

How the State Discourages Vigilantism - Evidence from a Field Experiment in South Africa*

Anna M. Wilke[†]

March 21, 2024

9,987 Words

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 participate in 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>.

[†]New York University, amw703@nyu.edu

Governments rely on citizen cooperation to function effectively. Examples include tax compliance, obedience to health regulations, and crime reporting. Yet, in many contexts, citizens are reluctant to cooperate with the state and turn to informal alternatives like gangs, traditional healers, or vigilante mobs. A large literature points to state performance as the source of citizen cooperation. Work on legitimacy suggests citizens rely on state institutions if they effectively deliver services (Risse and Stollenwerk, 2018; Brinkerhoff, Wetterberg, and Dunn, 2012); Levi (1988, 1997) argues trustworthy governments elicit “quasi-voluntary compliance” and Tyler (2003) holds citizens cooperate with institutions they perceive as procedurally just. The idea of a virtuous cycle – high quality service provision leads citizens to voluntarily turn to the state, which in turn strengthens state institutions – has informed major state building initiatives (Zoellick, 2008).

Empirical evidence on the link between state capacity and informal alternatives is mixed, however. Some find strengthening the state weakens informal provision (Acemoglu et al., 2020; Blair, Karim, and Morse, 2019), but expanding state services can also strengthen non-state actors (Cooper, 2019). One reason may be that, for a virtuous cycle to exist, citizens need to demand state services and see them as substitutes for informal ones. Yet, these services often differ in fundamental ways. Traditional healers may supply herbal remedies that vaccine-hesitant citizens prefer to government-mandated vaccines (Mugari and Obioha, 2021). Vigilante punishments are harsher than state sentences and sidestep due process rights of criminal suspects (Smith, 2019). If citizens inherently prefer informal to state services, the state may not be able to out-compete informal provision by becoming more effective. At the same time, strong states also use their capacity to regulate. States can declare an informal alternative illegal and dampen its supply. Where the taste for informal services is widespread, states may overcome informal alternatives through coercion.

This paper sheds new light on the relationship between state capacity and informal alternatives by focusing on law enforcement as a core state function. I investigate how an increase in police capacity shapes citizens’ willingness to rely on the state and participate

in mob vigilantism. In many contexts, spontaneously formed groups of citizens physically “punish” criminal suspects, often through gruesome assaults (Jung and Cohen, 2020). I conduct a field experiment in South Africa which has one of the highest rates of vigilantism worldwide even though its police force has above average capacity for Sub-Saharan Africa (UNODC, 2015; SAPS, 2022). A history of violence and strained citizen-police relations under the Apartheid regime are often cited as reasons (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 are historically entrenched, the same may hold where vigilantism arose more recently.

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. I leverage a similar shift, though on a much smaller scale. Together with a South African non-profit organization, I randomly assigned 100 of 250 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 common to South African townships and slums in other contexts, where police face a confusing street layout and a lack of street names and lights. Many citizens believe police would never arrive when called and are hesitant to rely on them. The alarm seeks to improve police’s familiarity with and ability to locate households, which may help police respond to burglaries that are the most frequent crime in the study precinct and a common trigger of mob vigilantism. Involvement in the alarm project may also signal 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 participate in vigilantism. The effects are concentrated among subgroups who were a priori pessimistic about police and relatively more willing to inflict harm on suspects. These results support the intuition that strengthening state institutions can reduce the popularity of informal alternatives and increase cooperation with the state.

I then explore the mechanisms behind these effects. A first mechanism corresponds to the logic of competition between service providers that underlies the virtuous cycle idea. More capable police may provide higher quality services. E.g., access to information from the alarm may allow police to quickly find a household which increases the likelihood of an 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. Even though South African police sometimes look the other way, 53% of baseline respondents in this study think it likely that vigilantism perpetrators would be arrested. 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 that owners' 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 participating in vigilantism.

I use three strategies to disentangle these mechanisms. First, I show 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. Both effects are concentrated among a priori pessimistic respondents who saw the biggest shifts in their willingness to rely on police and vigilantism. Second, I report results from an 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 alternative explanations including the worry that results may be driven by experimenter demand.

These findings demonstrate the relationship between the state and informal actors is not limited to competition. Where the state has declared an informal alternative illegal, increased police presence may induce citizens to abandon this alternative out of concern about state punishment. This insight illuminates the conditions under which state capacity may be most effective at weakening informal alternatives. Such alternatives are not always illegal. Rural traditional courts and chiefs, for example, are often recognized by the state (Cooper, 2019; Baldwin, 2016). If citizens have a taste for informal services and 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 deems 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 on legibility initiatives like biometric identification systems (Muralidharan, Niehaus, and Sukhtankar, 2016; Bossuroy, Delavallade, and Pons, 2019), highlights the benefits of effective service delivery. Yet, even democratic states also use their capacity for regulation. A shock to police capacity caused a priori pessimistic citizens to expect both better services and a greater risk of state punishment for illegal behavior. Vigilantism is one example of an illegal practice with 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

Formal justice institutions often co-exist with informal ones. The latter sometimes consist of non-state actors like chiefs or gangs (Baker, 2008). I focus on settings where punishments are meted out by ordinary citizens. 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. (...) 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; Kirsch and Grätz, 2010). Ordinary citizens directly participate in mob vigilantism by apprehending and inflicting violence on suspects. Anecdotal evidence suggests such participation can be widespread. Many men in the study precinct alluded in conversations to having beaten suspected criminals. Examples of women and even children inflicting harm exist as well. As a South African activist recalls: “[A]s a child you also run after the person and we also just start pulling on the person while the adults are hitting him, so the children are also helping.” (Khayelitsha Commission, 2014, p.111). Beyond those who mete out violence, vigilantism is fueled by citizens who support it in indirect ways. Someone typically calls the community for help, and incidents draw crowds of spectators who cheer on the violence. The presence of a crowd provides social support and protection for perpetrators, since police find it difficult to identify and link individuals to the harm inflicted (Roelf, 2012).

When asked why they participate in or support vigilantism, respondents in qualitative interviews often point to the police. One 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’ behavior change if police became more capable?

I presume citizens wish for some transgressions to be punished due to a desire for deterrence or vengeance (Becker, 2000). Qualitative evidence suggests four considerations that

may affect a citizen's choice between reporting to the police and rallying her community.

Nature of offense. State and community punishment may not both be viable. 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 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 like theft and burglary and violent crimes like 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. Many South Africans cite its harsh and public nature as an advantage of vigilantism. More than 50% of South Africans are dissatisfied with the courts, the most common complaint being sentences are too lenient (StatsSA, 2016/2017).

Probability that suspect will be punished. Community members are well positioned to apprehend suspects, because they are spatially proximate to the crime scene and well informed. On the state's side, more capable police may be better able to apprehend perpetrators and investigate crimes that citizens report. Of course, police do not alone determine the likelihood of conviction. Other institutions like courts also matter. Democratic justice systems typically uphold a set of due process protections, while communities often mete out punishment even if evidence of guilt is tenuous. The probability of a suspect being punished by a democratic state may then, by design, be lower than the probability of community punishment, even when police are highly effective. Nonetheless, police as frontline service providers loom large in the minds of citizens. Respondents in open-ended interviews frequently framed the perceived inability of the state to sanction perpetrators as police ineffectiveness, even when the hands of police may have been tied by a lack of evidence.

Risk of punishment for vigilantism. Vigilantism amounts to serious crimes like assault or

murder. Victims thus have legal recourse. A victim in the study precinct opened an assault case against his attackers. Despite the challenges involved in investigating vigilantism, arrests and convictions occur. In the study precinct, 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. Again, police do not alone determine the legal fate of vigilantism perpetrators, but seem to play an important role in citizens' perceptions of punishment risks. One respondent described such risks by pointing out: "The police do not want us to beat up criminals." An increase in the perceived risk of state punishment may be most relevant to the decision to inflict 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 tend to be the instigator's immediate community. Even the decision to instigate vigilantism may thus be affected by an increase in the risk that one's friends or family could be arrested.

This discussion suggests police capacity 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 perceived 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 perceived probability that participation in mob vigilantism leads to state punishment.

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 service delivery. 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 the control group in this study opposes prison sentences even for vigilantes who killed a criminal suspect.

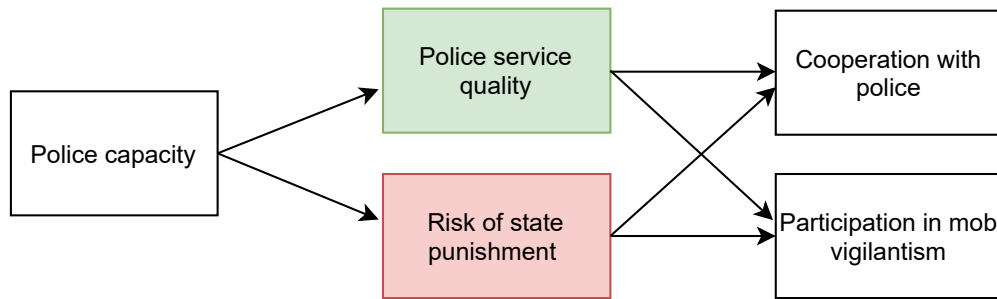


Figure 1: Police capacity and the choice between the state and mob vigilantism.

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 *out-compete* vigilantism. According to this logic, citizens 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 citizen compliance (e.g., [Risse and Stollenwerk, 2018](#); [Levi, 2019](#)) and relies on the assumption that citizens perceive state and non-state justice as substitutes. Recent work on vigilantism casts doubt on this assumption. Vigilante mobs punish offenses like witchcraft, mete out harsher punishments than state courts and do not protect the rights of suspects. [Smith \(2019\)](#) argues these characteristics drive support for vigilantism in South Africa. If so, the state may not be able to out-compete vigilante mobs by becoming more effective. The second mechanism suggests state capacity may nonetheless help the state to counter vigilantism, because it allows the state to *rule out* informal alternatives that it deems illegal. [Jaffrey \(2023\)](#), e.g., links vigilantism in Indonesia to the state’s inability to punish vigilantes. Rather than voluntarily, citizens may cooperate with the state if it can limit their options.

The two mechanisms also differ in their relevance for different kinds of involvement in vigilantism. The second mechanism suggests police capacity may most strongly affect the decision to mete out violence. Citizens’ willingness to support vigilantism in indirect ways may be affected through this mechanism as well but only if citizens perceive these ways to

carry legal risks or care about risks for others. Conversely, if an increase in police capacity convinces citizens that police services are so good that vigilantism is not needed anymore, citizens should abandon all kinds of involvement.

Whichever mechanism is at play, the effects of a capacity shock may be moderated by citizens' prior views about the police. Crime is a rare event and police capacity not directly observable. Hence, citizens' decisions are likely shaped by perceived rather than de facto capacity. Research suggests prior beliefs matter for how citizens update this perception as a result of a de facto capacity increase. E.g., shifts in government performance were found to affect views on legitimacy more strongly when changes were subjectively perceived as improvements over the status quo (McLoughlin, 2015), and prior beliefs moderate the effects of government performance information on vote choice (Arias et al., 2022). Citizens' ex ante perceptions of police may reflect personal experience with the state, observations of police operations, conversations with others and news stories. Where prior perceptions are heterogeneous, a capacity increase may most strongly affect a priori pessimistic citizens to whom it comes as a surprise. If perceptions of police capacity are important determinants of beliefs about the likelihood that a suspect is sanctioned through the state and the risk of state punishment for vigilantism, a capacity increase should then most strongly affect the behavior of citizens with low ex ante expectations about these police outputs – exactly the groups that one would expect to be a priori least likely to draw on the police and most likely to participate in vigilantism.¹ Moreover, since both are shaped by perceptions of capacity, citizens' views about these outputs will be correlated. Citizens who are a priori pessimistic about one output may then increase their perceptions of *both* outputs as a result of a positive capacity shock.

¹Appendix section A.3 formalizes this argument.

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 created under the Apartheid regime. This study takes place in a semi-urban, predominantly black, low-income township in the Northwest Province which the implementing partner and police selected because of a high rate of burglaries that the intervention may help address. Crime statistics from 2016 suggest burglaries at residential premises are the most common crime in the precinct, with a rate almost twice as high as that of the median precinct.

[Scott \(1998\)](#) argues communities have to be “legible” for state institutions to function effectively. Townships were designed to be easily policed, but tend to be difficult to “read” today. Many have grown substantially through the expansion of informal settlements. Street names are rare and houses are numbered within sections comprising thousands of dwellings. 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 their emergency response.

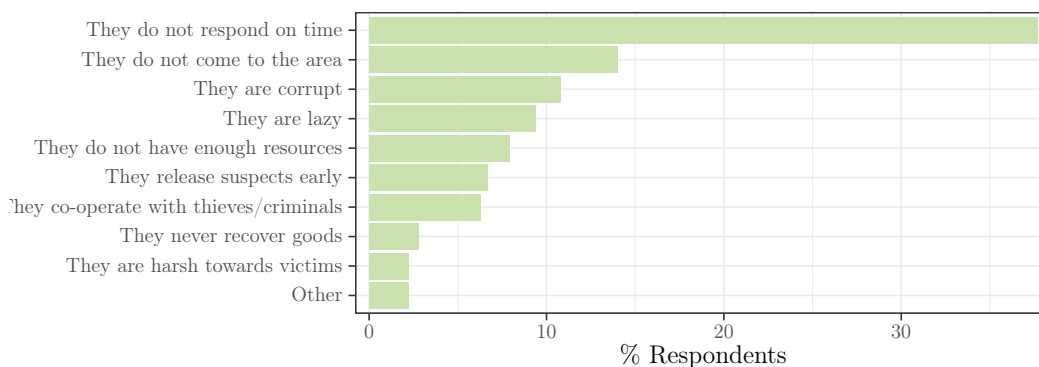


Figure 2: Slow response is main reason for dissatisfaction with police ($N = 8,906$)

Calculated among 43% of $N = 21,095$ respondents in a nationally representative sample who are dissatisfied with police [StatsSA \(2016/2017\)](#). Question asks about main reason for dissatisfaction.

Dissatisfaction with police is widespread. Figure 2 shows slow response times are nationwide the most prevalent grievance. 55% respondents in this study’s control group believe police would never come or take longer than two hours when called to an emergency.

Mob vigilantism is the primary alternative to state justice in the study precinct. Figure A1 in the appendix shows at least a quarter of respondents in most areas recalled one vigilante incident or more in their area between May and July 2018. In qualitative interviews, most respondents could describe at least one case that typically led to severe injuries or even the death of the accused. Anecdotes in which the accused was caught stealing household items from a residential dwelling were particularly common. The vast majority of incidents occurred in residential areas of the township. In fact, household surveying for this study had to be interrupted twice because enumerators suspected a vigilante incident may occur. Around 14% of control group respondents – 17% of men and 12% of women – report they would participate in beating a man who was caught stealing from a neighbor’s yard. This gender gap may seem surprisingly small but resonates with anecdotes about the broad participant base of mob vigilantism in South Africa (Khayelitsha Commission, 2014) and evidence that women are often particularly supportive of vigilantism (Wilke, 2023).

The approach of police to vigilantism is ambiguous. There is no shortage of anecdotes about police turning a blind eye, but arrests occur and appear to be highly salient for citizens. E.g., in qualitative interviews, all respondents from one part of the study precinct were aware that several men from this area were in prison for killing a suspected thief. Such incidents may shape citizens’ prior perceptions of the police.

Table 1 displays the joint distribution of baseline perceptions of police service quality and the risk of state punishment for vigilantism. The largest share of respondents – about one third – expect little on both dimensions. Around 16% have above median service quality expectations but do not think vigilantism perpetrators would be arrested, while about 27% consider legal repercussions for vigilantism likely but have little hope that police provide high quality services.

Risk of punishment MV	Police service quality		Total
	Low	High	
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 vigilantism perpetrators would be arrested. Respondents with high (low) service quality perceptions fall above (below) the median of an index of *Customer Service*, *Arrive quickly* and *Send guilty to prison*. See appendix section D.5 for question wording.

2.2 Intervention

The implementing partner developed the alarm together with the police. The alarm is installed in the house and can be triggered via a panic button, motion sensor or cell phone.² 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.³ Text messages indicate alarm owners’ names, phone numbers, and landmarks close to their 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. The partner registers all alarm panics through a back-end system.

Figure 3 depicts how the alarm may alter the de facto and perceived capacity of police to intervene in a household. First, the alarm aims to make households “legible” to police, i.e.,

²Households without electricity received a solar panel that only powers the alarm. The alarm has a 24-hour battery.

³Households can nominate two neighbors to receive text messages, but I find no evidence that the alarm changed community relations.

easier to locate and identify. Alerting police without an alarm requires calling a centralized emergency hotline or police station. Without reliable addresses, it can be difficult to explain one’s location to a call center agent unfamiliar with the 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 could find protected households from memory alone. Alarm recipients may thus expect a faster police response, which may increase the chance that perpetrators are apprehended and convicted.

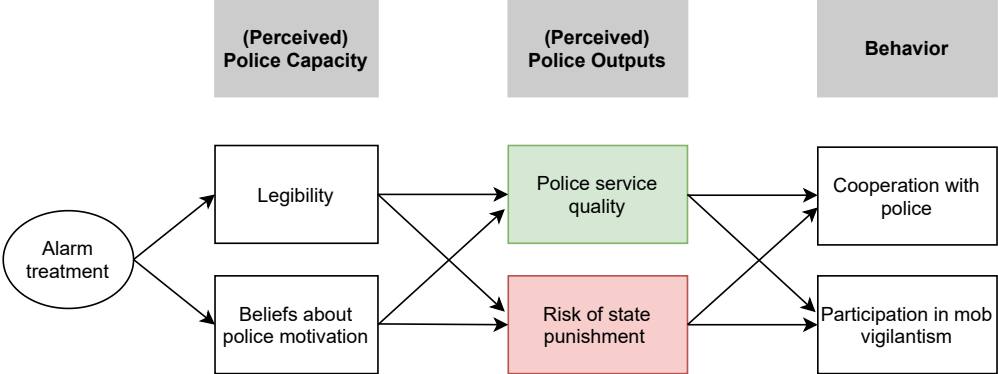


Figure 3: Hypothesized effects of police alarm.

Being known to police is beneficial if one reaches out for help, but may cause worry if one considers breaking the law. Many study participants reside in informal settlements and are 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 resulting shift may in part be psychological. Moreover, police often described an “anonymous crowd” that is unwilling to testify as an obstacle for investigations of vigilantism. Being known to police may help police identify household members in a mob situation. Hence, alarm recipients may worry about the legal consequences of participating in vigilantism.

Alarm owners may also update their more general views about 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, but motivated police as a welcome

way to improve performance. Police involvement in the project may thus act as a signal of police motivation.⁴ Learning may also occur through increased interactions with police. 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.

The alarm may thus affect the choice between the state and vigilantism through the two mechanisms outlined above. The belief that police can respond fast and are highly motivated may make reliance on the state more attractive. Importantly, the alarm may help police address burglaries, a common crime and trigger of vigilantism in the study context. Yet, being “known” to police who take their job seriously may also increase the fear of state punishment for vigilantism. As discussed above, these effects may be particularly strong for citizens with low prior expectations. I thus pre-registered sub-group analyses by prior beliefs about police service quality and the risk of state punishment for vigilantism.

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 community meetings. The remaining 115 households were chosen from a pool created by geo-locating every tenth house in the precinct’s eleven most high crime areas.

Households were selected in non-random ways to limit non-compliance, attrition and interference. To limit interference, I chose the largest sample such that each household is no closer than 150m to all other sampled households. Due to location inaccuracies, only 67% of households satisfy this constraint. To limit non-compliance, the sample excludes 27 households who were not interested in the alarm at baseline.⁵ To limit attrition, the sample

⁴Other residents may learn about the alarm project and draw the same inference. Yet, surveys with neighbors showed no evidence of such effects.

⁵Uninterested respondents were particularly pessimistic about police. Since treatment

excludes 77 households that were interviewed at baseline but could not be reached during back-checks. At baseline, I interviewed the woman most involved in household decisions in each household.⁶ Appendix section A.6 provides more details on sampling and section A.10 provides descriptive characteristics.

2.4 Random assignment

Households were organized into 50 blocks of 5. After dividing households into two groups by how they were sampled, I chose blocks to minimize the within-block Mahalanobis distance of four variables: baseline support for and the willingness to participate in vigilantism and the household's latitude and longitude. 100 households, two in each block, were assigned to the alarm treatment. 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. At midline, 93 of 100 households in the treatment group and none in control had received an alarm. Four households 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.

This widespread interest in the alarm may seem at odds with the argument that the alarm causes fear of state punishment for vigilantism. Why would respondents who support vigilantism agree to such a treatment? One explanation is that respondents are willing to trade off their freedom to participate in vigilantism for improved police protection against heavily armed perpetrators whom communities cannot confront. How citizens resolve this effects appear concentrated among pessimistic respondents, excluding uninterested respondents should, if anything, result in estimates that are too small (in absolute value).

⁶This rule ensured respondents could confirm their household's interest in the alarm. The implementing partner was most interested in women. Budget constraints prevented surveys with two household members at baseline. Mid- and endline surveys interview a woman and a man per household where available.

trade-off may depend on their beliefs about police services and taste for vigilantism. Figure A4 in the appendix shows those not interested in an alarm were 45% more likely to participate in vigilantism and less optimistic about the police’s emergency response at baseline.

Figure A5 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 2019.⁷ 72 of the 94 protected households triggered their alarm at least once. Not all panics are caused by crime. E.g., 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. 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 household surveys, respectively, one and eight months after treatment roll-out. The same respondents were interviewed at mid- and endline: the women sampled at baseline and one randomly selected man per household. In all-women households, a second woman was randomly selected. Since 23 of 250 households have only one member, the target sample size was $N = 477$. Response rates were 92% ($N = 438$) at midline and 85% ($N = 407$) at endline. Appendix section B.2 shows attrition seems unaffected by treatment. Additional respondents were interviewed if other household members were available during the interview. 45 respondents were added at midline and 39 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. Appendix section B.1 provides evidence of covariate balance. I impute missing values due to item non-response using multivariate imputation via chained equations within pre-specified families (e.g., “vigilantism related outcomes”) and create indices by averaging constituent items.

⁷Multiple panics in a household on the same day are collapsed into one incident.

2.7 Estimation and hypothesis tests

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

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

\mathbf{Y} is a vector of outcomes, α an intercept, τ the sample ITT, \mathbf{z} a vector of treatment assignments, \mathbf{n} a vector storing the number of respondents per household with associated coefficient δ , 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 may be biased if cluster size correlates with potential outcomes.

To estimate conditional ITTs and differences between them, I add 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. I rely on household-level measures of prior beliefs collected from one respondent per household at baseline. See appendix section [D.5](#) for the question wording and [Table 1](#) for the distribution of prior belief measures.

Results from a second pre-registered specification that controls for covariates selected through lasso regression are 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) effect for any unit. Appendix section [A.1](#) summarizes divergences from the pre-analysis plan.

2.8 Ethics

Appendix section [A.4](#) describes in detail the steps I took to address three sets of ethical concerns – the potential for adverse consequences of the intervention, the minimization of risks for respondents during data collection and the protection of study staff.

3 Main Results

Table 2 shows the alarm treatment seems to have increased citizens’ willingness to rely on police and decreased their willingness to participate in 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 proclivity to share crime-related information with police. Columns 1 and 2 suggest the alarm increased the willingness to rely on police by roughly one third of a control group standard deviation at midline and by around one quarter at endline ($p < 0.01$). Analyses in columns 5 and 7 allow effects to vary across prior beliefs. 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 $p < 0.1$).

Despite the apparent increase in respondents’ willingness to rely on police, Table A9 in the appendix provides no evidence of an increase in the share of respondents who recently spoke to police. This finding seems surprising. A possible explanation is that the alarm reduces the need to reach out to police, because it deters crime. Respondents in the treatment group indeed feel safer, but evidence of a reduction in victimization is limited (see Table A21). Appendix section A.5 provides a more detailed discussion of these results.

Turning to vigilantism, I first focus on respondents’ willingness to directly participate, since this outcome may be affected through both hypothesized mechanisms. I use one item at midline and an index of the same and another item at endline. Both measures elicit what respondents would do if the community apprehended a suspect. The item included at mid- and endline asks respondents to imagine the suspect had broken into their house and involves three ordered options reflecting increasing levels of participation: advocate for

handing the suspect over to police, let others beat the suspect but do not participate, or personally participate in the beating. The item added at endline asks whether respondents would join in inflicting harm on a suspect who stole from their neighbor.

Column 3 of Table 2 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 statistically insignificant. Perhaps respondents revert their perceptions of the police downwards as the novelty of the alarm wears off. As a result, effects may become undetectable, especially among respondents who were less affected to begin with. Once I allow for effect heterogeneity across prior beliefs, I find evidence of a negative effect among a priori pessimistic groups 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 by about one third of a control group standard deviation ($p < 0.05$). 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 but less pronounced ways (see column 8).

Table A14 in the appendix shows results are similar for both endline index components and Table A11 suggests effects on the outcome used at mid- and endline are driven by a reduction in the share of respondents who would personally participate in harming the suspect. Re-coding this outcome as a binary indicator for whether respondents chose this option over the alternatives to let the beating happen or advocate for police involvement, I find the alarm decreased the share of respondents who would join by around eight percentage points at midline ($p < 0.01$) and almost five percentage points at endline ($p < 0.1$).

	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*** (0.028)	0.075*** (0.031)	-0.078*** (0.032)	-0.012 (0.028)	0.132*** (0.044)	-0.100** (0.044)	0.126*** (0.042)	-0.042 (0.037)
Alarm \times High Prior Punishment					-0.108** (0.062)	0.158*** (0.057)		
Alarm \times High Prior Service							-0.104* (0.061)	0.066 (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 <i>p</i> -value Diff.	-	-	-	-	0.034	0.002	0.061	0.206
Hypothesis Diff	-	-	-	-	lwr	upr	lwr	upr
Number H _H s	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 participate in mob vigilantism.

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

The alarm seems to have discouraged vigilantism in the sense that it decreased respondents' willingness to directly participate. Effects on the willingness to indirectly support vigilantism appear less consistent (see Table A10). There is some evidence that the alarm reduced support for the participation of others at midline and among respondents with low prior beliefs about the risk of state punishment, but alarm owners do not appear less inclined to reach out to neighbors if someone enters their home to steal from them. This pattern provides a first clue regarding mechanisms. Merely alerting one's neighbors is not necessarily a crime. If the alarm's effects result from increased concerns about state punishment, the apparent absence of an effect on this outcome seems intuitive.

Taken together, the results suggest the alarm encouraged reliance on police and discouraged participation in vigilantism, especially among respondents who were pessimistic at baseline. Appendix Tables A3 and A4 show these respondents were a priori less likely to rely on police and more willing to inflict harm on suspects. These patterns support the notion that police capacity helps the state supersede informal alternatives.

4 Mechanisms

I now explore *how* the alarm produced its effects. I provide evidence using three empirical strategies each of which has its own advantages and limitations. Online appendix section A.3 formalizes how each strategy relates to my theory.

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

I first analyze the alarm's effects on potential mediators. Because the treatment's main effects appear concentrated among households who were a priori pessimistic, I estimate conditional effects on mediators as well. I report estimates at endline in the main text and estimates of unconditional effects at mid- and endline in online appendix section C.4.

Did the alarm indeed affect perceptions of police capacity? Table 3 presents evidence of the alarm's effects on perceptions of legibility and police motivation as a manipulation check. The outcome in columns 1 and 3 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 (1)	are motivated (2)	know HH (3)	are motivated (4)
Alarm	0.114** (0.068)	0.216*** (0.068)	0.077* (0.061)	0.186*** (0.063)
Alarm \times High Prior Punishment	-0.00003 (0.091)	-0.179** (0.095)		
Alarm \times High Prior Service			0.093 (0.090)	-0.109 (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 prior belief measures and Table 1 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. The alarm seems to increase the share of respondents who do not think a slow police response would result from a lack of motivation by roughly 20 percentage points among both low prior

groups ($p < 0.01$). The interaction terms suggest effects among those with high priors 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, but that on beliefs about police motivation seems concentrated among pessimistic respondents.

Next, I turn to the potential downstream effects of these changes. Did the alarm affect 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 respondents with high prior expectations.⁸

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 A16 in the appendix 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.

Respondents assigned to an alarm thus expect more from police, in terms of both service quality and punishment for vigilantism. These effects appear concentrated among the low prior groups that also saw the greatest shifts in the willingness to rely on police and participate in vigilantism.

⁸I also find evidence of a small upward shift in the perception that police would find out about illegal behaviors other than vigilantism (see appendix section C.4).

	<i>Risk of state punishment</i>		<i>Service quality</i>		<i>Risk of state punishment</i>		<i>Service quality</i>	
	Respond MV	Imprison MV	Service index	Respond MV	Imprison MV	Service index		
	(1)	(2)	(3)	(4)	(5)	(6)		
Alarm	0.043 (0.051)	0.091* (0.066)	0.085** (0.044)	0.023 (0.045)	0.110** (0.055)	0.085** (0.037)		
Alarm × High Prior Punishment	−0.074 (0.070)	−0.075 (0.088)	−0.094* (0.059)					
Alarm × High Prior Service				−0.040 (0.073)	−0.145** (0.089)	−0.091* (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 <i>p</i> -value Main	0.235	0.07	0.034	0.262	0.032	0.021		
Hypothesis Main	upr	upr	upr	upr	upr	upr		
RI <i>p</i> -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 prior belief measures and Table 1 their distribution. Appendix section A.2 provides details on model specification and testing, and appendix section D.3 on outcome question wording.

Moreover, expectations about one police output seem to condition effects on perceptions of another. E.g., respondents with low priors about police service quality see larger effects on perceptions of *both* service quality and the risk of punishment for vigilantism. Exploratory analyses in appendix section C.8 suggest the alarm most consistently affected respondents with low priors beliefs about both potential mediators. This subgroup was also a priori most pessimistic about police capacity in terms of legibility and police motivations (see Table A4). These patterns support the idea that effects of changes in state capacity are moderated by citizens' prior capacity perceptions and that capacity perceptions shape expectations of both the benefit of relying on the state *and* costs of breaking laws.

4.2 Effects of information treatments that vary perceptions of police outputs

So far, results are consistent with both hypothesized mechanisms. It is possible, however, for the alarm to affect a potential mediator on average even if this variable does not mediate the relationship between the alarm and final outcomes (Green, Ha, and Bullock, 2010). Moreover, because shifts in the two potential mediators are concentrated in the same subgroups, analyses of heterogeneous effects cannot illuminate whether one mechanism is more important than the other. I instead use two information treatments, each designed to change one mediator but not the other, to investigate how effectively shifts in perceptions of service quality and punishment risks discourage vigilantism. Provided that the alarm and information treatments produce similar changes in the mediators, this exercise sheds light on the likely mechanism behind the alarm's effects (Ludwig, Kling, and Mullainathan, 2011).

The treatments consist of news articles read to respondents during their endline interview (see appendix section A.9). 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 vigilantism perpetrators. The "Police fight crime" treatment depicts police efforts to convict perpetrators of crimes against women and children, a service that is in high demand. Moreover, 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 (see appendix Table A1). As pre-registered, my analyses include the entire endline sample ($N = 815$), which encompasses two members of a neighboring household for each of the 250 study households.⁹ Enumerators were unaware of the goal to understand effects on subsequent responses and thought the aim was to elicit opinions about the articles using open-ended questions. Due to time constraints, I only measure perceptions of police efforts and the willingness to participate in vigilantism. I estimate effects of each treatment by regressing outcomes on its assignment indicator, marginalizing across the other information and alarm treatments.

	Believes police fight crime		Believes police fight MV		Would participate MV	
	(1)	(2)	(3)	(4)	(5)	(6)
Police fight Crime	0.116** (0.058)		-0.001 (0.052)		-0.0005 (0.049)	
Police fight MV		0.035 (0.059)		0.087* (0.052)		-0.083** (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. Sample includes respondents from main and neighboring households with low priors about the risk of punishment for vigilantism measured at endline. 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.

I pre-specified to analyze the information treatments within two low prior subgroups should they appear ineffective in the entire sample. Table A18 in the appendix suggests

⁹See appendix section A.7.

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.¹⁰

Column 1 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" ($p < 0.05$). The outcome in column 3 asks respondents whether police "won't do anything", "may make some effort", or "do all they can" to send vigilantism perpetrators to prison. There is little evidence that the "Police fight crime" treatment affects this outcome. Columns 2 and 4 show a similar pattern for the "Police fight MV" 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.

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. I measure this outcome using an index of two items. The first asks whether respondents would participate in beating a criminal, would do so only if the criminal had hurt someone they know or would always do so. The second asks whether respondents would leave the scene, stay and watch or join the beating of a man who was caught breaking into an old lady's house. Column 5 contains little evidence that the "Police fight crime" treatment affected this index. 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 A19 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 but does not seem to affect the willingness to participate in vigilantism. The "Police fight MV" treatment appears to shift neither beliefs about police effort nor the willingness to participate

¹⁰I here use endline measures of priors that were asked before the information treatments.

in vigilantism among this subgroup.

To summarize, even though the “Police fight crime” treatment seems to improve service delivery expectations in both subgroups, there is little evidence that it discouraged vigilantism. The “Police fight MV” treatment seems to increase 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 discourage participation in vigilantism. Hence, increasing the perceived risk of state punishment seems more effective at discouraging vigilantism than improving 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 distinct, but leverage theoretical expectations about how these treatments should interact if either 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 is an important mediator, the alarm should be particularly effective at discouraging vigilantism if combined with a treatment that convinces respondents police make efforts to deliver high quality services. Likewise, if punishment risk is an important mediator, the alarm should be particularly effective if combined with a treatment that convinces respondents police make efforts to convict vigilantes.

I thus regress the measure of respondents’ willingness to participate in 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 assigned to the respective other information treatment.

	Would Participate MV			
	All (1)	Police fight MV = 0 (2)	All (3)	Police fight crime = 0 (4)
Alarm	-0.045 (0.050)	0.036 (0.072)	0.034 (0.050)	0.036 (0.072)
Alarm × Police fight Crime	0.061 (0.067)	0.0002 (0.093)		
Alarm × Police fight MV			-0.097* (0.068)	-0.169** (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 respondents from main households. Columns 2 and 4 exclude respondents assigned to the respective other information treatment. “Control Mean” pertains to respondents assigned to neither the alarm nor the 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 strengthened this effect. Columns 3 and 4 similarly suggest the alarm did not affect the willingness to participate in vigilantism among respondents not assigned to the “Police fight MV” treatment. Yet, here, both interaction terms are negative ($p < 0.1$ and $p < 0.05$), suggesting the “Police fight MV” treatment made the alarm more effective at

discouraging vigilantism. 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 reflect 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 is not obvious why respondents would censor only some of their opinions about the police or why only respondents with low expectations would do so.

One may worry most about experimenter demand driving effects on vigilantism, especially because respondents’ willingness to inflict harm appears most consistently affected while effects on indirect support that is not as clearly illegal seems muted. 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 alarms were installed in September and October 2018. The treatment group remembering or witnessing fewer incidents would thus suggest a treatment-induced reluctance to be linked to vigilantism. I find no evidence of an effect on these outcomes (see appendix Table A20). Even more convincing would have been to ask about respondents’ participation in vigilantism prior to treatment, but doing so would have created legal risks for respondents.

Another concern is that effects reflect changes among the control rather than the treatment group. Respondents not assigned to an alarm may have become frustrated with police. Alternatively, police may have focused efforts on alarm owners, neglecting other households. Yet, Figure A6 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 community punishments. Even though alarm owners seem to feel safer, I find little evidence of a reduction in crime victimization or in respondents' demand for harsh and immediate punishments per se (see appendix Table [A21](#)).

6 Discussion

Many have suggested the prevalence of informal alternatives to the state reflects the state's inability to provide citizens with high quality 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 a set of households, I find greater police capacity indeed encouraged reliance on police and discouraged participation in vigilantism. Effects are concentrated among citizens who were a priori pessimistic about police and relatively more willing to inflict harm on suspects. However, these respondents developed both more sanguine views of police service quality and a greater sense that the state punishes vigilantes. An information experiment indicates the risk of state punishment plays a bigger role in the decision to participate in vigilantism than police service quality.

Why may improvements in police service quality have limited effects on vigilantism? Perhaps such effects 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 \(2019\)](#) provides evidence that South Africans resort to vigilantism because they dislike due process protections and wish for punishments to be harsher than those provided by state courts. These conditions likely extend to contexts other than South Africa. Figure [A2](#) in the appendix shows discontent with state punishments is widespread in Sub-Saharan Africa and Figure [A3](#) demonstrates investigative capacity and protections for the rights of criminal

suspects tend to be positively linked. Indeed, organizations promoting security sector reforms typically work to strengthen state capacity in conjunction with due process protections (OECD, 2005; USAID, 2017). The results presented here suggest that in such contexts, state capacity is likely to help the state discourage vigilantism by regulating its supply and not by reducing demand through high quality services.

Similar logics may apply to other informal practices. Traditional healers 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 sometimes prevail despite increased availability of formal credit (Tsai, 2004). Notwithstanding high interest rates, borrowers may prefer loan sharks, e.g. because they do not require formal contracts. A lower prevalence of these informal services in high capacity contexts may reflect the state’s ability to regulate suppliers.

A broader implication is that state capacity can have downsides for citizens. Those who favor illegal practices like vigilantism may be wary of increased state presence, even if it improves government services. 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 installed its siren out of sight. This request seems counterintuitive if the alarm only protects against intruders. If the alarm also deters vigilantism, however, households may want to hide the alarm to enjoy improved police services while upholding the threat of community punishment. Indeed, I find no evidence of spillover effects on neighboring households.

The alarm’s effects are concentrated among households who were a priori pessimistic about the police. Of course, prior beliefs are not randomly assigned. Reassuringly, respondents with high and low prior beliefs have relatively similar demographic characteristics (see appendix Tables A3 and A4). These groups also differ in intuitive ways. Respondents with low priors about service delivery and punishment risks are less likely to have been sampled

through the police and to know a state official. They are also almost twice as likely to have lived in their community for less than 10 years indicating they may reside in the more recently created informal parts. These patterns suggest superseding informal alternatives requires extending the state's reach towards citizens who are a priori least connected to it, including into ungoverned spaces like informal settlements. Doing so on a large scale may be costly due to infrastructural conditions and remoteness. These challenges may be one reason why vigilantism continues to flourish, even in a relatively strong state like South Africa.

Indeed, the treatment studied here is unlikely to be as effective if rolled out widely. If every household received an alarm, police would be less familiar with and able to attend to alarm protected households. Fortunately, the 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 small number of households discourages vigilantism provides evidence that a successful extension of its presence can help the state supersede informal alternatives, even if a larger alarm roll-out would do little to further the state's reach.

Open questions remain, however. First, I do not have evidence on whether the alarm prevented vigilante violence from occurring. On the one hand, reductions in the willingness to participate seem concentrated among respondents who were a priori most willing to harm suspects. That said, only few households received alarms and neighbors appear unaffected. Since the presence of a crowd complicates police investigations, vigilantism may follow a coordination logic where individuals are willing to participate once they expect others to do so ([Kuran, 1991](#)). If so, the effects of state capacity on vigilantism may be highly non-linear. Small increases may not have big effects but as state presence grows, there may be a tipping point where state punishment for vigilantism becomes salient enough such that citizens doubt anyone would participate. The strategic forces that shape how changes in the willingness to participate aggregate up into a reduction in violent incidents remains an important topic for future research. A second question concerns the incentives of state

officials. Given widespread support for informal alternatives like vigilantism, why would politicians and bureaucrats enforce laws against them? Another next step will be to study the behavior not only of citizens but also of state officials.

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.
- Arias, Eric, Horacio Larreguy, John Marshall, and Pablo Querubín. 2022. “Priors Rule: When Do Malfeasance Revelations Help Or Hurt Incumbent Parties?” *Journal of the European Economic Association* 20 (4): 1433–1477.
URL: <https://doi.org/10.1093/jea/jvac015>
- 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, health-care 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>

- Brinkerhoff, Derick W, Anna Wetterberg, and Stephen Dunn. 2012. "Service Delivery and Legitimacy in Fragile and Conflict-Affected States: Evidence from water services in Iraq." *Public Management Review* 14 (2): 273–293.
- 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.
- 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.

- Kuran, Timur. 1991. "Now Out of Never: The Element of Surprise in the East European Revolution of 1989." *World Politics* 44 (1): 7–48.
- Levi, M. 2019. "Trustworthy Government and Legitimizing Beliefs."
- Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley and Los Angeles, California: University of California Press.
- Levi, Margaret. 1997. *Consent, Dissent, and Patriotism*. Cambridge: Cambridge University Press.
- 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.
- McLoughlin, Claire. 2015. "When Does Service Delivery Improve the Legitimacy of a Fragile or Conflict-Affected State?" *Governance* 28 (3): 341–356.
- Miguel, Edward. 2005. "Poverty and witch killing." *The Review of Economic Studies* 72 (4): 1153–1172.
- Mugari, Ishmael, and Emeka E Obioha. 2021. "African beliefs and citizens' disposition towards COVID-19 vaccines: The belief guided choices." *African Journal of Governance and Development* 10 (1.1): 277–293.
- Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2016. "Building state capacity: Evidence from biometric smartcards in India." *American Economic Review* 106 (10): 2895–2929.

- OECD. 2005. "Security System Reform and Governance." *DAC Guidelines and Reference Series*. Paris: OECD Publishing .
URL: <https://www.oecd-ilibrary.org/content/publication/9789264007888-en>
- Risse, Thomas, and Eric Stollenwerk. 2018. "Legitimacy in Areas of Limited Statehood." *Annual Review of Political Science* 21: 403–418.
- Roelf, Wendell. 2012. "In South Africa's slums, mob justice rules." *REUTERS July 12, 2012* .
URL: <https://www.reuters.com/article/us-safrica-crime-idUSBRE86B0YM20120712/>
- 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
- 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.
- 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. 2003. "Procedural Justice, Legitimacy, and the Effective Rule of Law,," *Crime and justice* 30: 283–357.
- UNODC. 2015. Data on Police Personnel. United Nations Office on Drugs and Crime.
URL: <https://dataunodc.un.org/data/crime/Police%20personnel>
- USAID. 2017. "Security Sector Governance." *United States Agency for International Development* .
URL: <https://tinyurl.com/3k8x3hua>
- Wilke, Anna M. 2023. "Gender Gaps in Support for Vigilante Violence." *Comparative Politics* 55 (2): 263–285.
- 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.
- Zoellick, Robert B. 2008. "Fragile States: Securing Development." *Survival* 50 (6): 67–84.