number HHs | 245 | 237 | 245 | 237 | 237 | 237 | 237 | 237 |
| Observations | 483 | 448 | 483 | 448 | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 17:** Effects of alarm on components of index "Rely Police" (see Table 2)
Outcomes range from 0 to 1. Analyses in columns 5 to 8 regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction, and the cluster size control. One respondent was interviewed per household at baseline and their response is interpreted as a household-level measure of prior beliefs. The measure of priors about punishment (columns 5 and 6) asks whether it is likely (unlikely) that vigilantism perpetrators would be arrested. The measure of service quality priors (columns 7 and 8) indicates whether respondents fall above the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See appendix section D.5 for question wording and Table 1 for the distribution of prior beliefs. The table displays randomization inference *p*-values and directions of hypothesis tests. Appendix section A.2 provides details on model specification and testing, and appendix section D.1 on outcome question wording and coding.

| | Join MV (Endline) | | | | | |
|---|---|---|---|---|---|---|
| | Join beating | Join mob | Join beating | Join mob | Join beating | Join mob |
| Alarm | −0.039 | 0.016 | −0.128*** | −0.071* | −0.085** | 0.0004 |
| | (0.035) | (0.033) | (0.051) | (0.052) | (0.045) | (0.044) |
| Alarm × High Prior Punishment | | | 0.165** | 0.150*** | | |
| | | | (0.069) | (0.067) | | |
| Alarm × High Prior Service | | | | | 0.101 | 0.031 |
| | | | | | (0.070) | (0.066) |
| Control Mean | 0.21 | 0.14 | 0.21 | 0.14 | 0.21 | 0.14 |
| Control SD | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 | 0.35 |
| RI p-value Main | 0.136 | 0.684 | 0.008 | 0.066 | 0.038 | 0.581 |
| Hypothesis Main | lwr | lwr | lwr | lwr | lwr | lwr |
| RI $p$-value Diff. | - | - | 0.016 | 0.008 | 0.118 | 0.434 |
| Hypothesis Diff | - | - | upr | upr | upr | upr |
| Number HHs | 237 | 237 | 237 | 237 | 237 | 237 |
| Observations | 448 | 448 | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 18:** Effects of alarm on individual items used to create the index "Join MV" at endline.
Outcomes range from 0 to 1. Analyses in columns 3 to 6 regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction, and the cluster size control. One respondent was interviewed per household at baseline and their response is interpreted as a household-level measure of prior beliefs. The measure of priors about punishment (columns 3 and 4) asks whether it is likely (unlikely) that vigilantism perpetrators would be arrested. The measure of service quality priors (columns 5 and 6) indicates whether respondents fall above the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See appendix section D.5 for question wording and Table 1 for the distribution of prior beliefs. The table displays randomization inference $p$-values and directions of hypothesis tests. Appendix section A.2 provides details on model specification and testing, and appendix section D.1 on outcome question wording and coding.

| | Police... | | | |
|---|---|---|---|---|
| | know name | know house | know name | know house |
| | (1) | (2) | (3) | (4) |
| Alarm | 0.097* | 0.130** | 0.068* | 0.086* |
| | (0.071) | (0.073) | (0.062) | (0.066) |
| Alarm × High Prior Punishment | 0.011 | −0.011 | | |
| | (0.094) | (0.100) | | |
| Alarm × High Prior Service | | | 0.080 | 0.106 |
| | | | (0.095) | (0.097) |
| Control Mean | 0.45 | 0.44 | 0.45 | 0.44 |
| Control SD | 0.46 | 0.5 | 0.46 | 0.5 |
| RI p-value Main | 0.091 | 0.037 | 0.094 | 0.068 |
| Hypothesis Main | upr | upr | upr | upr |
| RI *p*-value Diff. | 0.535 | 0.409 | 0.675 | 0.781 |
| Hypothesis Diff | lwr | lwr | lwr | lwr |
| Number HHs | 237 | 237 | 237 | 237 |
| Observations | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 19:** Effects of alarm on individual items used to create the index "Police know HH".
Outcomes range from 0 to 1. All analyses regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction, and the cluster size control. One respondent was interviewed per household at baseline and their response is interpreted as a household-level measure of prior beliefs. The measure of priors about punishment (columns 1 and 2) asks whether it is likely (unlikely) that vigilantism perpetrators would be arrested. The measure of service quality priors (columns 3 and 4) indicates whether respondents fall above the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See appendix section D.5 for question wording and Table 1 for the distribution of prior beliefs. The table displays randomization inference *p*-values and directions of hypothesis tests. Appendix section A.2 provides details on model specification and testing, and appendix section D.2 on outcome question wording and coding.

| | Arrive quickly | Take problem seriously | Send guilty to prison | Arrive quickly | Take problem seriously | Send guilty to prison |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Alarm | 0.129** | −0.020 | 0.146** | 0.125*** | −0.004 | 0.135** |
| | (0.055) | (0.042) | (0.072) | (0.047) | (0.038) | (0.061) |
| Alarm × High Prior Punishment | −0.068 | −0.024 | −0.189** | | | |
| | (0.072) | (0.058) | (0.095) | | | |
| Alarm × High Prior Service | | | | −0.051 | −0.034 | −0.187** |
| | | | | (0.073) | (0.058) | (0.095) |
| Control Mean | 0.45 | 0.74 | 0.46 | 0.45 | 0.74 | 0.46 |
| Control SD | 0.35 | 0.29 | 0.5 | 0.35 | 0.29 | 0.5 |
| RI $p$-value Main | 0.014 | 0.691 | 0.026 | 0.006 | 0.564 | 0.016 |
| Hypothesis Main | upr | upr | upr | upr | upr | upr |
| RI $p$-value Diff. | 0.192 | 0.347 | 0.022 | 0.258 | 0.348 | 0.022 |
| Hypothesis Diff | lwr | lwr | lwr | lwr | lwr | lwr |
| Number HHs | 237 | 237 | 237 | 237 | 237 | 237 |
| Observations | 448 | 448 | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 20:** Effects of alarm on on individual items used to create the "Service index."
All outcome measures range from 0 to 1. All specifications regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction between the two, and the cluster size control. Dichotomous baseline measures of prior beliefs are treated as household-level measurements, since only one respondent was interviewed per household at baseline. Prior beliefs about punishment (columns 1 to 3) are measured through an item that asks whether it is likely (unlikely) that participants in a hypothetical incident of vigilantism would be arrested. The measure of prior beliefs about service quality (columns 4 to 6) captures whether respondents fall above or below the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See section D.5 for details on question wording and Table 1 for the joint distribution of prior beliefs. Randomization inference $p$-values and directions of hypothesis tests are displayed in the table. Section A.2 of the appendix contains more details on model specification. See section D.3 for question wording and coding of outcomes.

## C.4 Unconditional effects on intermediate outcomes

| | Police know HH | Police are motivated | | Service index | | Would discover | Respond MV | Imprison MV |
|---|---|---|---|---|---|---|---|---|
| | Endline | Midline | Endline | Midline | Endline | Endline | Endline | Endline |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Alarm | 0.102** | 0.068* | 0.114** | 0.060** | 0.034 | 0.039* | 0.002 | 0.041 |
| | (0.047) | (0.048) | (0.048) | (0.026) | (0.029) | (0.026) | (0.034) | (0.043) |
| Control Mean | 0.44 | 0.42 | 0.47 | 0.43 | 0.55 | 0.78 | 0.67 | 0.71 |
| RI $p$-value | 0.015 | 0.081 | 0.014 | 0.024 | 0.131 | 0.067 | 0.464 | 0.204 |
| Hypothesis | upr | upr | upr | two | upr | upr | upr | upr |
| Number HHs | 237 | 245 | 237 | 245 | 237 | 237 | 237 | 237 |
| Observations | 448 | 483 | 448 | 483 | 448 | 448 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 21:** Effects of alarm on perceptions of police.
All outcome measures range from zero to one. Randomization inference $p$-values and directions of hypothesis tests are displayed in the table. Section A.2 of the appendix contains details on model specification. See appendix sections D.2, D.3, and D.8 for question wording and coding of outcomes.

### C.5 Additional results information treatments

| | Police Punish Criminals | Police Punish Mob Justice |
|---|---|---|
| Police Performance | 0.020 | |
| | (0.035) | |
| Police Oversight | | −0.009 |
| | | (0.028) |
| Control Mean | 0.459 | 0.652 |
| Control SD | 0.497 | 0.402 |
| RI $p$-value | 0.263 | 0.63 |
| Hypothesis | upr | upr |
| Observations | 815 | 815 |
| *Note:* | | $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01 |

**Table 22:** Effect of information treatments among all endline respondents
Sample includes all endline respondents from main and neighboring households. See appendix sections A.2 and D.4 on model specification and question wording.

| | Believes police fight crime | | Believes police fight MV | | Would participate MV | |
|---|---|---|---|---|---|---|
| Police fight Crime | 0.062* | | 0.008 | | 0.007 | |
| | (0.040) | | (0.038) | | (0.035) | |
| Police fight MV | | −0.019 | | 0.028 | | −0.025 |
| | | (0.040) | | (0.038) | | (0.035) |
| Control Mean | 0.2 | 0.25 | 0.57 | 0.56 | 0.34 | 0.36 |
| Control SD | 0.4 | 0.43 | 0.41 | 0.41 | 0.37 | 0.38 |
| RI $p$-value | 0.067 | 0.67 | 0.424 | 0.244 | 0.597 | 0.234 |
| Hypothesis | upr | upr | upr | upr | lwr | lwr |
| Observations | 453 | 453 | 453 | 453 | 453 | 453 |

*p<0.1; **p<0.05; ***p<0.01

**Table 23:** Effect of information treatments among respondents with low priors about police service
Sample includes respondents from main and neighboring households with low priors about police service quality as measured at endline. See appendix section D.5 for prior belief measures, section A.2 on model specification and section D.4 on outcome question wording.

## C.6 Ruling out alternative explanations

### C.6.1 Social desirability bias

| | All endline respondents | | | Low priors legal repercussions MV | | |
|---|---|---|---|---|---|---|
| | Any MV incidents | Number MV incidents | Witnessed any | Any MV incidents | Number MV incidents | Witnessed any |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Alarm | 0.042 | 0.267 | 0.053 | 0.048 | 0.304 | 0.040 |
| | (0.050) | (0.174) | (0.046) | (0.075) | (0.226) | (0.066) |
| Control Mean | 0.31 | 0.76 | 0.22 | 0.33 | 0.69 | 0.22 |
| Control SD | 0.46 | 1.39 | 0.41 | 0.47 | 1.18 | 0.42 |
| RI $p$-value | 0.802 | 0.95 | 0.884 | 0.74 | 0.906 | 0.734 |
| Hypothesis | lwr | lwr | lwr | lwr | lwr | lwr |
| Number HHs | 237 | 237 | 237 | 110 | 110 | 110 |
| Observations | 448 | 448 | 448 | 202 | 202 | 202 |

$^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

**Table 24:** Effect of the alarm treatment on recollection of incidents of mob vigilantism that happened *prior* to treatment
During the endline survey, respondents were asked "I would like you to think back to last year last winter, meaning May, June and July last year (2018). Can you recall any mob justice incidents that happened in your section during last winter?" If they answered "yes," they were asked "How many mob justice incidents can you recall from last winter?" as well as "Did you personally witness any of these mob justice incidents?" The outcome in columns 1 and 4 is an indicator variable for whether respondents can recall any incidents. The outcome in columns 2 and 5 is the number of incidents that a respondent can recall. The outcome in columns 3 and 6 is an indicator for whether a respondent reports having witnessed any incidents of vigilantism. Those who cannot recall an incident are coded as zero. Analyses in columns 1 to 3 are based on the entire sample. Analyses in columns 4 to 6 are subset to respondents from households with low priors about the likelihood of state punishment for vigilante violence. This subgroup is of relevance, because it sees the largest (in absolute value) treatment effects on the willingness to participate in vigilante violence. See section A.2 of the appendix for information on model specification.

### C.6.2 Changes among control group

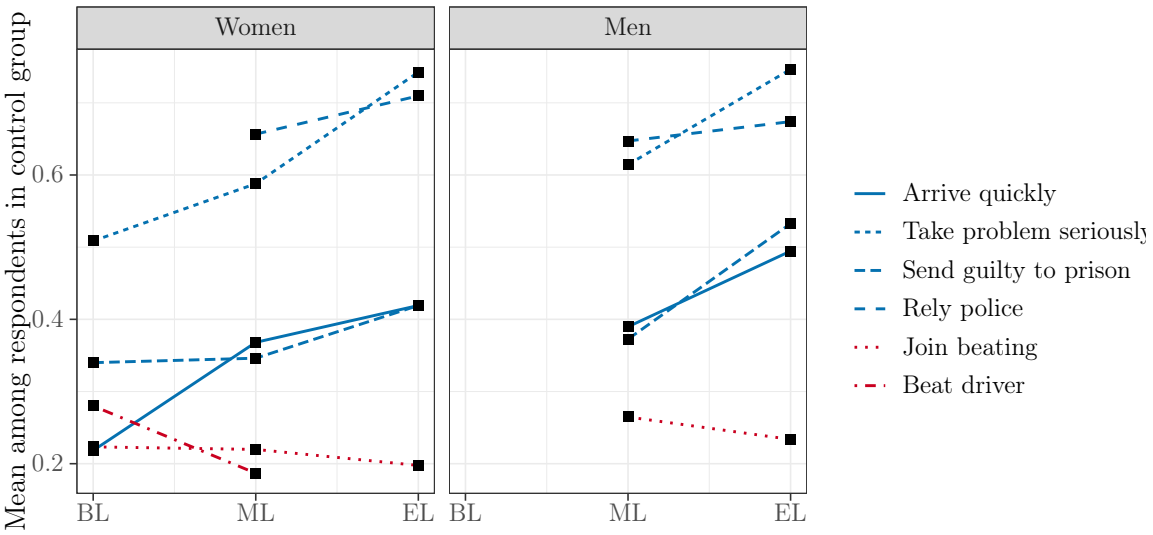


**Figure 9:** Change in outcomes in control group across survey waves by gender

Only women were interviewed at baseline. Outcomes in blue relate to police; outcomes in red relate to vigilantism. BL stands for baseline, ML midline and EL for endline. See sections D.1, D.3, and D.7 for question wording.

### C.6.3 Change in punishment preferences due to improved safety

| | Feel Safe | HH experienced crime | Punish more | Quick justice |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Alarm | 0.099*** | −0.018 | −0.009 | −0.014 |
| | (0.028) | (0.049) | (0.047) | (0.050) |
| Control Mean | 0.59 | 0.18 | 0.46 | 0.51 |
| Control SD | 0.29 | 0.38 | 0.5 | 0.5 |
| RI $p$-value | 0 | 0.368 | 0.86 | 0.778 |
| Hypothesis | upr | lwr | two | two |
| Unit of Analysis | Ind. | HH | Ind. | Ind. |
| Number HHs | 237 | 237 | 237 | 237 |
| Observations | 448 | 237 | 448 | 448 |

*p<0.1; **p<0.05; ***p<0.01

**Table 25:** Effect of the alarm treatment on safety and punishment preferences
Outcomes measured at endline. Question wording: *Feel safe*: Do you feel safe in your home during [at random: day/night] time? If yes: Do you feel just safe or very safe? If no: Do you feel just unsafe or very unsafe? 1 = Very safe, 0.66 = Just safe, 0.33 = Just unsafe, 0 = Very unsafe. *Crime Victimization:* Since last Christmas, did any crime happen in your house or yard? 1 = Yes, 0 = No. Answers have been collapsed to household level means. *Punish more:* Imagine you've been robbed at knifepoint and you report the robbery to the police. The robber took your belongings but did not hurt you. The police arrest the robber, and he will be kept in prison for 2 years. Is that a severe enough punishment, or should he have been punished more? 0 = It is severe enough., 1 = He should have been punished more. *Quick justice:* Please tell me which of the following statements comes closest to your view: 1 = Statement 1: The most important thing is that justice is served quickly. 0 = Statement 2: As long as the sentence is fair, I don't mind how long it takes for justice to be served. See section A.2 of the appendix for information on model specification.

# D Question Wording

The responses "don't know" and "refuse to answer" are coded as missing and imputed using multivariate imputation via chained equations. Indices are created by averaging across items.

## D.1 Table 2

*Column 1, 2, 5 and 7: Rely Police.* This measure is an index of one item and one sub-index:

- *Alert Police:* Suppose someone is trying to enter your home to steal something from you. Some people say that reaching out to the police in such situations is useless, because the

police won't arrive in time anyway. What about you, which of the following comes closest to what you would do? 0 =I would not rely on the police for help, 0.5 = I may alert the police later, but not right away, 1 = Before doing anything else, I would alert the police to come and help me.

- *Cooperate Police:* This measure is an index of the following three items:
  - *Report Police:* Please tell me which of the following statements comes closest to your view: 1 = If I see a crime, I will always report it to the police, 0 = I do not think it is worth reporting minor crimes to the police, because the police won't do anything anyway.
  - *Share Information (only part of the midline index):* Suppose you are aware that a member of your community is selling drugs. Which of the following are you most likely to do? 1 = I would report this person to the police, 0 = I would turn a blind eye, because I do not feel comfortable reporting criminals to the police.
  - *Report GBV:* Imagine you are at home watching TV in the afternoon. You hear your neighbor's wife screaming, because her husband is beating her. Which of the following are you most likely to do? 1 = I would alert the police, 0 = I would go to the neighbor's house and intervene, or, I would turn a blind eye.

*Column 3, 4, 6 and 8: Join MV.* This measure is a single item at midline and an index of two items at endline:

- *Join beating:* In this same situation, suppose some men from your community do get hold of the burglar who stole from you and that they want to beat him up. Which action are you most likely to take? 0 = I would try to calm the group down and tell them we should wait for the police, 0.5 = I would not join the group but allow the men to continue with the beating, 1 = I would join the group in beating up the thief.
- *Join mob (only endline):* Suppose you are on your way home. In your street, you encounter a group of [at random: 10/50] community members. The community members are beating a man who has been caught stealing from your neighbor's yard. Would you join the group? 1 = Yes, 0 = No

## D.2 Table 3

*Columns 1 and 3: Police know HH.* This measure is an index of two items:

- *Know Name:* Thinking about the police that work in your community. Do you think that someone from the police knows your name? If no: Do you think the police knows the name of someone else who lives in this household? 0 = Respondent answered no to both questions, 0.5 = Respondent said no to the first question but yes to the second, 1 = Respondent said yes to the first question.
- *Know House:* Do you think someone from the police knows your house? 0 = No, 1 = Yes

*Columns 2 and 4: Police are motivated.* I am now going to read you several statements. Please tell me which one comes closest to your view. 1 = Statement 1: If the police do not respond to incidents of crime in time, it is because they do not have enough cars, 0 = Statement 2: The police have enough cars and if they do not respond in time, it is because they cannot be bothered to do their jobs.

### D.3 Table 4

*Columns 1 and 4: Respond MV* Suppose such an incident [an incident of mob vigilantism] did happen in your street. Do you think the police would hear about the incident? If Yes: Will they be alerted while the incident is happening or will they hear about it later? If Yes: And are the police likely to arrive while the community members are still beating the criminal? 0 = The police would not hear about the incident, 0.33 = The police will hear about the incident but later, not while it is happening, 0.66 = The police will hear about the incident while it is happening but not arrive while the community members are still beating the criminal, 1 = The police will hear about the incident whilt it is happening and arrive while the community members are still beating the criminal.

*Columns 2 and 5: Imprison MV.* Which of the following statements comes closest to your view? 1 = Statement 1: The police do everything they can to ensure that those who take the law into their own hands receive a prison sentence, 0 = Statement 2: The police do not care much about sending those who take the law into their own hands to prison.

*Columns 3 and 6: Service Index* This outcome measure is an index of three items:

- *Take problem seriously:* When you or someone like you takes a problem to the police, how likely is it that the police take your problem seriously? 1 = Very likely, 0.5 = Somewhat likely, 0.25 = Not very likely, 0 = Not likely at all
- *Send guilty to prison:* Which of the following statements comes closer to your view? 1 =Statement 1: The police ensure that people who are guilty almost always go to prison, 0 = Statement 2: The police often let people who are guilty go free.
- *Arrive quickly:* Imagine you are at home and alert the police in an emergency. Do you think the police would come to your help?
  - If Yes or Maybe: Do you think the police would take more or less than an hour to come to your help? If you don't know, please give your best guess.
    - If More than an hour:
      - Do you believe the police would take more than two hours or less than that?
    - If Less than an hour:
      - Do you believe the police would take less than 30 minutes or more than that?
  - 0 = The police would not come
  - 0.25 = The police take more than two hours
  - 0.5 = The police would take more than one hour but less than two
  - 0.75 = The police would take less than one hour but more than 30 minutes
  - 1 = The police would take less than 30 minutes

### D.4 Table 5

*Columns 1 and 2: Believes police fight crime.* And finally, what about these two statements? 1 = The police do everything they can to ensure that criminals receive the punishment that they deserve, 0 = The police do not make much of an effort to ensure that criminals receive the punishment that

they deserve.

*Columns 3 and 4: Believes police fight MV.* Finally, which of the following do you believe the police will do? 1 = The police will do all they can to send those who beat the criminal to prison, 0.5 = The police may make some efforts to send those who beat the criminal to prison, but they will not try very hard, 0 =The police will not do anything to send those who beat the criminal to prison.

*Columns 5 and 6: Would participate MV.* This outcome is an index of the following two items:

- *Join Beating 3:* Some people we speak to say that they would definitely participate in beating a criminal if the community were to catch one. Others say that they would not participate in the physical punishment of a criminal. Which comes closest to your view? If would not participate: What if the criminal had hurt someone you know. Would you participate in beating the criminal? 0 = No, I would never participate, 0.5 = I would participate only if the criminal hurt someone I know, 1 = Yes, I would participate.
- *Join Beating 4:* Suppose someone in your community is known for breaking into the houses of old women. One day, your neighbors catch the guy red-handed as he is breaking into the house of an old lady in your street. A group of community members surrounds the thief and they start to beat him. Which of the following are you most likely to do? 1 = I would join the group in punishing the criminal, 0.5 = I would stay and watch but would not join the group, 0 = I would leave the scene.

## D.5 Measures of prior beliefs

### Alarm Treatment

Prior belief measures are taken from the baseline survey. Since only one woman was interviewed per household at baseline, prior belief measures are treated as household level measures.

*Prior beliefs about legal repercussions for MV:* Suppose such an incident (an incident of mob vigilantism) did happen in your community. How likely is it that the police would hear of the event and arrest the people who [beat/killed] the accused? High prior (1 =) "Very likely" or "Somewhat likely," Low prior (0 =) "Not very likely" or "Not likely at all"

*Prior beliefs about police service quality:* This item is an index using the following measurements:

- *Customer service:* When you or someone like you takes a problem to the police, how likely is it that the police [at random: take your problem seriously/ appear to know what they are doing]? 1 = Very likely, 0.5 = Somewhat likely, 0.25 = Not very likely, 0 = Not likely at all
- *Arrive quickly:* Imagine you are at home and alert the police in an emergency. Do you think the police would come to your help?
  - If Yes or Maybe: Do you think the police would take more or less than an hour to come to your help? If you don't know, please give your best guess.
    - If More than an hour:
      - Do you believe the police would take more than two hours or less than that?
    - If Less than an hour:
      - Do you believe the police would take less than 30 minutes or more than that?

  - 0 = The police would not come
  - 0.25 = The police take more than two hours
  - 0.5 = The police would take more than one hour but less than two
  - 0.75 = The police would take less than one hour but more than 30 minutes
  - 1 = The police would take less than 30 minutes

- *Send guilty to prison:* Which of the following statements comes closer to your view? 1 =Statement 1: The police and the courts ensure that people who are guilty almost always go to prison, 0 = Statement 2: The police and the courts often let people who are guilty go free.

High prior (1 =) respondent's score falls strictly above baseline sample median of index. Low prior (0 =) respondent's score falls below baseline sample median of index.

### Information Treatments

Analyses that draw on information treatments only (Tables 5 and 23) use prior belief measures asked during the endline survey prior to the administration of information treatments. These measures are available for every respondent including neighbors. Since these measures were collected after alarm installations, I do not rely on them when analyzing the alarm and information treatments together.

*Endline beliefs (prior to information treatment) about legal repercussions for MV:* Which of the following statements comes closest to your view? High prior (1 =): Statement 1: The police do everything they can to ensure that those who take the law into their own hands receive a prison sentence, Low prior (0 =): Statement 2: The police do not care much about sending those who take the law into their own hands to prison.

*Endline beliefs (prior to information treatment) about police service quality:* Which of the following statements comes closest to your view? High prior (1 =): Statement 1: The police ensure that people who are guilty almost always go to prison, Low prior (0 =): Statement 2: The police often let people who are guilty go free.

## D.6 Table 13

*Spoken to police (Midline).* I would like you to think about the last month. During this time, did you ever speak to someone from the police? 1 = Yes, 0 = No

*Spoken to police (Endline).* I would like you to think about the time since last Christmas. During this time, did you ever speak to someone from the police? 1 = Yes, 0 = No

## D.7 Table 15

*Column 1,2,5,7: Support MV* An index of five items at midline and two items at endline:

- *Not arrest mob:* Sometimes communities beat criminals to death and then the police begin to investigate. Do you think the police should arrest community members who beat criminals to death? 0 = Yes, 1 = No

- *Beat known thief:* Someone in your community is known to be involved in stealing cars and plasma TVs. One day, the community catches him red-handed as he is breaking into a house. Which of the following do you believe the community members should do? 0 =

The community should call the police and leave it to them to deal with the thief, 1 = The community members should beat the thief there and then.

- *Beat petty thief (only midline):* Finally, imagine the following: A [at random: man/woman] from your community is blowing the whistle, because [he/she] saw someone stealing food and a box of cold drinks from [his/her] yard. The neighbors come running and one them gets hold of the thief. Again, which of the following do you believe the neighbors should do? 0 = The neighbors should call the police and leave it to them to deal with the thief, 1 = The neighbors should beat the thief there and then.

- *Beat driver (only midline):* Imagine the following situation: A truck driver drove 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? 0 = The group should leave it to the police to investigate, 1 = The group of men should beat the truck driver to teach him a lesson.

- *Community deal crime (only midline):* Some people think that, if people want to stop crime in their neighborhood, it is best for community members to deal with criminals themselves. Others think that these matters are best left to the police. Which comes closest to your view? 1 = Community members should deal with criminals themselves, 0 = These matters are best left to the police.

*Column 3,4, 6 and 8: Call Comm.* This measure is an index of two items:

- *Alert community:* What about your neighbors and other community members. If someone is about to enter your home to steal from you, would you reach out to the community for help? If YES: Would you want to alert the entire community or just the people you know best? 0 = No, 0.5 = People I know best, 1 = Entire community.

- *Alert neighbors:* Imagine you come home and you see a burglar leaving your Yard. Would you want to alert your neighbors [at random: even though, if the community gets hold of the man, they may beat him very severely]? 1 = Yes, 0 = No

### D.8 Table 14

*Police would discover.* This outcome measure is an index of the following two items:

- *Discover stolen car:* We do not mean to say that you would ever do something like this. However, suppose you bought a stolen car and you tried to hide it from the police. How likely do you think it is that the police would find out about that? 1 = Very likely, 0.5 = Somewhat likely, 0.25 = Not very likely, 0 = Not likely at all

- *Discover illegal immigrant:* Again, we do not mean to say that you would ever do something like this.However, suppose you had a tenant who is an illegal immigrant without papers and you want to hide that from the police. How likely do you think it is that the police would find out about that? 1 = Very likely, 0.5 = Somewhat likely, 0.25 = Not very likely, 0 = Not likely at all