

# How Does the State Replace the Community?

## Experimental Evidence on Crime Control from South Africa\*

Anna M. Wilke<sup>†</sup>

July 26, 2021

### Abstract

Throughout the developing world, citizens often rely on mechanisms other than the state's justice system to deal with crime. A common form of informal crime control is mob vigilantism, the physical punishment of criminal suspects by groups of ordinary citizens. This paper sheds light on the relationship between state capacity and citizens' willingness to rely on state rather than informal justice mechanisms. Citizens who perceive an increase in the capacity of state institutions may expect them to provide better enforcement services and may voluntarily substitute away from vigilantism. A more capable state may also have a greater ability to punish perpetrators of vigilante violence. I present results from a field experiment in South Africa that creates variation in the capacity of police to locate households. Findings from mid- and endline surveys suggest households exposed to an increase in police capacity became more willing to rely on police and less willing to resort to vigilantism. Results from a mechanism experiment point towards increased fear of state punishment for vigilante violence rather than improved service quality as the link between state capacity and 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. 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, Nicholas Rush Smith, John Huber, John Marshall, Alexandra Scacco 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>.

<sup>†</sup>[amw2229@columbia.edu](mailto:amw2229@columbia.edu) Ph.D. Candidate, Columbia University.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Theory</b>	<b>5</b>
2.1	The choice between state and community justice . . . . .	5
2.2	Crime and policing in South Africa . . . . .	9
2.3	Vigilantism in South Africa . . . . .	11
2.4	Increasing police capacity through a home-based police alarm . . . . .	13
<b>3</b>	<b>Experimental Design</b>	<b>16</b>
3.1	Household sampling and baseline survey . . . . .	16
3.2	Random assignment . . . . .	18
3.3	Treatment take-up and compliance . . . . .	18
3.4	Outcome measurement . . . . .	20
3.5	Estimation and hypothesis tests . . . . .	21
3.6	Ethical considerations . . . . .	21
<b>4</b>	<b>Main Results</b>	<b>23</b>
4.1	Usage of the police alarm . . . . .	23
4.2	Effects on the willingness to rely on police and participate in mob vigilantism . . . . .	24
<b>5</b>	<b>Mechanisms</b>	<b>28</b>
5.1	Conceptual framework . . . . .	28
5.2	Effects on perceptions of police . . . . .	30
5.3	Mechanism experiment . . . . .	35
5.3.1	Design . . . . .	35
5.3.2	Results . . . . .	36
<b>6</b>	<b>Alternative Explanations</b>	<b>40</b>
<b>7</b>	<b>Discussion</b>	<b>42</b>
	<b>Appendices</b>	<b>51</b>
<b>A</b>	<b>Supplementary Information</b>	<b>51</b>
A.1	Pre-registration . . . . .	51
A.2	Explanatory note for regression tables . . . . .	54
A.3	Additional plots and figures . . . . .	55
A.4	Sampling strategy for households . . . . .	58
A.5	Sampling strategy for neighbors . . . . .	60
A.6	Sampling strategy overview . . . . .	61
A.7	Descriptive statistics . . . . .	62
A.8	Information treatments . . . . .	62
<b>B</b>	<b>Identification</b>	<b>63</b>
B.1	Covariate balance . . . . .	63
B.2	Attrition . . . . .	70
B.3	Additional respondents . . . . .	71

<b>C</b>	<b>Additional Analyses</b>	<b>72</b>
C.1	Additional outcomes . . . . .	72
C.2	Robustness . . . . .	74
C.3	Disaggregating indices . . . . .	77
C.4	Additional results information treatments . . . . .	80
C.5	Ruling out alternative explanations . . . . .	82
	C.5.1 Social desirability bias . . . . .	82
	C.5.2 Changes among control group . . . . .	83
	C.5.3 Change in punishment preferences due to improved safety . . . . .	84
<b>D</b>	<b>Question Wording</b>	<b>84</b>
D.1	Table 22 . . . . .	84
D.2	Table 3 . . . . .	85
D.3	Table 4 . . . . .	89
D.4	Tables 6 and 29 . . . . .	89
D.5	Table 7 . . . . .	90
D.6	Measures of prior beliefs . . . . .	90
D.7	Table 18 . . . . .	92
D.8	Table 19 . . . . .	92

## 1 Introduction

Maintenance of order is a core task that has been attributed to governments since classic accounts by [Hobbes \(1958\)](#) and [Wilson \(1889\)](#). Yet, throughout the developing world, citizens often choose mechanisms other than the state’s justice system to deal with crime. One widespread form of informal crime control is mob vigilantism. In many contexts, spontaneously formed groups of ordinary community members physically “punish” criminal suspects.<sup>1</sup> Even though such informal mechanisms may help deter crime in places where the state is weak, this tendency of citizens to bypass the state has severe downsides. First, vigilante mobs commit gruesome assaults, often in response to minor offenses. In South Africa, vigilantism results in an average of two deaths every day ([SAPS, 2018/2019](#)). Second, the reluctance to cooperate with state law enforcement may undermine the effectiveness of formal legal institutions. The police and courts, for instance, rely on information provided by citizens to function effectively ([Tyler and Huo, 2002](#)).

Both the reluctance to cooperate with state institutions and the popularity of informal alternatives are often attributed to state weakness.<sup>2</sup> Implicit in such accounts is the idea that, as state institutions like the police increase in capacity, they will ultimately supersede informal alternatives like vigilantism. In this paper, I shed light on the microdynamics that underlie the relationship between state capacity and citizens’ willingness to, respectively, rely on state institutions and engage in vigilante violence. I present theoretical arguments and empirical evidence collected during thirteen months of field work in South Africa on how and why citizens may change their views and behavior as they come to perceive a greater ability of police to affect their lives.

Police are tasked with apprehending suspects and gathering evidence to ensure that perpetrators of crime will be convicted by the courts. Insofar as police target perpetrators whom citizens would like to see punished, police thus provide a service that citizens demand. To the extent that citizens wish to engage in behavior that lies outside the law, however, police act not as service providers but as regulators of behavior. The perception that police have become more capable to affect their

---

<sup>1</sup>In South Africa, the police registered almost six mob vigilantism cases per day in 2018, with about two cases per day resulting in the death of the accused ([SAPS, 2018/2019](#)). Official police statistics suggest that mobs kill more than one person per day in Uganda ([Uganda Police, 2013](#)). In the main city of Tanzania, Dar es Salaam, and in one province of Papua New Guinea, groups killed one person every two days over the course of a year ([Ng’walali and Kitinya, 2006](#); [CLRC, PNG, 2009](#)). These numbers are likely underestimates of the true prevalence of mob vigilantism, because many incidents do not come to the attention of authorities.

<sup>2</sup>See for example [Acemoglu et al. \(2020\)](#), [Baker \(2002\)](#), [Sekhonyane and Louw \(2002\)](#) and [Tankebe \(2009\)](#).

lives, I argue, may thus change citizens' response to crime through two mechanisms. A first link arises from the logic of competition between service providers. Community members who take the law into their own hands apprehend and punish alleged law-breakers. They hence produce services that, to some degree, resemble those provided by the state. If increased police capacity results in improvements in the quality of state services, it may cause citizens to voluntarily substitute away from vigilantism and towards state institutions. A second link arises from the logic of regulation. Vigilantism is also a crime. Where improved police capacity increases the chance that those who commit vigilante acts will go to prison, citizens may become reluctant to engage in vigilantism.

Throughout history, the capacity of state agents to provide services and regulate citizens' behavior has been shaped by technological innovations. Examples range from the ability to create cadastral maps (Scott, 1998) and the invention of the telegraph (Martland, 2014) to facial recognition software (Xu, 2020) and biometric identification systems (Breckenridge, 2014). These technologies have expanded the reach of the state by making it easier for state officials to identify and locate citizens or, as Scott (1998) puts it, by making citizens more "legible" to state agents.<sup>3</sup> I leverage a similar shift, albeit on a much smaller scale. This study makes use of a unique opportunity to create exogenous variation in the extent to which households are legible to police in order to study the effects on views about state and vigilante justice.

I create this variation in the context of a field experiment that I implemented in partnership with a South African non-profit organization that works closely with South African police. Even though South Africa falls at the upper end of the distribution of conventional measures of state capacity in Sub-Saharan Africa, police capacity remains low in many parts of the country (Khayelitsha Commission, 2014). Many South Africans are hesitant to rely on police and instead draw on various types of informal enforcement mechanisms (Baker, 2008). Vigilantism is widespread, especially in urban and semi-urban areas of the kind in which this study is set.

The experiment equips households in one police precinct with an alarm system that links them

---

<sup>3</sup>Scott (1998, chapter 1) argues that cadastral maps allowed the state to identify property owners, thereby expanding its capacity for revenue collection. Martland (2014, p. 298) describes how the speed in communication associated with the telegraph allowed police in Chile to curb crime by locating bandits involved in cattle theft. Curry, Phillips, and Regan (2004) show how geo-location technology improved the efficiency of state provided emergency services by allowing first responders to locate households in an emergency. Barnwal (2017) and Muralidharan, Niehaus, and Sukhtankar (2016) show how biometric identification systems improve the efficiency of government transfer schemes by making it possible to uniquely identify beneficiaries. See Lee and Zhang (2017) and Brambor et al. (2020) for general accounts of how state capacity depends on the information that states have about citizens.

to the police. When activated, the alarm sends text messages to various police personnel. Names, contact details and the location of households that own an alarm are on file at the police station and also contained in the text message. The system is meant to improve police service delivery by making it easier for police to quickly find households if they are subject to crime. That their names and contact details are on file at the police station, however, may also worry households who would like to engage in illegal activities such as vigilantism. Being more legible to police may increase the perceived risk that police would find out if one broke the law.

This study is based on 250 households, 100 of which were randomly assigned to receive an alarm system. I measure outcomes during a midline ( $N = 483$ ) and an endline ( $N = 448$ ) survey that were conducted, respectively, one and eight months after treatment roll-out. The findings suggest that respondents from households that received an alarm became more willing to rely on police. In both mid- and endline surveys, respondents in the treatment group are more likely to say that they would reach out to police if they experienced a crime, for example. Evidence of effects on the willingness to engage in vigilantism is mixed in the sample as a whole. A pre-registered subgroup analysis, however, indicates a sizable negative effect on this outcome among parts of the sample who had particularly low prior beliefs about police at baseline.

Delving into the exploration of mechanisms, I find that respondents from households that received an alarm developed both a more positive view of the quality of services that police provide and a greater sense of being regulated by police. For example, respondents in the treatment group believe that police would respond more quickly if called, but also perceive it more likely that police would find out if respondents engaged in illegal behavior.

To better understand the relative importance of these changes in bringing about the apparent decrease in the willingness to participate in vigilantism, I subsequently present results from a mechanism experiment that was part of the endline survey. The experiment featured two information treatments designed to induce independent changes in respondents' beliefs about police efforts to provide services that citizens desire and to convict those who participate in vigilantism.

Two findings emerge. First, only the treatment that highlights police efforts to bring perpetrators of vigilantism to book appears to decrease the willingness to participate in vigilantism. Even though respondents who were assigned to the service quality treatment are more likely to believe that police make efforts to arrest criminals whom citizens would like to see convicted, there is little

evidence that this change translated into a reduction in respondents' proclivity to participate in vigilantism. Second, priming respondents to think that police make efforts to convict perpetrators of vigilante violence appears to increase the extent to which the alarm treatment discourages vigilantism. Priming police's commitment to service delivery seems to have, if anything, the opposite effect. Both patterns point to the importance of an increased expectation of state punishment as a link between police capacity and citizens' reduced proclivity to engage in vigilantism.

These results speak to a growing literature on the interplay between formal and informal justice institutions.<sup>4</sup> A prominent idea in this literature is that the existence of multiple ways to resolve disputes and punish wrong-doers leads citizens to forum shop (e.g. Sandefur and Siddiqi, 2011). In other words, when faced with more than one provider of enforcement services, citizens choose to rely on whichever provider promises to deliver the best outcome at lowest cost. State capacity plays a role in these accounts because it allows the state to expand formal service provision to previously unserved areas (see e.g. Cooper, 2018) or to alter the characteristics of state services so that they compare more favorably with those provided through informal channels (see e.g. Sandefur and Siddiqi, 2011). The state is thus treated as one of many entities that compete for the demand of citizens who act as consumers in a market for enforcement services.

The findings presented in this paper suggest that the relationship between the state and informal enforcement mechanisms is not limited to one of competition. Even though states sometimes choose to recognize informal actors as legitimate providers of enforcement services (Oomen, 2005), highly decentralized forms of crime control such as vigilante mobs typically fall outside the confines of the law. Where the state has declared an informal alternative illegal, an increase in police presence may induce citizens to abandon this alternative not because of an improvement in state services but out of a concern about state punishments for those who deal with crime in extra-judicial ways. Increased police capacity may be particularly effective at reducing vigilantism if citizens are convinced that police are determined to convict those who commit vigilante acts.

A second take-away from this study concerns the upsides and downsides of state capacity from the perspective of citizens. Much of the recent literature on state capacity focuses on its importance for economic development and government service provision (Berwick and Christia, 2018). The

---

<sup>4</sup>See for example Acemoglu et al. (2020), Cooper (2018), Blair (2018), Blair, Karim, and Morse (2019), Sandefur and Siddiqi (2011), Lazarev (2017), Magaloni, Franco Vivanco, and Melo (2018), Aldashev et al. (2012), D'Aoust and Sterck (2016).

results presented here shed light on potential downsides of state capacity for certain groups of citizens. Even democratic states use their capacity not only to provide citizens with services but also to limit citizens' choices by placing sanctions on those who engage in practices that the state deems illegal. Vigilantism is only one example of an illegal practice that is widespread. Other examples include electricity theft (Smith, 2004b), unlicensed street vending (Holland, 2016) or tax evasion (Ali, Fjeldstad, and Sjursen, 2014). An increase in the capacity of state institutions like the police is likely to be a double-edged sword for those who support such activities.

This paper is structured as follows. Section 2 discusses the theoretical link between police capacity and the choice between state and community justice. This section also relates the theoretical framework to the study context and the police alarm intervention. Section 3 describes the design of the experiment and section 4 presents its main results. Section 5 lays out theoretical predictions and empirical results that help distinguish the two hypothesized mechanisms. Section 6 considers alternative explanations, chief among which is the worry that effects may be driven by experimenter demand. Section 7 concludes.

## 2 Theory

### 2.1 The choice between state and community justice

According to Max Weber's (1946) classic notion of the state as an organization that claims the monopoly of the legitimate use of violence, the placement of sanctions on those who break the law is a task fulfilled by state institutions. In practice, formal justice institutions often co-exist with informal enforcement mechanisms. In some contexts, such informal mechanisms take the form of non-state actors including traditional authorities, gangs or vigilante organizations (Baker, 2008). I here focus on settings where the dominant mechanism of informal crime control involves ordinary community members meting out punishments against criminal suspects, while centralized providers of informal enforcement services are absent or weak. A report from South Africa contains a detailed description of an example of such vigilante violence:

Let me give you an example of what happened in my neighborhood just this morning at 5.00am! We heard a woman screaming *i-Bag yam? I-Bag yam? Nal'isela* (My bag! My Bag! Here's a thief!). In no time, I mean, in no time, everybody was coming out, slamming doors behind them. I mean, it was like a split second – and they were all dressed in their clothes, not pyjamas. It was as if they were waiting, ready all night for

exactly this kind of thing to happen. 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 no time – they had finished him. Think about it, 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)

Accounts of incidents in which spontaneously formed groups of citizens hurt or kill suspects of criminal behavior are abundant throughout the developing world ([Human Rights Watch, 2010](#); [Smith, 2004a](#); [Adinkrah, 2005](#); [Schubert, 2013](#); [Kirsch and Grätz, 2010](#); [Gross, 2016](#)). When asked why they participate in or support vigilantism, respondents in qualitative interviews often point towards unsatisfactory police performance. As one South African respondent explained: “It is not a good thing to take the law into your own hands, but since the police is not doing a good job, people have no other option.”<sup>5</sup> A similar sentiment was expressed by a market vendor in Uganda: “The police does not help us. So, we have become police ourselves.”<sup>6</sup> How may citizens’ views on crime control change as police become more capable?

I take as a starting point the assumption that citizens wish for some set of transgressions against the law to be punished. This demand may reflect various desires including a preference for the deterrence of future crime or a taste for vengeance ([Becker, 1968](#)). Consider a citizen who has information about a crime. To fix ideas, imagine that the citizen, in the absence of a centralized informal alternative such as a gang or vigilante organization, has two principal options to ensure that the perpetrator of this crime is punished. She can either report what she knows to the police or rally her family, friends, and neighbors.<sup>7</sup> Inspired by qualitative evidence from South Africa, I discuss four considerations that may affect the citizen’s choice.

*Nature of the offense.* The set of offenses for which it is indeed true that both state and community punishment are viable options will often be a subset of the actions that citizens would like to see punished. Witchcraft, for example, is not typically punishable by law. Still, the demand for punishment of alleged witches is high in many contexts including South Africa ([Smith, 2019](#); [Miguel, 2005](#)). The apprehension of perpetrators who are heavily armed, on the other hand, requires coercive capacities that typically go beyond those of ordinary community members. Drug dealers,

---

<sup>5</sup>Interview with South African resident of study precinct in Northwest Province on 9 August 2018.

<sup>6</sup>Focus group with market vendors in Owino market, Kampala, Uganda on 31 May 2017.

<sup>7</sup>In practice, citizens also have the option not to report the crime or to rely on both the community and police.

for instance, are rarely the target of vigilantism, because they tend to be connected to heavily armed criminal networks. Despite these examples, the set of offenses that get addressed through both the state’s justice system and vigilantism remains large and includes minor crimes such as petty theft and burglary as well as violent offenses like sexual assault and rape.

*Expected punishment for criminal suspect.* Provided that the offense in question may be punished by both the state and the community, an important consideration may be the kind of punishment that perpetrators would receive under either scenario. Countless anecdotes suggest that vigilantism results in harsh and often gruesome punishments.<sup>8</sup> The contrast between state and community punishment is particularly stark when it comes to petty crimes. If dealt with through the state’s justice system, such crimes would, in most contexts, result in not more than a short prison sentence. When dealt with through vigilantism, they often end in grave injury or death.<sup>9</sup> Aside from being harsh, community punishments also tend to be more public than state sanctions. Vigilante incidents often attract large crowds of spectators.<sup>10</sup> How these differences shape the choice between state and community justice may vary with tastes for punitiveness and instrumental concerns about deterrence. Many South African respondents cited its harsh and public nature as a reason to support vigilantism, for example. Others opposed vigilantism on the same grounds.<sup>11</sup>

*Probability that criminal suspect will be punished.* Another concern may be whether suspects of crime will indeed be punished when dealt with by the state or the community. Communities appear particularly effective in this regard. Being spatially proximate to the crime scene and well informed about community structures, community members likely have an easier time apprehending alleged perpetrators. Communities are also not bound by due process requirements and often mete out punishment even if evidence of guilt is tenuous.<sup>12</sup> Turning to the state’s justice institutions, it is

---

<sup>8</sup>A re-occurring form of punishment, both in South Africa and other contexts, is “necklacing” which involves putting a rubber tire filled with petrol around a person’s chest and arms and setting it on fire.

<sup>9</sup>Examples from the study precinct (Northwest Province, South Africa) include a person who was beaten to death for having stolen two chairs; a person who suffered serious injury when being assaulted for stealing sneakers and two people who were burnt to death for stealing food and drinks.

<sup>10</sup>One respondent estimated crowd sizes of sixty to ninety community members for various incidents in the study precinct (interview on December 3 2018, study precinct, Northwest Province, South Africa).

<sup>11</sup>Various qualitative interviews with residents of study precinct in Northwest Province, 2018.

<sup>12</sup>The following anecdote illustrates the fragility of the evidence base on which vigilantism can be based. During the baseline survey of this project, residents of the study precinct assaulted a man who had come to the precinct from a neighboring community, because they found that the man had in his possession a phone, which had previously belonged to a community member. The previous owner of the phone claimed that the phone had been stolen by a group of men known in the community for their addiction to nyaope (a prevalent drug in South Africa). Since those engaged in substance abuse often exchange stolen goods for drugs, community members concluded that the man must be a drug dealer. Fortunate for the man, police arrived before he was injured severely. Police could not

here where police most obviously play a role. The severity of a sentence will be decided by judges who are bound, to some extent, by a country's penal code. A precondition for the ability of courts to hand out any sentence at all, however, is that perpetrators have been apprehended and that enough evidence has been collected to prove their guilt. Both outcomes are direct outputs of police activity. As police forces increase in capacity, they may be better able to provide these services.

*Risk of state punishment for vigilantes.* Finally, citizens may consider the legal repercussions of their engagement with non-state mechanisms of enforcement. Such mechanisms do not always fall outside the law. Rural traditional courts, for example, are non-state actors but still recognized by the South African constitution. If a traditional court finds someone guilty of stealing and determines that the person's family has to pay a fine, the state treats this judgment as a legitimate use of authority and not as a robbery (see Oomen, 2005, p. 202). Acts of mob vigilantism that are perpetrated by ordinary community members, however, amount to serious crimes like assault or murder.<sup>13</sup> As a consequence, targets of vigilantism have access to legal recourse. One survivor of a vigilante incident in the study precinct, for example, opened an assault case against his attackers. In another case, several individuals were sent to prison because they participated in the burning of two suspected thieves.<sup>14</sup> This risk of state punishment may be most relevant for the decision of whether to actually inflict violence – an act that is often, though not exclusively, carried out by young men.<sup>15</sup> The woman who screamed “My bag! My Bag! Here's a thief!” in the anecdote above, for example, is unlikely to be held accountable for the death of the accused if she did not participate in the assault. That said, those who inflict vigilante violence will often be the instigator's immediate community – her family, friends and neighbors. Hence, even in deciding whether to encourage or support vigilantism, citizens may take into account the chances that doing so will cause their husbands, sons and brothers to get arrested. Just as a police force with more capacity will be better able to apprehend perpetrators whom citizens would like to see convicted, more police capacity may also enable police to more effectively investigate cases of vigilantism.

---

ascertain any evidence that the man was involved in the drug business. It remained unclear whether his phone was indeed the phone that had been stolen from the community member and whether this phone or a different phone had been exchanged for drugs. The victim of the vigilante incident opened a case against the community members who assaulted him (Interview with police, 8 May 2018, Northwest Province).

<sup>13</sup>The South African police, for example, conducted a nationwide analysis of dockets for cases of murder, attempted murder and assault to determine the share of dockets that pertain to incidents of vigilantism (SAPS, 2018/2019).

<sup>14</sup>Interviews with 4 men and 2 women, August 7 and 8 2018.

<sup>15</sup>In the control group of this study, the share of respondents who stated that they would personally participate in the beating of a suspect was 21% among men under the age of 35 and 10% among the rest of the sample.

This discussion suggests that police capacity may matter to the extent that it allows police to increase two core outputs, the probability that a perpetrator will be convicted if a crime is reported to the police and the probability that community members will be sanctioned if they, instead, take the law into their own hands. In other words, an increase in police capacity may affect the choice between state and community justice through two mechanisms:

1. *Improved Quality of Police Service:* An increase in police capacity may encourage cooperation with police and discourage mob vigilantism by increasing the probability that a criminal who is reported to the police gets sanctioned through the formal justice system.
2. *Increased Risk of State Punishment:* An increase in police capacity may encourage cooperation with police and discourage mob vigilantism by increasing the probability that participation in vigilante violence has legal repercussions.

While these two mechanisms are not mutually exclusive, each points towards a qualitatively distinct reason for why state institutions may be “the only game in town” in certain places but not others. The first mechanism implies that citizens of strong states voluntarily choose to consume services offered by the state, because police capacity makes it possible for the police to *out-compete* the community as a service provider. The second mechanism implies that citizens in strong states refrain from participating in vigilante violence, because a capable police force helps to effectively *rule out* such behavior.

## 2.2 Crime and policing in South Africa

This study is set in South Africa, where both violent and property crime are widespread.<sup>16</sup> Problems are particularly acute in the country’s townships. A result of urban planning policies implemented under South Africa’s Apartheid regime,<sup>17</sup> townships are racially segregated semi-urban areas that are typically located at the outskirts of major towns and cities. Not unlike urban slums in other parts of the world, townships tend to suffer from limited access to government services including

---

<sup>16</sup>According to the United Nations Office on Drugs and Crime (UNODC), South Africa recorded 36 homicides per 100,000 people in 2017; roughly five times the worldwide average of 5 homicides per 100,000 people. Homicide rates are often said to be the most reliable indicator of the overall levels of crime, because homicides tend to come to the attention of authorities irrespective of reporting decisions. UNODC suggests that South Africa experienced 2 robberies per 100,000 people in 2016 as compared to a worldwide average of 0.3 robberies per 100,000 people.

<sup>17</sup>See [Mahajan \(2014\)](#) for more on township characteristics and history.

water, sanitation and electricity. Many townships now contain large informal settlements that remain entirely without access to these services.

The same is true for the predominantly black, low income township in the Northwest Province where this study takes place.<sup>18</sup> Around 60% of respondents in this study do not own a flushing toilet, for example, and more than 90% do not have running water inside the house. The threat of becoming a victim of crime is a constant reality in the study precinct. Figure 8 in section A.3 of the appendix details the most common kinds of crimes. Burglary, robbery and serious assault are especially prevalent and happen at a rate that far exceeds the respective country-wide median.<sup>19</sup> In the baseline survey of this study ( $N = 250$ ), roughly 44% of respondents report that someone in their household has experienced a crime over the past year.

Police performance is a concern throughout South Africa,<sup>20</sup> and police stations in low income townships tend to face the biggest obstacles.<sup>21</sup> While the study precinct does not fall far from the countrywide median when it comes to police personnel per capita,<sup>22</sup> the conditions in this precinct, as in many townships, are not set up to facilitate the work of police. Scott (1998) famously argued that the ability of state institutions to function effectively is linked to the extent to which communities are “legible” to state agents. Even though townships under Apartheid were designed to be legible spaces that could be easily policed, townships tend to be difficult to “read” today. Many townships have grown substantially over recent years. Instead of following a carefully thought-out urban design, this growth typically happened organically through the expansion of informal settlements. The result tends to be a confusing layout of streets and the absence of a

---

<sup>18</sup>Data from the 2011 census suggest that the median income in the study precinct is slightly below (80%) the median income across all police precincts.

<sup>19</sup>Since crimes that result from mob vigilantism are not counted separately, the high rate of assaults may in part reflect the prevalence of mob vigilantism. Moreover, actual crime rates may be even higher than suggested by official crime statistics insofar as not all crimes are reported to police. That said, these numbers are also based on population estimates from the 2011 census, and the population of the study precinct has almost certainly grown.

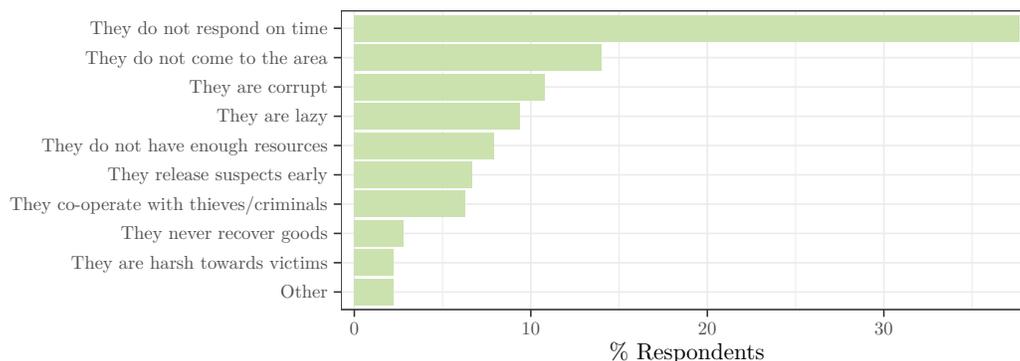
<sup>20</sup>A government-mandated study that sought to determine reasons for the prevalence of armed violence points to “a general lack of capacity for investigating and prosecuting perpetrators of crime” (CSVR, 2010, 52).

<sup>21</sup>A commission of inquiry that was established in 2012 as the result of public outrage about the quality of policing services in one particularly crime-ridden township found that police stations that serve low income communities are most understaffed (Khayelitsha Commission, 2014). The judge presiding over a resulting court case in front of South Africa’s inequality court ruled in 2018 that the way in which the police allocate resources across police precincts discriminates against poor and black communities.

<sup>22</sup>The median police precinct had 2.9 assigned police officers per 1,000 people in 2015/2016. The study precinct had 2.4 assigned police officers per 1,000 people in the same year. These are data on the number of police officers *assigned* to a police station, which do not necessarily reflect the number of police officers actually present at a given station. Data have been obtained from the Social Justice Coalition, an activist group involved in the Khayelitsha Commission and the Equality Court case on the discriminatory distribution of police resources.

well organized address system. Street names, for example, are rare in the study precinct. Instead, houses are numbered within sections that comprise hundreds if not thousands of houses. To make matters worse, there are three different numbering systems and the numbers within a given system are not necessarily in sequence. Even during the day, it is thus difficult to locate specific houses. Problems are more severe at night, because street lights are rare. In addition, most police officers do not live in the communities that they serve which limits their insight into community structures.

Dissatisfaction with police is widespread throughout the country.<sup>23</sup> Figure 1 shows that slow response times are the most important grievance, which is perhaps not surprising given the infrastructural conditions.<sup>24</sup> Other complaints are that one’s area is not served by the police at all and that police are corrupt. Opinions in the study precinct mirror these nationwide concerns. 55% of respondents in the control group of this study, for example believe that police would never come or take longer than two hours to arrive when called in an emergency. A related complaint is that police are generally ineffective at ensuring that those who are guilty of crime are convicted.<sup>25</sup>



**Figure 1:** Slow response is main reason for dissatisfaction with police service ( $N = 8,906$ ). Calculated among 43% of respondents who said they are not satisfied with police in their area. Respondents were asked to select the main reason for their dissatisfaction. Data taken from the nationally representative Victims of Crime Survey 2016/17 by StatsSA.

### 2.3 Vigilantism in South Africa

Centralized providers of informal enforcement services such as traditional courts, gangs and vigilante groups exist in various parts of South Africa. Nonetheless, there are many townships like the

<sup>23</sup>A recent nationally representative opinion poll ( $N = 21,095$ ) indicates that 43% of South Africans are dissatisfied with police services in their area (StatsSA, 2016/2017).

<sup>24</sup>See Curry, Phillips, and Regan (2004) for arguments about how a swift emergency response depends on well organized address systems and the availability of location data.

<sup>25</sup>61% of control group respondents believe that “the police and the courts often let people who are guilty go free.”

study precinct in which the dominant form of informal crime control is mob vigilantism. Tribal authorities, for example, have little influence in the precinct where this study is set. Neither is there an organized vigilante group or a gang that residents could turn to with concerns about crime.<sup>26</sup>

The term “mob justice” is widely used throughout the study township to refer to incidents in which community members physically punish suspects of criminal behavior. In fact, many households own a whistle for the purpose of summoning the community in emergencies.<sup>27</sup> Figure 9 in section A.3 of the appendix plots data from the endline survey of this study in which respondents were asked how many vigilante incidents they could recall happening in their area between May and July 2018. In most areas, at least a quarter of respondents could recall one incident or more. In qualitative interviews, most respondents were able to describe at least one if not multiple cases in their area. Most anecdotes involved victims who had been accused of burglary or theft, often in the form of stealing items from people’s yards or houses. Some of the more serious accusations were related to sexual violence including rape. Many, though not all incidents of vigilante violence led to severe injuries of the accused; in the most extreme cases, the accused had been killed.

The approach of police in South Africa to mob vigilantism is ambiguous. While there is no shortage of anecdotes about police turning a blind eye or even supporting vigilantism, arrests and convictions do happen.<sup>28</sup> Many police officers seem aware that police action against vigilantism has the potential to spark public outrage since many support such violence.<sup>29</sup> Around 40% of control group respondents, for example, support the beating of someone who is “known to be involved in stealing cars and plasma TVs” and almost half oppose prison sentences for perpetrators of vigilantism even in cases where the criminal suspect was killed. Yet, many police personnel seem also convinced that giving in to community pressures would be a failure to perform one’s duty and could result in disciplinary consequences.

Table 1 displays the joint distribution of perceptions about the two theoretically relevant outputs

---

<sup>26</sup>There is an organization that deals specifically with cattle theft which however happens rarely (see figure 8).

<sup>27</sup>Examples given of situations in which the whistle would be used include someone noticing a stranger in their yard or discovering that their house had been broken into. Whistles are also used to summon the community for other reasons such as the start of a community meeting. Hearing a whistle in the evening hours, however, is often interpreted as a call for help, since meetings do not typically happen at that time.

<sup>28</sup>Interviews with police in Northwest Province in 2018 and 2019.

<sup>29</sup>Several individuals who were involved in the vigilante incident described in footnote 12, for example, were subsequently held by the police on assault charges. A group of angry community members marched to the police station, demanding the release of the perpetrators of vigilantism. A patrol officer described responding to a different incident of vigilantism and facing a large group of angry community members as a frightening experience.

of police activity, the quality of service delivery and the risk of state punishment for vigilantism, in the baseline survey of this study. The table suggests that expectations about these two quantities are not as correlated as one may expect. For example, around 16% of respondents fall above the median in terms of their expectations about the quality of services that police provide but do not think it likely that those who participate in mob vigilantism would be arrested. Conversely, almost one third of respondents perceive it likely that there would be legal repercussions for vigilante violence while, at the same, having little hope that reliance on police would result in high quality services like a speedy response. This latter group may include particularly disgruntled residents who, in qualitative interviews, described the police as doing little for the community while also depriving community members of the possibility to protect themselves against crime.

Risk of punishment MV	Police service quality		Total
	Low	High	
High	27.2%	26.0%	53.2%
Low	31.2%	15.6%	46.8%
<b>Total</b>	58.4%	41.6%	$N = 250$

**Table 1:** Baseline perceptions of police service quality and risk of state punishment for mob vigilantism.

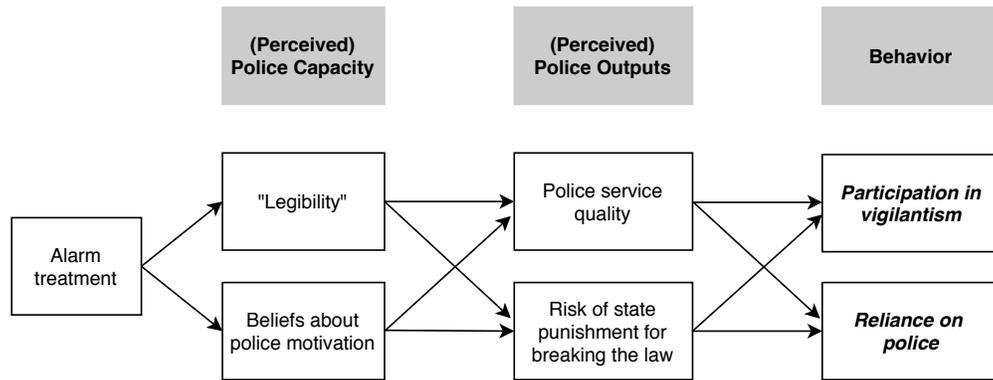
At baseline, one woman was interviewed in each household. Percentages indicate shares of respondents. Respondents who answered “not very likely” or “not likely at all” when asked whether police would arrest perpetrators of a hypothetical vigilantism incident in the respondent’s street are coded as perceiving a low risk of punishment for mob vigilantism. Whoever answered “somewhat likely” or “very likely” was coded as perceiving a high punishment risk. Perceptions of police service quality are measured through an index of three items: *Customer Service*, *Arrive quickly* and *Send guilty to prison*. Respondents are coded as perceiving police service quality as high or low depending on where they fall relative to the baseline sample median. See appendix section D.6 for details on question wording.

## 2.4 Increasing police capacity through a home-based police alarm

The intervention that is the subject of this study consists of a home-based alarm system that was developed by a South African non-profit organization in collaboration with South African police. The alarm is a small electronic device that is installed in the house and can be triggered via a panic button, remote control or cell phone.<sup>30</sup> The system also includes a motion sensor that can be activated if no one is at home. The alarm can be triggered silently or such that a bright light

<sup>30</sup>Households without access to electricity received a solar panel to power the alarm. The solar panels are specifically designed to power the alarm system and, beyond that, do not provide general access to electricity. Moreover, the alarm is equipped with a 24-hour battery that can power the alarm during electricity outages.

flashes outside the alarm owner’s house and a siren sounds. In either case, the alarm sends text messages to various police personnel including members of the police station management,<sup>31</sup> police officers currently on duty at the police station and senior members of the Community Policing Forum (CPF), a volunteer organization that serves as liaison between police and community.<sup>32</sup> Text messages indicate the alarm owner’s name, contact numbers, and landmarks close to the alarm owner’s home. This information is also on file at the police station. All instances in which an alarm is triggered are registered in the back end system of the non-profit organization.



**Figure 2:** Hypothesized effects of police alarm.

Figure 2 summarizes two ways in which this alarm treatment may alter the de facto or perceived ability of police to affect the lives of household members. First, households that are protected by an alarm may become more “legible” for police, i.e., police may be better able to identify and locate households and their members. A core purpose of the alarm is to make it easier for police to find households that reach out to police in an emergency. Alerting the police without an alarm requires calling a centralized emergency hotline or the closest police station. When calling from a landline phone, calls to the emergency hotline are free of charge.<sup>33</sup> Given the absence of a reliable address system and the confusing layout of township streets, however, it can be difficult to explain the location of a specific house over the phone. This is especially true if the person on the phone is a call center agent who lacks familiarity with the specific township. The absence of street lights further complicates the process of locating households at night time. The alarm system was

<sup>31</sup>The study precinct is partitioned into four sectors and the alarms send text messages to sector managers.

<sup>32</sup>Alarm owners can also link up to two neighbors to their alarm, who will receive text messages when the alarm is triggered. Neither surveys with households who were eligible for the alarm treatment nor surveys with neighbors provided any evidence of a change in community relations as a result of the alarm treatment.

<sup>33</sup>See [https://www.saps.gov.za/services/cc\\_10111.php](https://www.saps.gov.za/services/cc_10111.php).

designed to address these challenges by sending location details directly to the local police station, flashing a bright light and sounding a siren. Moreover, this study installed only 100 alarms in a precinct with more than 42,000 residents. Anecdotal evidence suggests that, by the end of this study, some police officers would have been able to find households with an alarm from memory alone. Alarm recipients may thus expect a faster police response, which may increase the chance that perpetrators are apprehended and convicted – an improvement in police service quality.

On the flip side, alarm recipients may also feel more watched by police. Being easy to locate for police has its upsides if one is subject to crime, but it may also increase the chances that police are able to find out and successfully intervene if a household member breaks the law. Many residents of the study precinct live in informal settlements with little or no access to government services. Hence, local authorities such as police have limited records of who these people are, where they live and what they do. Names and contact details of alarm owners, however, are on file at the police station irrespective of whether the alarm is triggered. In part, the effect of this change may be psychological. The alarm console, for example, shows the logo of the South African Police Service,<sup>34</sup> and hence may serve as a daily reminder that someone at the police station now has a record of this household and its residents. This thought may be unsettling to someone who engages in or lives with family members who engage in activities like vigilantism that fall outside the law, even without it being immediately obvious how the police will use this information.

A second possibility is that the alarm treatment causes household members to update their views of police more generally in ways that are not a direct result of the functionality of the alarm. Community members may expect the alarm project to have differential costs and benefits for different “types” of police, for example. That the alarm gives community members an easy way to contact police and provides information about incidents to an outside party may be perceived as a nuisance by unmotivated police.<sup>35</sup> Highly motivated police, on the other hand, may welcome the project as a way to increase police output. Households who receive the alarm may thus perceive the police’s involvement in the alarm project as a signal of police motivation.<sup>36</sup> Learning may also

---

<sup>34</sup>See figure 10 in the appendix for a picture of the alarm console.

<sup>35</sup>The concern that the alarm system creates too much “demands” on police has sometimes been voiced by police and there is variation across police precincts in the willingness to collaborate with the non-profit organization.

<sup>36</sup>Of course, other community members could draw the same inference to the extent that they can spot the alarm sirens or come to know about the project in other ways. Yet, a survey with neighbors showed little evidence of indirect effects of the alarm treatment on surrounding households.

occur through interactions with police that arise as a result of the alarm treatment. The belief that police are highly motivated to perform may again make alarm recipients more optimistic about the services that police provide but also more worried about the consequences of breaking the law.

In line with the central argument of this paper, the alarm treatment may thus affect the choice between formal and informal crime control through two logics. On the one hand, the belief that police are able to respond quickly when called to the home and highly motivated may make reliance on the state more attractive. Importantly, the alarm system is well set up to improve police performance with regard to burglaries, a crime that is both common in the study precinct and often dealt with through mob vigilantism. The perception that police take their job seriously together with the belief that one has become “known” to police, on the other hand, may make participation in vigilantism, like involvement in any other kind of crime, appear more risky.

### **3 Experimental Design**

#### **3.1 Household sampling and baseline survey**

The police precinct in which this study is set has been selected in collaboration with the implementing partner and the South African police. A main reason was the high rate of burglary and robbery at residential premises in this precinct – crimes that the alarm intervention appears well suited to address. The precinct is located in the Northwest Province and covers a semi-urban area that is roughly thirteen kilometers long and, at its widest point, six kilometers wide. The experiment involves a sample of 250 households which were selected as part of a baseline survey that took place between May and July 2018. See figure 3 for more details on the study timeline.

Households were sampled in two ways. 135 households were chosen from a list of vulnerable homes provided by the police. This procedure corresponds to the implementing partner’s usual way of selecting beneficiaries. In theory, the police was supposed to create this list by consulting their records of households that recently experienced crime. In practice, the CPF was heavily involved in the creation of this list and appears to have selected households whose members are regular attendees of crime-related community meetings. The remaining 115 households were chosen from a pool of households that was created independently from the police. A team of enumerators walked the streets of the eleven most high crime areas of the precinct and geo-located every tenth house.

Households were selected from these two pools in non-random ways designed to limit non-

compliance, attrition and interference. To limit interference, a stochastic algorithm was used to select the largest sample of households in each pool such that each household is located no closer than 150 m to any other house in the sample. Due to inaccuracies in the geo-location procedure, around 67% of the final sample satisfies this constraint. To limit non-compliance, 27 households who indicated in the baseline survey that they would not be interested in an alarm system were excluded. To limit attrition, 77 households that were interviewed during the baseline survey but could not be reached via phone or in person during subsequent back-checks were excluded from the final sample. More details on sampling can be found in sections A.4 and A.6 of the appendix.

**Figure 3:** Study timeline

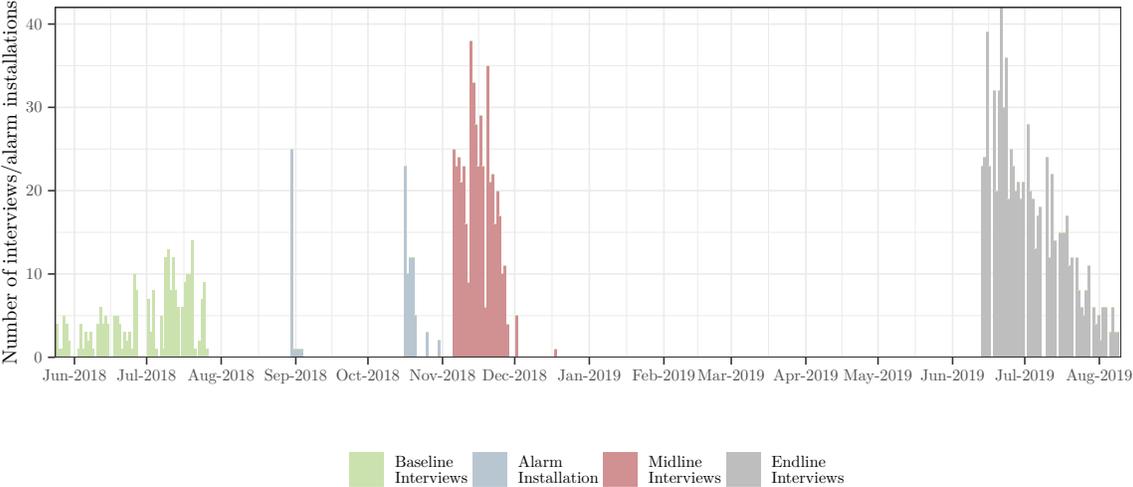
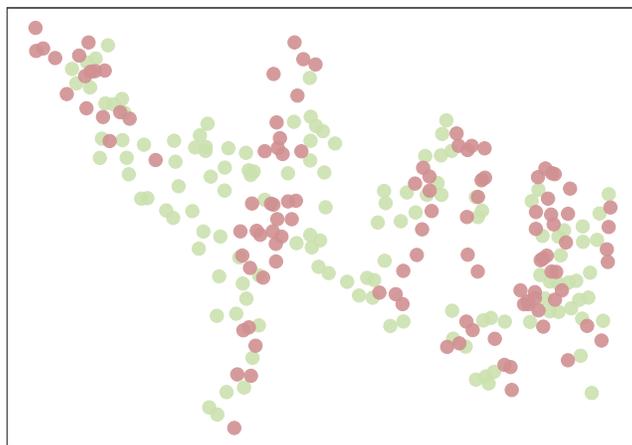


Figure 4 gives a sense of the locations of households in the final sample. Precise locations are not shown in order to protect the identity of study participants. There is spatial overlap across the two sets of households – those that were sampled through the police and those that were sampled through the listing exercise. Households sampled through the police are more geographically dispersed, because the listing exercise was only conducted in specific areas that the police identified as most crime-ridden. During the baseline survey, one woman was interviewed in every household.<sup>37</sup> Enumerators were instructed to interview the woman who is most involved in household

<sup>37</sup>In other projects, the implementing partner focuses on the safety of women and was hence most interested in safety-related views of women. To accommodate these preferences, the baseline survey included only female respondents. Budget constraints made it impossible to survey two household members at baseline. Mid- and endline surveys interview a woman and a man in each household where available. The sample includes a larger share of women because some households sampled at baseline turned out to be all-women households.

decisions.<sup>38</sup> Table 8 in section A.7 of the appendix compares baseline responses across respondents that were sampled, respectively, through the police and the listing exercise. While the two groups appear similar in many respects, respondents who were sampled through the police have higher expectations of police outputs and a greater baseline willingness to turn to the police.

**Figure 4:** Households in study sample



● Police Sample ● Listing Sample

### 3.2 Random assignment

Prior to random assignment, households were organized into 50 blocks of 5. To do so, the sample was divided into two sets according to the way in which households had been sampled. Within each set, blocks were formed to minimize the within-block multivariate Mahalanobis distance of four variables: baseline measures of support for and the willingness to participate in vigilantism and the household's latitude and longitude.<sup>39</sup> 100 households, two in each block, were assigned to receive the alarm treatment. Alarm installations took place in September and October 2018.<sup>40</sup>

### 3.3 Treatment take-up and compliance

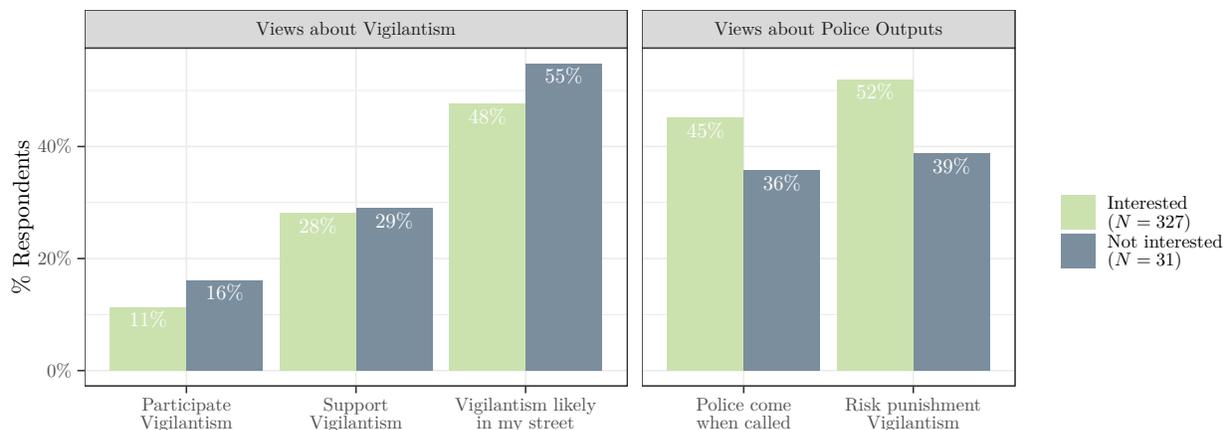
Interest in the police alarm was widespread. Only a small share of respondents (27 out of 358) was excluded from the study prior to random assignment because of a lack of interest in the alarm at

<sup>38</sup>This rule ensured that respondents would be able to confirm their household's interest in the alarm system.

<sup>39</sup>Blocking on the sampling strategy and on latitude and longitude ensured buy-in from the police and the CPF, who wanted alarms to be spread across the precinct and to be distributed mainly to households on the police list.

<sup>40</sup>Due to supply chain problems, treatment roll-out was interrupted after the first set of alarms had been installed in September and only continued in the second half of October.

baseline. At the time of the midline interview, 93 of 100 households in the treatment group and no household in the control group were equipped with an alarm. Among the seven households in the treatment group that did not comply with their assigned treatment, four households refused the alarm, one dismantled it after installation and two did not receive an alarm due to administrative errors. Of the latter two households, one received an alarm before the endline interview.<sup>41</sup>



**Figure 5:** Baseline views by interest in police alarm.

The sample includes 358 baseline respondents. 250 of these are part of the experimental sample. See section 3.1 and section A.4 of the appendix for 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 interview, but had changed their mind when they were contacted during subsequent telephonic back-checks. Four households refused the alarm after they had been randomly assigned to receive it.

To some extent, the widespread interest in the police alarm may seem at odds with the argument that the alarm may increase the perceived risk of state punishment for illegal activities. Given that many of them support vigilantism, why would respondents agree to a treatment that stores their contact details at the police station? One explanation is that respondents trade off the downsides of increased police supervision with the promise of improved service delivery. The latter concern may weigh heavily given that vigilantism is not a panacea for all kinds of crime. Gun violence is common in the study precinct. Since communities lack the coercive capacity to mete out punishments against groups of heavily armed perpetrators, many may be willing to limit their freedom to engage in vigilantism in exchange for improved police protection.

How citizens resolve this trade-off may depend on their expectations about police service delivery

<sup>41</sup> A regression of an indicator for a household’s receipt of the alarm on a treatment assignment indicator yields an *F*-statistic of roughly 1977, suggesting that treatment assignment remains a strong instrument for treatment received.

and their taste for vigilantism. Figure 5 plots responses from all baseline respondents, including those who did not become part of the study sample. The figure compares respondents who showed interest in the alarm to those who refused it, either at baseline or during a later stage.<sup>42</sup> Putting aside the caveat that the number of households who refused the alarm at any point is very small ( $N = 31$ ), the plot suggests that these groups differ in intuitive ways. For example, those not interested in an alarm are 45% more likely to say they would participate in vigilantism and appear less optimistic that police would arrive when called in an emergency.

### 3.4 Outcome measurement

I measure outcomes using two waves of household surveys; a midline survey that took place roughly one month and an endline survey that took place around eight months after the installation of alarms. The target was to interview the same two members in each household at midline and endline: the woman who had been interviewed at baseline and one randomly selected adult man. In all-women households, a second woman respondent was selected at random. Since 23 out of 250 households have only one member, the target sample size was 477 respondents. 438 of these respondents could be interviewed at midline giving a response rate of 92%. At endline, the response rate was 85% with 407 respondents completing the interview. Section B.2 of the appendix shows that the treatment does not appear to affect rates and patterns of attrition.

In addition to the so sampled respondents, more respondents were interviewed if other household members were available at the time of the interview.<sup>43</sup> As can be seen in section B.3 of the appendix, there is no statistically significant relationship between treatment and whether or how many additional respondents were interviewed in a given household. In line with the pre-analysis plan, analyses in this paper are based on all 483 respondents interviewed at midline and all 448 respondents interviewed at endline. As I show in section C.2 of the appendix, all main results are robust to sub-setting the sample to the two initially sampled respondents per household. See section B.1 of the appendix for evidence of covariate balance across experimental conditions and section A.7 for sociodemographic characteristics of the midline and endline samples.

Missing values due to non-response to outcome questions are imputed using multivariate im-

---

<sup>42</sup>This figure should be interpreted as suggestive, because it is unknown whether households assigned to the control group may also have refused the alarm system had they been assigned to it.

<sup>43</sup>During the midline survey, 45 additional respondents were interviewed across 39 households. During the endline survey, 39 additional respondents were interviewed across 38 households.

putation via chained equations. Outcomes are imputed within pre-specified outcome families (e.g. “vigilantism related outcomes”) and the imputation procedure does not condition on treatment status or covariates. All outcome measures range from zero to one. Individual items have been imputed prior to the creation of indices and indices are created by averaging constituent items. See section [D](#) of the appendix for the question wording and coding of all outcomes.

### 3.5 Estimation and hypothesis tests

The main analyses below seek to estimate the intent-to-treat (ITT) effect among respondents in the sample using the following linear regression specification:

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

$\mathbf{Y}$  here is a vector of outcomes;  $\alpha$  is an intercept;  $\tau$  is the ITT among respondents in the sample;  $\mathbf{z}$  is a vector of treatment assignments;  $\mathbf{n}$  is a vector of cluster sizes (number of respondents from a given household included in the analysis) and  $\delta$  is the associated coefficient;  $\boldsymbol{\epsilon}$  is a vector of error terms that allows for clustering at the household level. I control for the number of respondents per household, since heterogeneity in cluster size may bias estimates of the finite sample ITT if cluster size correlates with potential outcomes. In addition to this specification, the pre-analysis plan also contains a specification that controls for a set of covariates selected through lasso regression. Here, I focus on estimates based on the simpler of the two pre-registered specifications. Hypothesis tests are based on  $p$ -values calculated via randomization inference by permuting treatment assignment 2000 times in order to simulate the sampling distribution of the estimator under the sharp null hypothesis of no (positive or negative) treatment effect for any unit. Unless stated otherwise, the direction of hypothesis tests (upper, lower or two tailed) follows the pre-registration. See section [A.1](#) of the appendix for a summary of divergences from the pre-analysis plan.

### 3.6 Ethical considerations

In this section, I describe various ethical considerations and efforts that have been made to address these. A first question is whether the alarm treatment may have been likely to produce adverse effects. Given that it comes with a siren, for example, one may worry that the alarm could be used to instigate mob vigilantism. It is impossible, however, to trigger the alarm’s siren without also

sending text messages to the police. For this reason, the implementing partner who had already installed almost 2,000 alarm systems throughout South Africa prior to this study did not think it likely that the alarm would be used in this way. To guard against adverse consequences for alarm recipients, households were informed in detail about the functionality of the alarm and were given ample opportunity to refuse it. Only a small proportion of households did so.

In evaluating the ethical implications of the intervention, it is also important to keep in mind that the implementing partner would have installed the one hundred alarms in this precinct even if this study had not taken place. Similarly, this study did not artificially limit the number of alarms to be distributed to create a control group. Doing so would be ethically problematic if the alarm system had important welfare benefits. At the time of the study, the implementing partner did not have funding for more than one hundred alarms for this precinct.

Turning to the data collection process, two risks for the well-being of respondents were re-traumatization through sensitive survey questions and implication of respondents in illegal behavior. Regarding the former, survey questions were vetted extensively through pretesting and discussions with the local research team. Questions about crime victimization were kept at a high level of generality. For example, questions asked about household level rather than individual experiences with crime and did not go into detail about the kinds of crimes that respondents and their family members experienced. Women respondents were matched to women enumerators and enumerators were trained to interview respondents in private. With regard to the risk of implicating respondents, no questions were asked about respondents' participation in vigilantism or other illegal activities. Questions on such topics made use of hypothetical scenarios or asked respondents whether they knew off or witnessed incidents irrespective of their own participation.

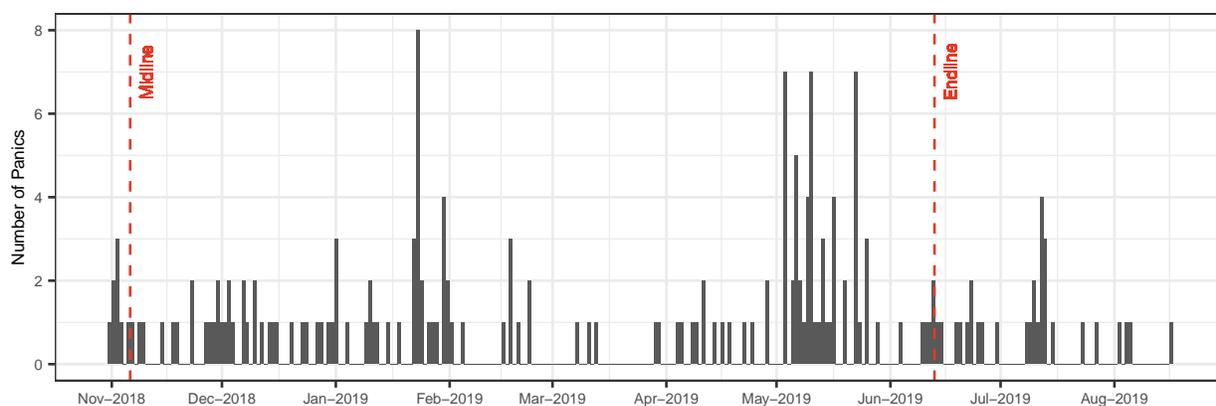
Finally, given the high crime rate, it was crucial to take steps to ensure the safety of study staff. In all survey waves, enumerators were residents of the study community, even though they were assigned to work in sections other than their own. Prior to starting work in a given section, local approvals were obtained from the relevant community leaders. Where local communities seemed uninformed about the survey or hostile, enumeration was stopped until problems were solved with local authorities. Police were aware of survey activities at all times. Enumerators always worked in pairs and were not allowed to enter a home by themselves. The times where enumerators had to walk in between households were kept to a minimum and enumerators were given bags such that

they would not carry expensive equipment such as tablets in the open. Wherever possible, the team stopped work before nightfall. Where enumerators conducted surveys after dark, a car was kept close and enumerators were dropped at their home after interviews were finished.

## 4 Main Results

### 4.1 Usage of the police alarm

Before considering how the alarm treatment affected household members, I provide a sense of the extent to which households made use of the alarm. Figure 6 plots the number of times any of the alarms in the study precinct was triggered per day across the study period.



**Figure 6:** Alarm panics over time.

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

The back end system of the implementing partner registered 159 alarm panics in the study precinct between 1 November 2018 and the start of the endline survey in mid-June 2019.<sup>44</sup> At least one panic was registered in 72 of the 94 households that received an alarm. In the endline survey, respondents from only 15 of the households in the treatment group reported that a household member had experienced a crime since the previous Christmas. A subset of registered panics are likely false alarms.<sup>45</sup> Alarms may also have been triggered in emergencies other than crime. In one case, a household triggered an alarm because the neighboring house was on fire. Finally,

<sup>44</sup>I collapse multiple panics registered for the same household on the same day into one panic, since such patterns likely result from a household member triggering the alarm more than once in the same situation.

<sup>45</sup>The motion sensor of the alarm is very sensitive and false alarms happen. Even those who are not the target of crime may thus receive a call from a police officer or have police come to their house as a result of the alarm treatment. One alarm recipient, for example, told the enumeration team that her child had silently triggered the alarm. The alarm owner expressed her surprise about finding police officers outside her door soon after.

alarms may sometimes be triggered as part of maintenance procedures of the implementing partner. Importantly, even panics that are not related to crime may result in a police response.

## 4.2 Effects on the willingness to rely on police and participate in mob vigilantism

Table 22 displays estimates of the effect of the alarm treatment on the two main outcomes of interest, citizens' willingness to rely on police and to participate in mob vigilantism. The main takeaway from the table is that the alarm treatment seems to have increased the willingness to rely on police and decreased the willingness to participate in 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. The first is an item that captures respondents' inclination to reach out to police if someone is trying to enter their home to steal from them. The second is a subindex of multiple items that measure respondents' general proclivity to share information about crime with police. Column 1 suggests that the alarm treatment increased the willingness to rely on police at midline by roughly one third of a control group standard deviation ( $p < 0.01$ ). Column 2 shows that the estimated effect at endline is slightly smaller but highly statistically significant ( $p < 0.01$ ). In short, the alarm treatment seems to have encouraged reliance on police in both the short and medium term. Table 23 in section C.3 of the appendix shows that estimates are similar for both index components.

Estimates shown in columns 5 and 7 stem from models that allow treatment effects to vary across prior beliefs. To the extent that it causes household members to learn about police, the alarm treatment may create surprises among those who expected little from police at baseline, while reaffirming beliefs or even disappointing among those who expected a lot. As pre-registered, I consider prior beliefs about the two theoretically relevant police outputs – police service quality and the risk of state punishment for mob vigilantism. Prior beliefs are captured using household-level measures from the baseline survey (see table 1). Column 5 presents estimates from an analysis that regresses the outcome on the treatment indicator, an indicator for high prior beliefs about the risk of state punishment for vigilantism and the interaction between the two. Column 7 presents estimates from the equivalent specification for prior beliefs about police service delivery.<sup>46</sup>

Both columns indicate that effects on the willingness to rely on police are concentrated among

---

<sup>46</sup>This analysis marginalizes across the rows/columns of table 1. Due to limited power, I did not pre-specify that I would analyze effects within all four cells of this table.

those who expected little from police at baseline. Among those who had low prior beliefs about, respectively, the risk of being arrested for vigilantism and police service delivery, the alarm treatment 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 that effects among the two high prior groups are statistically significantly smaller and close to zero ( $p < 0.05$  and  $p < 0.1$ ).

Next, I turn to effects on the willingness to participate in mob vigilantism. I measure this outcome using one item at midline and an index of the same and another item at endline. Both measures present respondents with a scenario in which the community has gotten hold of a criminal and ask respondents about their likely course of action. The item that was measured in both survey waves gives respondents 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 second item asks only whether respondents would participate in beating the suspect.

Column 3 suggests that the alarm treatment decreased the willingness to participate in vigilantism at midline by roughly one fifth of a control group standard deviation. The estimate is highly statistically significant ( $p < 0.01$ ) and implies a reduction in the proclivity to inflict vigilante violence by around thirty percent relative to the control group mean. The estimated effect at endline is substantially smaller and falls short of statistical significance.<sup>47</sup>

Once I allow for treatment effects to vary across low and high prior groups, however, I find evidence of a negative effect even at endline. Column 6 suggests that the alarm treatment decreased the willingness to participate in vigilantism among those with low prior beliefs about the likelihood of being arrested for doing so by about one third of a control group standard deviation ( $p < 0.05$ ). This decrease is of roughly of the same relative magnitude as the increase in the willingness to rely on police among this subgroup. The interaction term indicates that the alarm treatment's effect on the willingness to participate in vigilantism is statistically significantly less negative among the corresponding high prior group ( $p < 0.01$ ). As can be seen in column 8, prior beliefs about service delivery appear to condition effects in similar ways, but the patterns are less pronounced.

---

<sup>47</sup>Table 25 in section C.3 of the appendix shows that the apparent absence of an effect at endline is not an artifact of the inclusion of another item in the outcome index. Using only the item that was measured in both waves, the effect estimate at endline is only half the size of the estimated effect at midline and falls short of statistical significance.

	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
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 treatment on respondents' willingness to rely on police and participate in mob vigilantism.

All outcome measures range from 0 to 1. The results in columns 5 to 8 are based on specifications that 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 5 and 6) 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 7 and 8) 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.6 for 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 and testing. See section D.1 for question wording and coding of outcomes.

Taken together, the results presented in table 22 suggest that the alarm treatment encouraged reliance on police and discouraged vigilantism, especially among parts of the sample that expected little from police at baseline. These findings are in line with the central prediction that an increase in police capacity helps state institutions to supersede informal enforcement mechanisms.

In the appendix, I present evidence of the extent to which these results generalize to other, related outcomes. Table 18 in section C.1 of the appendix contains little evidence of a treatment effect on the share of respondents who have recently spoken to police. The apparent absence of an increase in contact with police may appear surprising given the high number of alarm panics and 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, which would reduce the need for households in the treatment group to reach out to police. Respondents in the treatment group are indeed more likely to say they feel safe. The evidence of a reduction in crime victimization, however, is limited (see table 31 in section C.5.3 of the appendix).

Table 19 in section C.1 of the appendix suggests that the alarm treatment 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 punishment for vigilantism.<sup>48</sup> That said, those who were assigned to receive an alarm do not appear less inclined to reach out to their neighbors in case of a criminal attack. In short, even though respondents in the treatment group have become more reluctant to personally engage in and support vigilantism, these respondents would still rely on their community for help in emergencies.

This pattern of results may be interpreted as a first clue regarding the question of mechanisms. Fear of state punishment should matter most for outcomes that concern vigilante violence. Merely alerting one's neighbors if one is at risk of being robbed is not a crime. Hence, if the effects of the alarm system are due to increased concerns about state punishment, the apparent absence of an effect on this last outcome is no surprise. If the alarm system worked by convincing citizens that police provide high quality services, on the other hand, one would expect it to reduce demand for *all* kinds of community involvement, irrespective of whether such involvement is illegal.

In the next section, I provide more evidence on the mechanisms through which the alarm system may have produced its two main effects – an increase in the willingness to rely on police and a

---

<sup>48</sup>Tables 24 and 25 in section C.3 of the appendix show results for individual items.

decrease in the willingness to engage in vigilantism.

## 5 Mechanisms

### 5.1 Conceptual framework

Why does the alarm treatment discourage vigilantism and encourage reliance on police? Subsequently, I state my theory in formal terms to derive predictions that distinguish between the two hypothesized mechanisms. Consider again the choice between reporting a criminal to police or taking the law into one’s own hands. Above, I have described several factors which may affect this choice. First, state and community punishments differ in harshness and publicity which may cause citizens to favor one or the other. Let  $x_v$  be the utility that citizens derive from community punishment and  $x_s$  their utility from state punishment. Second, citizens may consider the probability that calling the police or rallying a vigilante mob does indeed result in punishment of the criminal. I denote these probabilities by  $p_s$  and  $p_v$ . A final consideration is that those who participate in vigilantism may receive a prison sentence. I denote by  $q$  the probability of state punishment for vigilantism and by  $y$  the resulting utility loss.

Both  $p_s$ , the probability that reliance on the state results in punishment of the criminal, and  $q$ , the probability that the citizen faces punishment for her involvement in vigilantism are outputs of police activity. Specifically, it will be helpful to think of these outputs as a function of two inputs, police capacity and police effort. By police capacity, I refer to factors that help police be more effective at whatever task they choose to pursue.<sup>49</sup> For example, the more personnel a police chief has at her disposal and the more motivated officers are to perform, the better police may be able to solve crimes of any kind. Whether crimes of a particular kind get solved will also depend on how much effort police dedicate to this crime as opposed to others. Let  $e_s$  be the effort that police would make to convict the perpetrator that the citizen would like to see convicted and  $e_v$  be the effort that police will make to convict the citizen if she engages in vigilantism. Denoting police capacity by  $\theta$ , the two outputs of interest can be written as  $q(e_v, \theta)$  and  $p_s(e_s, \theta)$ . Both functions are assumed to increase in their two arguments.

Taken together, a citizen will choose to take the law into her own hands if the expected utility

---

<sup>49</sup>See [Berwick and Christia \(2018\)](#) for an overview of different ways to conceptualize the capacity of state institutions. The notion put forward here is similar to [Hanson and Sigman \(2013\)](#)’s definition of state capacity as the “ability of state institutions to effectively implement official goals.”

of community punishment net of the expected legal costs exceeds the expected utility of reporting the crime to police:

$$U_v - U_s = \underbrace{p_v x_v}_{\text{exp. utility of MV}} - \underbrace{q(e_v, \theta)y}_{\text{exp. legal cost of MV}} - \underbrace{p_s(e_s, \theta)x_s}_{\text{exp. utility of state justice}} \geq 0. \quad (1)$$

The key idea is that the police alarm serves as a shock to perceptions of police capacity. That they have become more legible to police may convince households that police have acquired a greater *general* ability to affect the lives of household members. While not impossible, it seems less likely that the alarm treatment would affect perceptions of what *specific* kind of actions police would prioritize. For example, even though I find that the alarm treatment increased respondents’ stated willingness to rely on police, I do not find evidence of an increase in actual contact with police that could serve as an opportunity to learn about police priorities. Neither was the alarm system presented to respondents as a solution to a particular kind of crime. Equation (1) makes explicit that a shock to police capacity may affect citizen behavior through both, its effect on the expected costs of vigilantism and its effect on the expected utility of reliance on the state. As a first step towards understanding the relevance of these mechanisms, I will test whether the alarm treatment did indeed affect citizens’ perceptions of police service quality  $p_s$  and the risk of state punishment for vigilantism  $q$ .

Merely showing that the alarm treatment changed citizens’ views of police outputs, however, says little about the extent to which these changes contributed to the effect of the alarm treatment on the willingness to rely on police and participate in vigilantism. I use two strategies to generate additional evidence on the relative importance of the two mechanisms.

First, equation (1) suggests that treatments which individually manipulate perceptions about police effort levels  $e_s$  and  $e_v$  would trigger only one of the mechanisms. Such treatments would thus allow for an assessment of how effective each of the two mechanisms is at discouraging vigilantism. Section 5.3 presents results from a mechanism experiment that was designed to achieve this aim.<sup>50</sup>

Second, I leverage theoretical expectations about how police efforts and capacity may interact to affect the willingness to participate in vigilantism. In the context of the alarm treatment shock

---

<sup>50</sup>This exercise is similar in spirit to what has been termed a “mechanism experiment” by Ludwig, Kling, and Mullainathan (2011). It also bears resemblance to the notion of implicit mediation analysis (Gerber and Green, 2012).

to police capacity, it seems intuitive that police capacity and efforts would be complements, i.e., in formal terms that  $\frac{\partial^2 q(e_v, \theta)}{\partial \theta \partial e_v} > 0$  and  $\frac{\partial^2 p_s(e_s, \theta)}{\partial \theta \partial e_s} > 0$ . The alarm should facilitate a speedier police response to crime, for example, but only to the extent that police attempt to find the household in question. Similarly, members from households with an alarm may think that police could use information about the identity of household members to identify and arrest them for participating in vigilantism. This change should worry respondents only to the degree that they expect police to make efforts to investigate cases of vigilantism.

Suppose that the effect of the alarm treatment is mediated to some degree by an increase in the risk of punishment for vigilantism. If so, the discouraging effect of the alarm system on vigilantism should be stronger the more convinced citizens are that police invest significant effort  $e_v$  to bring perpetrators of vigilantism to book. Similarly, to the extent that the effect of the alarm system is mediated by its effect on the expected utility of relying on police, the alarm treatment should be more effective at discouraging vigilantism if citizens are convinced that police make significant efforts  $e_s$  to provide services that citizens desire. I make use of the mechanism experiment to test these predictions.

Subsequently, I first present estimates of the effects of the alarm treatment on perceptions of police outputs before I describe the design and results of the mechanism experiment.

## 5.2 Effects on perceptions of police

Table 3 presents estimates of the effects of the alarm treatment on perceptions of police. All outcomes are coded such that positive estimates indicate a treatment-induced increase in respondents' expectations about police capacity and outputs.

The outcome in column 1 is an index of two items. The first asks respondents whether someone from the local police knows the location of their house, while the second asks whether local police know the respondent's name or the name of another household member. This outcome can be thought of as a manipulation check. Did the alarm treatment indeed convince citizens that they have become more "legible" to police? The coefficient in column 1 indicates a positive shift in respondents' sense of being known to police by a little more than one fifth of a control group standard deviation ( $p < 0.05$ ). In relative terms, this change corresponds to an increase of over 20% from a control group mean of 0.44.

	<i>Legibility (<math>\theta</math>)</i>		<i>Police service quality (<math>p_s</math>)</i>			<i>Risk of state punishment (<math>q</math>)</i>		
	Know HH	Arrive quickly		Service quality		Would discover	Respond MV	Imprison MV
	Endline	Midline	Endline	Midline	Endline	Endline	Endline	Endline
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Alarm	0.100** (0.047)	0.059** (0.033)	0.091*** (0.036)	0.067** (0.029)	0.041 (0.034)	0.039* (0.026)	0.004 (0.035)	0.041 (0.043)
Control Mean	0.44	0.38	0.45	0.44	0.56	0.78	0.67	0.71
RI $p$ -value	0.016	0.035	0.005	0.022	0.12	0.067	0.448	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 3:** Effects of alarm treatment 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 section D.2 for question wording and coding of outcomes.

Columns 2 to 5 suggest that the alarm treatment caused respondents to perceive a better quality of police service. Columns 2 and 3 provide evidence that the police alarm had enduring effects on expectations about the speed with which police would respond when called to an emergency in one’s home. The outcome is a scale that ranges from “The police would not come” coded as 0 to “The police would come in less than thirty minutes” coded as 1. At midline, the alarm treatment is estimated to have increased this outcome by around 16% from a control group mean of 0.38 ( $p < 0.05$ ). The effect appears, if anything, slightly larger at endline.

The outcomes in columns 4 and 5 are indices that combine a range of items designed to capture respondents’ more general expectations about police service quality. Tables 26 and 27 in section C.3 of the appendix show estimates of treatment effects on individual items. There is some evidence of increased expectations of police service quality that go beyond improved response times. The midline results indicate that the alarm treatment increased the police service quality index by around 15% ( $p < 0.05$ ), for example, and the share of midline respondents who think that police are successful at sending those who are guilty to prison is almost ten percentage points higher in the treatment than in the control group ( $p < 0.05$ ). Changes seem less pronounced at endline, suggesting, perhaps, that respondents lowered some of their expectations as the novelty of the alarm treatment wore off. Column 5 points towards an increase in the police service quality index by roughly 0.04 scale points at endline, but the estimate falls short of statistical significance.

Turning to perceptions of the likelihood of punishment for illegal behavior, the outcome in column 6 is an index of two items which ask respondents how likely it is that police would find out if respondents bought a stolen car or rented space to an illegal immigrant. These behaviors were selected to avoid offending respondents by asking them to imagine having committed acts that are unambiguously illegal. The estimate in column 6 is indicative of an increase in the perception that police would find out about these behaviors of around 0.04 scale points ( $p < 0.1$ ).

Finally, columns 7 and 8 pertain to perceptions that relate directly to the risk of state punishment for vigilantism. The outcome in column 7 measures respondents’ perceptions of the speed with which police would respond to a vigilante incident and the outcome in column 8 captures whether respondents believe that police will make sure that those who take the law into their own hands go to prison. Contrary to the outcome in column 6, both measures are not specific to the legal risk faced by the respondent him- or herself. Hence, any changes in these outcomes would

have to be a result of respondents updating their general views about police. The results shown in columns 7 and 8 provide little evidence of such updating in the sample as a whole. Even though the share of respondents who believe that police ensure that those who take the law into their own hands go to prison is around four percentage points higher in the treatment than in the control group, this estimate falls short of statistical significance.

Recall, however, that learning about police may be conditional on prior beliefs and that effects on the main behaviors of interest were concentrated among households with low prior beliefs about police outputs at baseline. Table 4 therefore, again, allows treatment effects to vary across prior beliefs about police service quality and the risk of state punishment for vigilantism.

Columns 1 and 4 of this table contain little evidence of a treatment effect on the perceived speed with which police would respond to vigilantism, even among those who expected little from police at baseline. Estimates of effects among low prior subgroups in columns 2 and 5, however, indicate a statistically significant increase of roughly ten percentage points in the share of respondents who think that police ensure that perpetrators of vigilante violence go to prison ( $p < 0.1$  and  $p < 0.05$ , respectively). The interaction terms suggest that treatment effects on this outcome are close to zero in both high prior subgroups. Columns 3 and 6 show similar patterns in terms of effects on perceptions about police service quality. Again, treatment effects appear concentrated among those with low priors and effects among high prior subgroups are statistically significantly smaller.

To some extent, it may seem surprising that prior beliefs about one kind of police output appear to condition treatment effects on perceptions of another kind of police output. For example, those with low prior beliefs about police service quality see statistically significantly larger treatment effects than those with high prior beliefs about police service quality on perceptions of both police service quality *and* the proclivity of police to send perpetrators of vigilantism to prison. It is important to keep in mind, however, that the two dimensions of prior beliefs are not independent. As suggested by equation 1, low expectations of one kind of police output may be reflective of the general conviction that police have little capacity  $\theta$ . Updating beliefs about  $\theta$  may lead to increased expectations with regard to various kinds of police outputs.

	<i>Risk of state punishment (q)</i>		<i>Service quality (p<sub>s</sub>)</i>	<i>Risk of state punishment (q)</i>		<i>Service quality (p<sub>s</sub>)</i>
	Respond MV	Imprison MV	Service index	Respond MV	Imprison MV	Service index
	(1)	(2)	(3)	(4)	(5)	(6)
Alarm	0.047 (0.052)	0.091* (0.066)	0.117*** (0.045)	0.026 (0.045)	0.110** (0.055)	0.110*** (0.040)
Alarm × High Prior Punishment	-0.078 (0.070)	-0.075 (0.088)	-0.114** (0.062)			
Alarm × High Prior Service				-0.043 (0.073)	-0.145** (0.089)	-0.095* (0.062)
Control Mean	0.67	0.71	0.53	0.67	0.71	0.53
RI <i>p</i> -value Main	0.216	0.07	0.009	0.238	0.032	0.006
Hypothesis Main	upr	upr	upr	upr	upr	upr
RI <i>p</i> -value Diff.	0.153	0.136	0.032	0.21	0.047	0.072
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:** Heterogeneity in effects of alarm treatment on perceptions of police at endline by prior beliefs

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.6 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.

The results in this section provide evidence that the alarm system caused respondents to expect more from police, both in terms of the quality of service police provide and in terms of the extent to which police are able or inclined to ensure that respondents face punishment for illegal behavior including participation in vigilantism. Effects of both kinds appear concentrated among the low prior groups that also saw the greatest increase in the willingness to rely on police and the largest decrease in the willingness to participate in vigilantism. The results presented so far thus provide evidence consistent with the relevance of both mechanisms.

### 5.3 Mechanism experiment

#### 5.3.1 Design

Even though both mechanisms may be at play, they may not be equally important. To shed light on the effectiveness of each mechanisms, I designed a mechanism experiment that features two information treatments in the form of local news articles (see section A.8 of the appendix). The first article is meant to convince respondents that police make effort to deliver services that citizens desire (“Police fight crime” treatment). It focuses on two cases of rape and the police’s efforts to convict perpetrators of crimes against women and children. Pre-testing among respondents from a similar township close by suggested that the demand for conviction of perpetrators of such crimes is high and that most respondents view the described police efforts as an important part of service delivery.<sup>51</sup> The second article is meant to affect beliefs about police effort to convict perpetrators of mob vigilantism (“Police fight MV” treatment).

Information treatments were administered during the endline survey and assigned on the respondent level using simple random assignment. Respondents were assigned to one, both or none of the two information treatments, creating a full factorial design. Information treatments were randomized across all respondents in the endline sample which includes two members of one neighboring household for each of the 250 study households.<sup>52</sup> In total, the endline sample comprises  $N = 815$  respondents. Analyses that pertain to the information treatments alone are based on respondents from main and neighboring households as pre-registered.

Enumerators read out the relevant article(s) during the endline interview and asked several

---

<sup>51</sup>Note also that mob vigilantism rarely targets women and children, which reduces the likelihood that this treatment would affect perceptions about the police’s approach to mob vigilantism.

<sup>52</sup>See sections A.5 and A.6 of the appendix for details on the sampling strategy for neighbors.

open ended questions about respondents’ opinions in order to encourage engagement with article content. Enumerators were not aware that the goal was to understand the effects of the information treatments on subsequent survey responses and, instead, thought that the aim was to elicit respondents’ views about the articles. Outcomes were measured in later parts of the questionnaire. Given limited questionnaire space, outcome measures are limited to respondents’ perceptions of police effort and their willingness to participate in mob vigilantism.

Table 5 shows the breakdown of respondents across information treatment conditions. In practice, all subsequent analyses will marginalize across one of the two dimensions of the factorial design. Effects of the “Police fight crime” treatment, for example, will be estimated by regressing outcomes on an indicator for whether a respondent was assigned to this treatment. Effects of the “Police fight mob vigilantism” treatment will be estimated in an analogous way.

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

**Table 5:** Number of respondents across information treatment conditions.

The first number in each cell pertains to the total number of respondents from main households and neighbors in that condition. The number in parentheses pertains to respondents from main households only. The  $p$ -value of a randomization inference based Chi-squared contingency table test is 0.294.

### 5.3.2 Results

I begin by analyzing effects of the information treatments, marginalizing across assignment to the the alarm treatment. Table 28 in section C.4 of the appendix suggests that neither of the information treatments is effective at shifting respondents’ expectations about police effort in the endline sample as a whole. Again, new information may be most effective at shifting beliefs among those who, a priori, expect police effort levels to be low. Hence, the pre-analysis plan specified that, in the event that information treatments appeared ineffective, their effects would be analyzed within two low prior subgroups. Analyses in table 6 are based on the sub-sample of respondents with low prior beliefs about the police’s inclination to ensure arrest of participants of mob vigilantism.<sup>53</sup>

Column 1 of table 6 suggests that the “Police fight crime” treatment is effective at shifting respondents’ beliefs about police effort in this subgroup. Respondents who received this treatment

<sup>53</sup>Analyses that do not involve the alarm treatment make use of measurements of priors from the endline survey that were asked prior to the administration of the information treatments.

are almost twelve percentage points more likely to believe that “the police do everything they can to ensure that criminals receive the punishment that they deserve” ( $p < 0.05$ ). This shift represents an increase in the perception of police service delivery effort of around 50%. As can be seen in column 3, the estimated effect of this treatment on respondents’ expectations about whether police ensure that perpetrators of mob vigilantism go to prison, however, is small and statistically insignificant. 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 perpetrators of mob vigilantism to prison by around 22% ( $p < 0.1$ ), there is no evidence of an effect of this treatment on expectations about the inclination of police to make sure that “criminals receive the punishment that they deserve.”

	Believes police fight crime	Believes police fight MV	Would participate MV			
Police fight Crime	0.115** (0.058)	-0.005 (0.052)	-0.005 (0.049)			
Police fight MV	0.034 (0.059)	0.083* (0.052)	-0.087** (0.048)			
Control Mean	0.23	0.28	0.42	0.37	0.39	0.43
RI p-value	0.032	0.276	0.553	0.051	0.458	0.039
Hypothesis	upr	upr	upr	upr	lwr	lwr
Observations	243	243	243	243	243	243

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

**Table 6:** Effect of information treatments among respondents with low priors about risk of state punishment for mob vigilantism

All outcome measures range from 0 to 1. Analyses are run on a subset of respondents from main and neighboring households that has low priors about the risk of legal repercussions for mob vigilantism as measured during the endline survey, prior to the administration of information treatments. See section D.6 for the endline item used to measure prior beliefs. Columns 1, 3 and 5 regress the outcome on an indicator for assignment to the ‘Police fight crime’ treatment. Columns 2, 4 and 6 regress the outcome on an indicator for assignment to the ‘Police fight MV’ treatment. Standard errors are heteroskedasticity robust. See section A.2 of the appendix for information on model specification and section D.4 of the appendix for outcome question wording and coding.

These results suggest that the information treatments were successful at independently shifting perceptions about the two kinds of police effort. The question of interest is the extent to which these shifts translate into changes in the willingness to participate in mob vigilantism. Column 5 contains little evidence that the “Police fight crime” treatment has an effect on the willingness to participate in vigilante violence. Column 6, however, suggests that the “Police fight MV” treatment

has decreased the willingness to participate in vigilantism by around 20% ( $p < 0.05$ ).

These findings are mirrored to some degree in table 29 in section C.4 of the appendix which presents estimates among respondents with low prior beliefs about service delivery. Here, too, the “Police fight crime” treatment appears to have caused an upward shift in expectations about service delivery; yet, this shift does not appear to translate into a reduction in the willingness to participate in mob vigilantism. The “Police fight MV” treatment, on the other hand, seems 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 produced a shift in beliefs about service delivery efforts in both low prior subgroups, there is little evidence that either of these shifts translated into a change in the willingness to participate in vigilantism. The “Police fight MV” treatment seems to have produced a shift in beliefs about police efforts to sanction perpetrators of vigilantism only among those who, a priori, did not expect police to make such effort. Among this subgroup, the “Police fight MV” treatment also seems to have discouraged participation in vigilante violence. Taken together, these findings suggest that an increase in the perceived risk of state punishment for vigilantism is, at least in the short run, more effective at discouraging vigilantism than an increase in the perception that police are committed to convicting perpetrators that citizens would like to see punished.

Finally, I turn to the interaction between the alarm and information treatments. Table 7 shows estimates from specifications that regress the measure of respondents’ willingness to participate in vigilantism on an indicator for assignment to the alarm treatment, an indicator for assignment to one of the information treatments as well as the interaction between the two. Models in columns 1 and 3 are based on endline data from respondents in the 250 main study households. Models in columns 2 and 3 exclude respondents who were assigned to one of the information treatments.

The first row in columns 1 and 2 shows that the estimated effect of the alarm treatment on the willingness to participate in vigilantism among those who were not assigned to the “Police fight crime” treatment is small and not statistically significant. The interaction terms in these columns provide no evidence that the “Police fight crime” treatment made the alarm treatment more effective at discouraging vigilantism. The interaction term in column 1 is positive rather than negative, while that in column 2 is close to zero. Both estimates fall short of statistical significance.

	Would Participate Vigilantism			
	All respondents	Police fight MV = 0	All respondents	Police fight crime = 0
	(1)	(2)	(3)	(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
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 7:** Interactive effects of alarm and information treatment on willingness to participate in mob vigilantism

All outcome measures range from 0 to 1. Analyses in columns 1 and 3 are based on all respondents from main households in the endline survey. Columns 2 and 4 subset this sample to those who were not assigned to one of the information treatments. All models regress the outcome on an indicator for assignment to the alarm treatment, an indicator for assignment to the respective information treatment and the interaction between the two. In addition, all specifications control for cluster size. The row labeled “Control Mean” shows the average outcome among respondents who were neither assigned to the alarm treatment nor to the respective information treatment. See section D.5 in the appendix for wording and coding of outcome questions. Standard errors are clustered on the household level. Hypothesis tests are based on randomization inference. See section A.2 in the appendix for details on model specification.

Turning to columns 3 and 4, the estimated effect of the alarm treatment on the willingness to participate in vigilantism among those who were not assigned to the “Police fight MV” treatment is again small and not statistically significant. Yet, in this case, both interaction terms are negative and statistically significant, suggesting that the “Police fight MV” treatment increased the extent to which the alarm treatment discourages vigilantism ( $p < 0.1$  and  $p < 0.05$ ). The model in column 3 suggests that the alarm treatment decreased the willingness to participate in vigilantism among those who were assigned to the “Police fight MV” by around 0.06 scale points ( $p < 0.1$ ). This model marginalizes across the “Police fight crime” treatment, meaning that some of those who were assigned to the “Police fight MV” treatment were also assigned to the “Police fight crime” treatment. The model in column 4 subsets to respondents who were not assigned to the “Police fight crime” treatment and hence provides a cleaner comparison in cases where the two information treatments interact. This model suggests that the alarm treatment has decreased the willingness to engage in vigilantism by around 0.13 scale points among those who were also assigned to the “Police fight MV” treatment ( $p < 0.05$ ).

Priming respondents to believe that police make efforts to convict those who take the law into their own hands thus appears to make the alarm treatment more effective at discouraging vigilantism. Priming citizens to think that police are committed to sanctioning perpetrators whom most citizens would like to see convicted appears to have, if anything, the opposite effect. This pattern of results again points towards the importance of an increase in the perceived risk that vigilantism may lead to a prison sentence rather than to improved satisfaction with police service quality as a link between police capacity and mob vigilantism.

## 6 Alternative Explanations

In this section, I consider the possibility that the estimated effects of the alarm treatment may be due to factors other than the two mechanisms discussed in this paper and present evidence that speaks against major alternative explanations.

Given that all outcome measures in this project are survey-based, an obvious concern is that results may be driven by experimenter demand. Respondents were asked during the baseline survey whether they would be interested in a police alarm. Hence, respondents were likely able to draw a connection between the alarm treatment and outcome measurement.

While it seems difficult to rule out completely, several observations speak against the interpretation that the findings are driven solely by experimenter demand. First, several steps were taken to limit the perceived connection between the alarm treatment and subsequent surveys. Respondents were never asked any questions about the alarm during mid- and endline interviews and enumerators were not aware that the purpose of the study was to assess effects of the alarm. Second, if results were driven by experimenter demand, it appears difficult to account for the apparent concentration of treatment effects on certain outcomes and among certain subgroups. It seems unclear, for example, why respondents would see the need to censor their views about some aspects of police performance but not others. Similarly, it is not obvious why respondents with low priors about the police would be most susceptible to such bias.

Given that it is a crime, one may be most worried about experimenter demand driving observed differences in outcomes that relate to vigilantism. To assess this possibility, the endline survey asked respondents whether and how many incidents of mob vigilantism they recall happening in their section during May, June, and July 2018 and whether they personally witnessed any of these. Given that alarm installations took place in September and October of the same year, the alarm treatment could not have affected these outcomes. A tendency among respondents in the treatment group to report that they remember or witnessed fewer incidents than respondents in the control group would thus be evidence of a treatment-induced reluctance to be associated with mob vigilantism. Yet, as can be seen in table 30 in section C.5 of the appendix, there is no evidence of such a tendency, neither among the sample as a whole, nor among the subgroup that sees the largest decrease in the willingness to participate in mob vigilantism.

A second concern may be that treatment effect estimates reflect changes among the control rather than the treatment group. For example, control group respondents may have been disgruntled because they did not receive a police alarm which may have caused them to become more frustrated with police. Alternatively, police may have focused all their efforts on households that received an alarm, neglecting other households in the precinct. Figure 12 in section C.5 of the appendix shows, however, that men and women in the control group developed more positive views about the police and became less inclined towards mob vigilantism over time. If anything, these patterns suggest that their knowledge of the alarm project may have led control group respondents to shift their views in ways that are similar to the changes that occurred in the treatment group.

If so, the estimates reported here would understate the true effects of the alarm treatment.

A final possibility is that observed differences in the willingness to engage with police and in views about vigilantism are driven by changes in respondents' preferences about the nature of punishment. For example, the alarm treatment may make household members feel safer, which in turn may decrease their demand for maximal deterrence through harsh and immediate punishments in the form of vigilante violence. Table 31 in section C.5 of the appendix shows that, even though there is little evidence of a reduction in household level crime victimization, members of households that were assigned to an alarm are indeed more inclined to report that they feel safe in their home. Beyond that, however, the estimates in the table provide little support for the hypothesis that the alarm treatment caused a reduction in respondents' demand for harsh and immediate punishments.

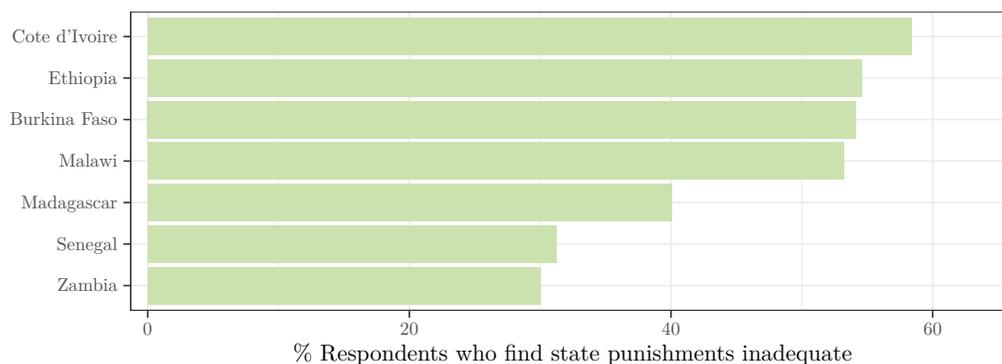
## 7 Discussion

This paper investigates the microdynamics that underlie the relationship between state capacity and the popularity of community-based alternatives to the state's law enforcement system. Many have suggested that the prevalence of informal ways to deal with crime is indicative of the inability of state institutions to provide citizens with effective law enforcement services (see e.g. [Acemoglu et al., 2020](#); [Baker, 2008](#); [Tankebe, 2009](#)). If state institutions such as police were more effective, these arguments seem to imply, citizens would voluntarily substitute away from informal alternatives such as vigilantism and choose to rely on the state.

Drawing on experimental variation in the ability of police to intervene in certain households but not others, I show that an increase in police capacity did indeed encourage reliance on police and discourage vigilantism. That said, citizens who were assigned to a positive shock to police capacity developed not only more sanguine views of police service quality. They also became more convinced that participation in vigilante violence may lead to state punishment. Results from a mechanism experiment suggest that the risk of legal repercussions may play a bigger role in the decision of whether to participate in vigilantism than expectations about police service delivery.

Why may improvements in citizens' expectations about the quality of police services have limited effects on vigilantism? One possibility is that this finding results from characteristics of the research design. The results presented in this paper have a relatively short time horizon, for example; and the effects of improvements in service delivery may take longer to materialize. Higher quality services

may also be more effective at discouraging vigilantism if such improvements result directly from actions taken by state officials and not, as is the case here, from a technological windfall provided by a non-profit organization.



**Figure 7:** Views on state punishment in Sub-Saharan Africa

Data are taken from a 2017 survey by the [World Justice Project](#). For each country, the sample consists of around 1,000 respondents from major urban centers. Respondents were asked: “Please tell us how confident you are that the criminal justice system as a whole gives punishments which fit the crime?” The figure displays the proportion of respondents who chose “Not very confident” or “Not at all confident.”

A more structural explanation may be that state justice, even if administered effectively, provides an imperfect substitute for community justice. [Smith \(2019\)](#) argues, for example, that South Africans resort to vigilantism because they wish for criminal suspects to enjoy few due process protections and for punishments to be harsher than those provided by the state. Nationally representative surveys indeed suggest that more than 50% of South Africans are dissatisfied with the courts, the most frequent complaint being that sentences are too lenient.<sup>54</sup> Figure 7 shows that such discontent with the state’s punishment regime is widespread also in other countries in Sub-Saharan Africa. Where citizens are fundamentally opposed to how the state treats law-breakers once they have been apprehended, states are unlikely to be able to out-compete community punishment by improving police service delivery.

The results presented here suggest that police capacity may nonetheless play a role in discouraging vigilantism, because it allows the state to effectively place sanctions on those who engage in vigilante violence. Reluctance of citizens in high capacity states to engage in vigilantism need not reflect that citizens are satisfied with police services. Such reluctance may also be driven by the

<sup>54</sup>These numbers are based on the Victims of Crime Survey 2016/17 by StatsSA. Out of 54% of respondents who are dissatisfied with the courts, 46% point to lenient sentences as the reason.

expectation that those who take the law into their own hands will go to prison.

Beyond informal enforcement mechanisms like vigilantism, similar logics may also apply to other informal practices that are widespread in low capacity contexts. Unlicensed health providers, for example, are ubiquitous throughout the developing world. To some degree, this prevalence may be driven by limited access to formal health care. Yet, such providers sometimes supply controversial remedies,<sup>55</sup> 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 the quality of government services improves. Another example are unlicensed moneylenders and loan sharks 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. As in the case of vigilantism, a lower prevalence of these informal services in high capacity contexts may in part reflect the state's ability to crack down on suppliers.<sup>56</sup>

A broader implication is that state capacity can have downsides for certain groups of citizens. Those who favor vigilantism or other informal practices that the state deems illegal may be wary of increased state presence, even if it comes with improvements in the quality of government services. Residents of the study precinct seemed cognizant of this trade-off. Not only was the willingness to participate in vigilantism particularly widespread among those who refused a police alarm, but those who received an alarm sometimes asked to have its siren installed at the back of their house. Installing the siren out of sight seems counterintuitive if the police alarm's only effect is better police protection against intruders. If the police alarm is known to also deter vigilantism, however, the decision to hide the alarm may reflect a desire to enjoy the benefits of improved police service delivery while upholding the threat of community punishment. In line with the interpretation that alarm owners sought to hide the alarm, I find little evidence that effects of the alarm spilled over to neighboring households.

A final question concerns the incentives of state officials. Given widespread public support for

---

<sup>55</sup>In Tanzania, for example, traditional healers have been involved in killings of albinos whose body parts are widely believed to have healing powers. See <https://www.npr.org/sections/goatsandsoda/2015/01/16/377505104/can-a-new-ban-on-witchcraft-protect-the-albinos-of-tanzania> for details.

<sup>56</sup>Attempts to regulate informal services are common throughout the world. Tsai (2004), for example, describes how governments in India and China have attempted to ban informal lending and the government of Tanzania has attempted to ban traditional medicine in an attempt to stop the killing of albinos.

vigilantes, why do politicians and bureaucrats choose to enforce laws against vigilantism? Alternative courses of action include simply looking the other way or even changing the state's punishment regime in such a way that it aligns more closely with the preferences of vigilantes. An important next step towards understanding how states supersede informal approaches to crime control will thus be to consider the incentives 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.
- Aldashev, Gani, Imane Chaara, Jean-Philippe Platteau, and Zaki Wahhaj. 2012. "Using the law to change the custom." *Journal of Development Economics* 97 (2): 182–200.
- Ali, Merima, Odd-Helge Fjeldstad, and Ingrid Hoem Sjørusen. 2014. "To pay or not to pay? Citizens attitudes toward taxation in Kenya, Tanzania, Uganda, and South Africa." *World Development* 64: 828–842.
- 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*. Nordiska Afrikainstitutet.
- Barnwal, Prabhat. 2017. Curbing leakage in public programs: Evidence from Indias direct benefit transfer policy. Technical report International Growth Centre (IGC). Working paper E-89111-INC-1.
- Becker, Gary S. 1968. "Crime and punishment: An economic approach." In *The economic dimensions of crime*. Springer pp. 13–68.
- Berwick, Elissa, and Fotini Christia. 2018. "State capacity redux: Integrating classical and experimental contributions to an enduring debate." *Annual Review of Political Science* 21: 71–91.
- Blair, Robert A. 2018. "International Intervention and the Rule of Law after Civil War: Evidence from Liberia." *International Organization* pp. 1–34.
- 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.

- Brambor, Thomas, Agustín Goenaga, Johannes Lindvall, and Jan Teorell. 2020. “The lay of the land: information capacity and the modern state.” *Comparative Political Studies* 53 (2): 175–213.
- Breckenridge, Keith. 2014. *Biometric state*. New York: Cambridge University Press.
- CLRC, PNG. 2009. “Review of the law on sorcery and sorcery related killings.” Constitutional and Law Reform Commission, Papua New Guinea.
- Cooper, Jasper. 2018. “State capacity and gender inequality: Experimental evidence from Papua new Guinea.” *Unpublished Manuscript* .
- CSVSR. 2010. Tackling Armed Violence: Key Findings and Recommendations of the Study on the Violent Nature of Crime in South Africa. Technical report. The Centre for the Study of Violence and Reconciliation/Department Safety and Security: Republic of South Africa.  
**URL:** [https://www.csvr.org.za/docs/study/6.TAV\\_final\\_report\\_13\\_03\\_10.pdf](https://www.csvr.org.za/docs/study/6.TAV_final_report_13_03_10.pdf)
- Curry, Michael R, David J Phillips, and Priscilla M Regan. 2004. “Emergency response systems and the creeping legibility of people and places.” *The Information Society* 20 (5): 357–369.
- D’Aoust, Olivia, and Olivier Sterck. 2016. “Who Benefits from Customary Justice? Rent-seeking, Bribery and Criminality in sub-Saharan Africa.” *Journal of African Economies* 25 (3): 439–467.
- Gerber, Alan S, and Donald P Green. 2012. *Field experiments: Design, analysis, and interpretation*. New York: WW Norton.
- Gross, Mark. 2016. “Vigilante violence and ?forward panic? in Johannesburg?s townships.” *Theory and Society* 45 (3): 239–263.
- Hanson, Jonathan K, and Rachel Sigman. 2013. Leviathan’s latent dimensions: measuring state capacity for comparative political research. In *APSA 2011 Annual meeting paper*.
- Hobbes, Thomas. 1958. *Leviathan, reprinted from the edition of 1651; with an essay by the late W.G. Pogson Smith*. Oxford: Clarendon Press.
- Holland, Alisha C. 2016. “Forbearance.” *American Political Science Review* 110 (2): 232–246.
- Human Rights Watch. 2010. “Mob Justice in Burundi. Official Complicity and Impunity.”  
**URL:** <https://www.hrw.org/sites/default/files/reports/burundi0310webwcover.pdf>

- 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.khayelitshacommission.org.za/final-report.html>
- Kirsch, Thomas G, and Tilo Grätz. 2010. *Domesticating vigilantism in Africa*. Boydell & Brewer.
- Lazarev, Egor. 2017. "Choosing among laws: Preferences for alternative legal systems in Chechnya." *Unpublished Manuscript* .
- Lee, Melissa M, and Nan Zhang. 2017. "Legibility and the informational foundations of state capacity." *The Journal of Politics* 79 (1): 118–132.
- Ludwig, Jens, Jeffrey R Kling, and Sendhil Mullainathan. 2011. "Mechanism experiments and policy evaluations." *Journal of Economic Perspectives* 25 (3): 17–38.
- Magaloni, Beatriz, Edgar Franco Vivanco, and Vanessa Melo. 2018. "Killing in the Slums: Social Order, Criminal Governance, and Police Violence in Rio de Janeiro." *Unpublished Manuscript* .
- Mahajan, Sandeep. 2014. *Economics of South African Townships: Special Focus on Diepsloot*. Washington, DC: International Bank for Reconstruction and Development/The World Bank.
- 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.
- Ng'walali, Paul M, and James N Kitinya. 2006. "Mob justice in Tanzania: a medico-social problem." *African health sciences* 6 (1): 36–38.
- Oomen, Barbara. 2005. *Chiefs in South Africa: Law, Power & Culture in the Post-Apartheid Era*. New York: Palgrave.

- Sandefur, Justin, and Bilal Siddiqi. 2011. "Forum Shopping & Legal Pluralism." *Unpublished working paper*.
- SAPS. 2017/2018. Annual Crime Report 2017/2018. Addendum to the SAPS Annual Report. Technical report. South African Police Service.  
**URL:** [https://www.saps.gov.za/services/annual\\_crime\\_report2019.pdf](https://www.saps.gov.za/services/annual_crime_report2019.pdf)
- SAPS. 2018/2019. Crime Statistics: Crime Situation in Republic of South Africa Twelve (12) Months (April to March 2018/2019). Technical report. South African Police Service.  
**URL:** [https://www.saps.gov.za/services/april\\_to\\_march2018\\_19\\_presentation.pdf](https://www.saps.gov.za/services/april_to_march2018_19_presentation.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*. Yale University Press.
- Sekhonyane, Makubetse, and Antoinette Louw. 2002. *Violent Justice: Vigilantism and the State's Response*. Pretoria: Institute for Security Studies.
- Smith, Daniel Jordan. 2004a. "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.
- Smith, Thomas B. 2004b. "Electricity theft: a comparative analysis." *Energy policy* 32 (18): 2067–2076.
- StatsSA. 2016/2017. Victims of Crime Survey. Technical report. Statistics South Africa.  
**URL:** <http://www.statssa.gov.za/?p=10521>
- 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.

Uganda Police. 2013. "Annual Crime and Traffic/Road Safety Report."

Weber, Max. 1946. *From Max Weber: essays in sociology. Translated, edited, and with an introduction, by H. H. Gerth and C. Wright Mills*. New York: Oxford University Press.

Wilson, Woodrow. 1889. *The state. Elements of historical and practical politics. A sketch of institutional history and administration*. Boston, D.C.: Heath & Co.

Xu, Xu. 2020. "To Repress or to Co-opt? Authoritarian Control in the Age of Digital Surveillance." *American Journal of Political Science* .

# Appendices

## A Supplementary Information

### A.1 Pre-registration

Pre-analysis plans (PAPs) and addendums for this study can be found at <https://osf.io/87u4f>. Below, I describe ways in which the analyses diverge from what has been pre-registered. Most divergences arise because of inconsistencies in the analysis procedures that have been pre-specified, respectively, prior to the midline survey and in between the mid- and endline surveys. As a general rule, I follow the endline PAP that was pre-registered in between the mid- and the endline survey.

**Regression Specification.** The midline PAP pre-specified a regression specification that includes block fixed effects. The endline PAP changed the regression specification to one without block fixed effects. Conditioning on block fixed effects is not required for unbiasedness (treatment assignment probabilities do not vary across blocks). Due to the large number of small blocks (50 blocks with 5 units each), the inclusion of block fixed effects leads to a substantial increase in the number of parameters to be estimated. It also runs the risk that entire blocks drop out of the analysis, especially in the presence of attrition and when estimating conditional effects. To harmonize analyses across survey waves and in keeping with what has been pre-registered prior to the endline survey, none of the analyses in this paper condition on block fixed effects. Both PAPs pre-specified two regression specifications, a barebones specification that does not control for covariates and one that controls for a set of covariates selected through a LASSO procedure. I prioritize results based on the barebones specification for transparency but show the robustness of main results to the other specification in section C.2 of the appendix.

**Construction of indices.** Some items that are used to create indices are only available in the midline or only available in the endline survey. All items available in a given survey wave have been used to create the indices as they were specified in the PAP that pertains to the particular survey wave. As a consequence, indices do not always contain the same items across waves. Detailed information on how the indices for each wave were created can be found in section D. The following list details ways in which outcome construction diverges from the midline PAP:

- *Alert Police.* The midline PAP specified that this item would be combined with another variable into an index. The second variable is an indicator that takes the value 1 if 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 the respondent would do if attacked in her home. The estimated treatment effect on this item is substantial. Yet, since this measure conflates availability of the alarm with willingness to rely on police, it was excluded from the endline survey. In keeping with the endline PAP, it is not part of the analysis presented in this paper.
- *Support MV.* This index was only pre-registered at endline. The midline PAP pre-specified that the constituent outcomes of this index would be analyzed separately. Analyses of constituent items are included in the appendix.
- *Service quality.* This index was only pre-registered at endline. The midline PAP pre-specified

that the constituent outcomes of this index would be analyzed separately. Analyses of constituent items are included in the appendix.

All outcome measures are constructed in keeping with the endline PAP. The only exception is the index *Rely Police*. Both the end- and midline PAPs pre-specified that the constituent outcomes of this index (*Alert Police* and *Cooperate Police*) would be analyzed separately. Analyses of constituent items are included in the appendix.

**Direction of hypothesis tests.** The direction of all hypotheses tests follows the endline PAP. This creates the following divergences from the midline PAP:

- *Support MV*. In keeping with the endline PAP, I conduct a one-tailed test (lower). The midline PAP pre-specified a two-tailed test.
- *Call Com*. In keeping with the endline PAP, I conduct a one-tailed test (lower). The midline PAP pre-specified a two-tailed test.
- *Service Quality*. In keeping with the endline PAP, I conduct a one-tailed test (upper). The midline PAP pre-specified a two-tailed test.
- *Join MV*. In keeping with the endline PAP, I conduct a one-tailed test (upper) with regard to 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 that I would analyze treatment effect heterogeneity with regards to prior beliefs about service quality on service quality outcomes (column 6) and with regards to prior beliefs about the risk of punishment on outcomes that pertain to beliefs about this risk (columns 1 and 2). Yet, as explained in the text, when taking into account that the two dimensions of prior beliefs are not independent, it seems intuitive that prior beliefs about one kind of police output may condition treatment effects on beliefs about another kind of police output. While not pre-specified, the analyses in columns 3, 4 and 5 thus arise from the same logic.
- *Table 6*. Analyses in columns 2 and 3 have not been pre-specified, because the PAP focused on the question of whether information treatments shift the beliefs that they are supposed to shift. Analyses in columns 2 and 3 seek to demonstrate that the information treatments did not affect beliefs that they were not intended to affect.
- *Table 7*. Analyses in columns 2 and 4 that subset to those who were not assigned to the respective other information treatment have not been pre-specified. They are included here because they provide a cleaner comparison in the event that the two information treatments interact to effect the willingness to participate in mob vigilantism.

**Registered analyses that are not reported in this paper.** The following analyses have been pre-registered but are not reported here:

- *IV estimation*. The midline PAP also pre-specified an IV estimator. Given the high rate of compliance, this analysis has been omitted from the endline PAP and from the paper.

- *Spillover analyses.* Both PAPs pre-specified that I would analyze spillover effects. The midline PAP specified an analysis that makes use of the spatial layout of households, while the endline PAP specified an analysis that makes use of the sample of neighbors. Given the superiority of the second approach in terms of statistical power, I do not rely on the first approach. I report in the text that I find very little evidence of any kinds of spillover using the second approach. Results are available upon request.
- *Omnibus tests.* The PAPs pre-specified two omnibus tests of the joint significance of two subsets of hypotheses as an additional strategy to shed light on the relevance of the two core mechanisms. It has become clear that these tests are not well suited to discriminate between these mechanisms. One major reason is that, as mentioned in the main text, it is plausible that prior beliefs about one kind of police output may condition treatment effects on both kinds of police outputs under both kinds of mechanisms. Given its limited probative value, this analysis has been omitted here.
- *Behavioral measure.* The endline PAP pre-specified a behavioral measure in the form of respondents' choice between two different kinds of t-shirts that were given to respondents as a thank you gift. The results of this t-shirt experiment do little to strengthen or counter the results presented here and are hence omitted, but are available upon request.
- *Effects of information treatments on demand for policing.* The endline PAP pre-specified hypotheses about the effect of the information treatments on measures of demand for policing that serve to inform a follow-up project but are not directly related to what is presented in this paper.

## A.2 Explanatory note for regression tables

**Effects of alarm treatment.** Unless otherwise indicated, the unit of analysis is always the respondent. Standard errors in all analyses that involve the alarm treatment allow for clustering on the household level so long as the analysis does not collapse the dataset to the household level. As pre-specified, all analyses that involve the alarm treatment control for cluster size, i.e. the number of respondents interviewed in a given household. Unless indicated otherwise, no additional covariates are included. All  $p$ -values are calculated, as pre-specified, using randomization inference by permuting treatment assignment 2,000 times to simulate the sampling distribution of the estimator under the sharp null hypothesis of no (positive/negative) treatment effect. The row labeled “hypothesis” in each table indicates the direction of hypotheses tests which follows the pre-analysis plans (see section A.1 of the appendix for exceptions).

Heterogeneous effects analyses such as those shown in tables 22 and 4 make use of the pre-registered interaction specification where the outcome is regressed on an indicator for treatment assignment, the moderator and the interaction between the two as well as the cluster size control. As pre-specified, all hypothesis tests that involve conditional intent-to-treat (ITT) effects and differences between them are based on randomization inference.  $p$ -values for hypotheses involving conditional ITTs are calculated by sub-setting to the respective group and using the same procedure of permuting treatment assignment 2,000 times described above. Tests of hypotheses involving the difference between conditional ITTs are based on the sharp null hypothesis that the treatment effect for each unit is equal to the estimate of the ITT in the sample as a whole. The testing procedures proceeds as follows:

1. Adjust outcomes in the treatment group based on the assumption that the ITT estimated for the sample as a whole is the true unit-level effect.
2. Permute treatment assignment 2,000 times.
3. Estimate the ITT in each subgroup and the difference between the two ITTs for every permutation.
4. Compare the observed difference in conditional ITT estimates to the simulated sampling distribution to calculate a  $p$ -value.

The resulting  $p$ -values can be slightly different 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).

**Effects of information treatments.** The information treatments were randomized across the entire endline sample ( $N = 815$ ) including 448 respondents from main households and 376 respondents from neighboring households. Analyses in tables 6, 28 and 29 pertain to the information treatments only. These analyses marginalize across the alarm treatment and do not differentiate between respondents from main or neighboring households. All such analyses estimate the effect of one factor of the full factorial design while marginalizing over the other factor. Hence, these analyses rely on pre-specified specifications of the form

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

where  $\mathbf{Y}$  is a vector of outcomes;  $\alpha$  is an intercept;  $\tau$  is the ITT of either the “Police fights crime” or the “Police fights mob vigilantism” prime;  $\mathbf{z}_{info}$  is a vector of indicators of assignments to

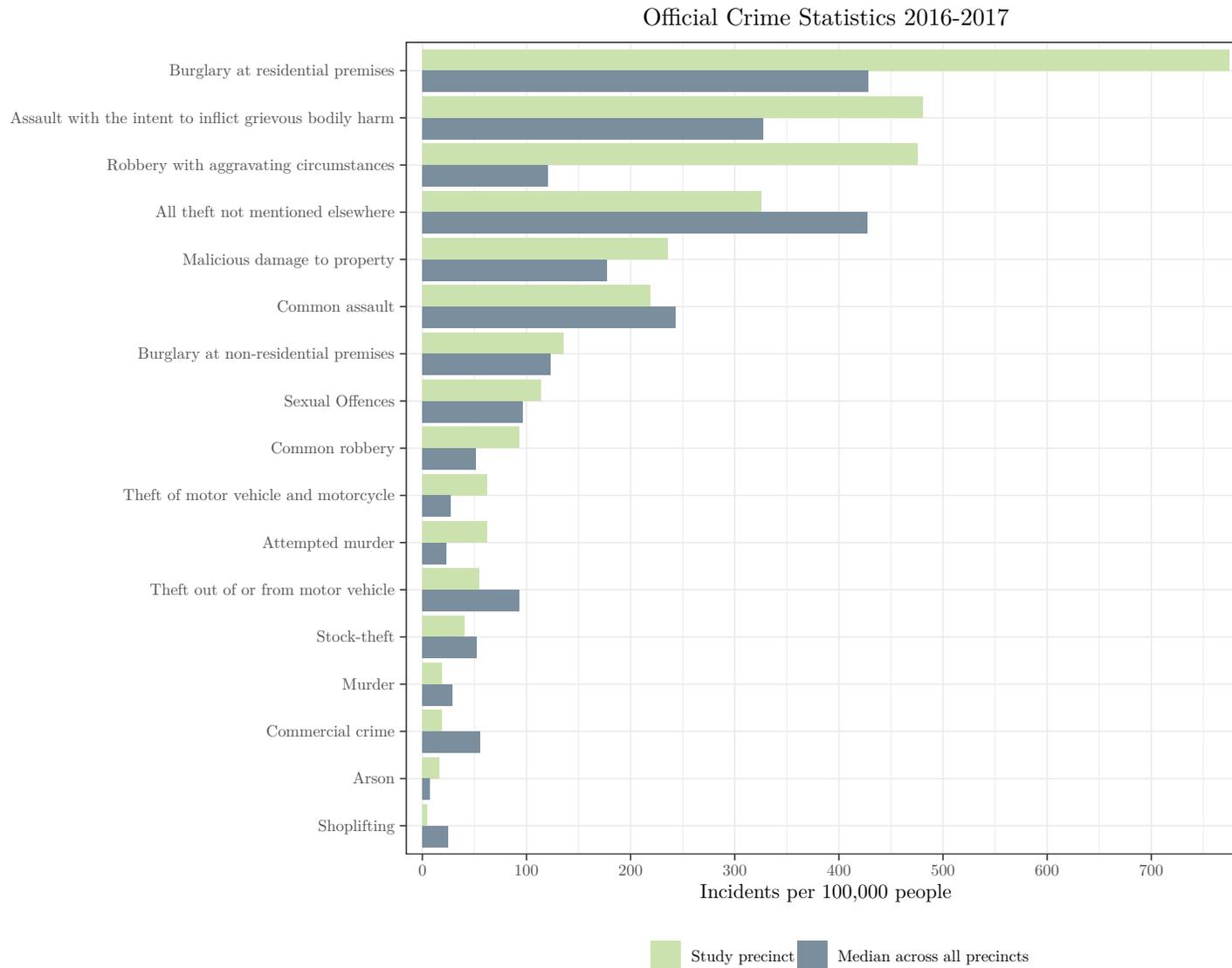
the respective prime; and  $\epsilon$  is a vector of error terms. The unit of analysis is the respondent. Since information treatments were assigned on the respondent and not the household level, this specification does not control for cluster size. For the same reason, standard errors are heteroskedasticity robust but do not allow for any clustering on the household level. Hypothesis tests are based on randomization inference drawing on the same simple random assignment procedure that was used to assign the information treatments in the first place.

**Interaction between alarm and information treatments.** Table 7 shows estimates from the following pre-registered regression specification:

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

where  $\mathbf{Y}$  is a vector of outcomes;  $\alpha$  is an intercept;  $\tau_1$  is the ITT of the alarm treatments among those who were not assigned to the respective information treatment;  $\mathbf{z}_{alarm}$  is a vector of indicators of assignments to the alarm treatment;  $\tau_2$  is the ITT of the respective information treatment among those who were not assigned to the alarm treatment;  $\mathbf{z}_{info}$  is a vector of indicators of assignments to the respective information treatment;  $\tau_3$  is the difference in the effect of the alarm treatment across those who were and were not assigned to the respective information treatment;  $\mathbf{n}$  is a vector of cluster sizes (number of respondents from a given household included in the analysis) and  $\delta$  is the associated coefficient;  $\epsilon$  is 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 treatment or respondents to the information treatment. The  $p$ -value regarding the interaction term 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. This  $p$ -value is calculated using the same procedure as described immediately above, this time permuting both assignment of the alarm and of the information treatments.

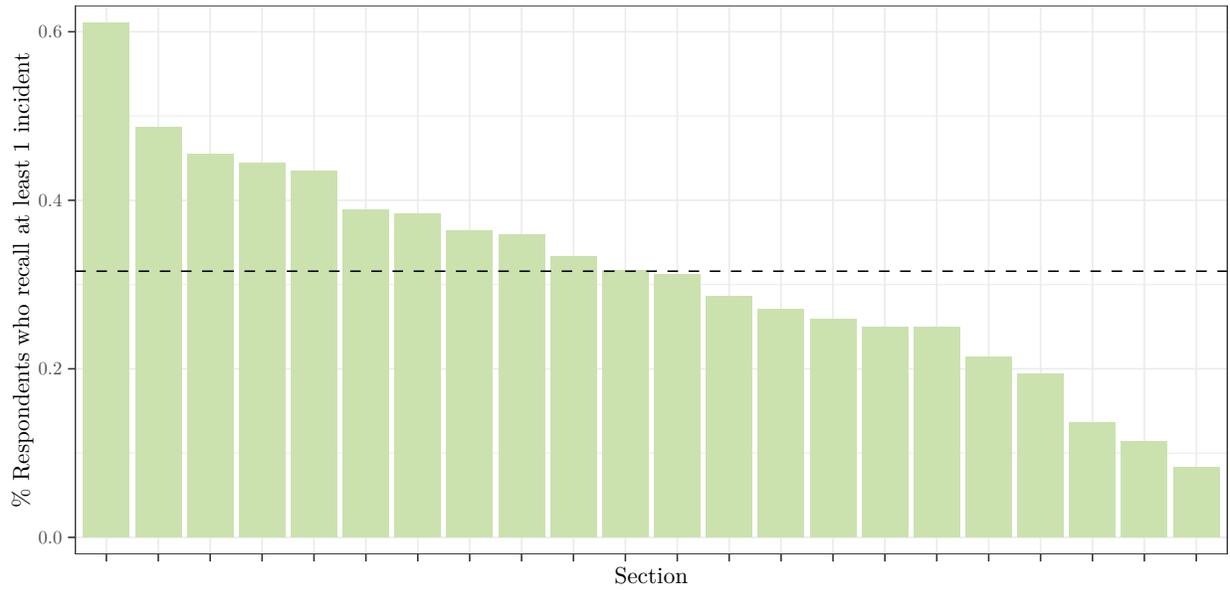
### A.3 Additional plots and figures



**Figure 8:** Official rate of various crimes in study precinct compared to median across all precincts

Crime incident data have been taken from official crime statistics by the SAPS. Crime rates have been calculated using data on population by police district that have been calculated based on the 2011 census and can be found at <https://datahub.io/pi/police-district-population-nometa>. The plot reports data for all crime categories defined as “community reported serious crime” by SAPS (2017/2018).

**Figure 9:** 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 10:** The Police Alarm System

#### **A.4 Sampling strategy for households**

Sampling of households during the baseline survey was based on two different sampling strategies. The original intention was to sample 150 households through each of the strategies.

##### **Police list**

150 households were sampled from a list of households provided by the police station. This is how the implementing partner usually selects alarm recipients. In total, the police provided a pool of 390 names and contact details. The survey team managed to geo-locate 336 of these 390 households. 54 households could either not be found or turned out to be duplicates of houses that had already been located. Among the geo-located households, a stochastic algorithm was used to identify sets of households that satisfy the constraint that every household is located at least 150 m apart from all other households in the set. This distance constraint was imposed to limit spillover effects. Among the sets identified through a large number (at least 1M) of iterations of the algorithm, 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. Due to practical and time constraints, the selection of households started prior to the complete pool of names having been received from the police and before all households had been geo-located. The above procedure was therefore applied repeatedly to subsets of households. For example, in the first round of sample selection, the random walk algorithm was run on the first 92 households that had been geo-located before the start of the survey. From the 92 households, a set of 67 units satisfying the 150 m distance constraint was selected. Subsequently, as additional households were geo-located, the random walk algorithm was

run again to select sets of additional households subject to the constraint that these households would be at least 150 m apart from each other and from the 67 households in the initially selected set.

### **Listing of households in high crime areas**

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

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

### **Household replacement criteria**

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

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

### **Baseline sample**

Additional challenges arose during the implementation of the above described sampling strategy for households: Towards the end of the baseline survey, surveying became impossible in two areas due to opposition from community members who felt suspicious towards the survey team. Households in these areas were replaced with more households from the other high crime areas in the sample. In some cases, the list of names provided by the police was outdated and people had moved. In these cases, the new residents of the address listed on the list received from the police were interviewed. Finally, the existence of three separate house numbering systems and double-numbering (more than one house with the same house number) led to inaccuracies in the geo-coordinates captured during the listing exercise. As a result of these problems, not all households in the sample satisfy the 150 m distance constraint. Additional households were included in the baseline sample to alleviate this problem. In total, we interviewed 358 respondents in the baseline survey, 171 from the police list and 187 that were sampled from high crime areas. 15 respondents refused to be interviewed and 16 respondents could not be found after 3 visits.

## Selection of experimental units

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

- *Exclusion of households not interested in alarm.* In order to avoid non-compliance, respondents were asked during the baseline survey whether they would be interested in receiving an alarm system.<sup>57</sup> 15 households (7 from the police and 8 from the listing in high crime areas) indicated that they would not want to be considered for an alarm. A further 12 households said they were interested during the baseline interview, but had changed their mind by the time they were contacted again during a subsequent round of telephonic back-checks.
- *Exclusion of CPF leaders.* The sample contained 5 of the 10 executive members of the CPF. The implementing partner decided to non-randomly allocate 10 alarm units to these executive members in order to ensure buy-in for the project.
- *Exclusion of households that could not be reached during back-checks.* Given the challenges of locating households and due to concerns about data quality, it was decided to conduct telephonic back-checks of all remaining households that would be considered for randomization. Households that could not be reached via phone were visited in person. Only households that could be reached during the back-check phase and still showed interest in the alarm system were included in the final experimental subject pool.

### A.5 Sampling strategy for neighbors

In the endline survey, one neighboring household was sampled for each of the 250 main household in the sample. Enumerators were instructed to always survey the household to the right of the main household. If there was no neighbor to the right (e.g. because the house was at a corner), enumerators were allowed to replace the house to the right with the house to the left. Within each neighboring household, one randomly selected adult woman and one randomly selected adult man were interviewed. In all-male and all-female households, two men or, respectively two women were interviewed. In single member households, only one respondent was interviewed. A relatively high number of 84 neighboring households turned out to be one-member households. Taking into account that only one respondent could be interviewed, the target sample size was  $N = 416$  neighbors. 84% of those ( $N = 349$ ) could be interviewed. As in main households, additional respondents were interviewed in a given neighboring household if available during the time of the interview. In total, 18 additional neighbors were sampled in this fashion, giving a total sample size of  $N = 367$  neighbors.

---

<sup>57</sup>Specifically, respondents were asked: 'Before I leave, I would like to know whether you may be interested in receiving a home alarm system to protect your household. The system is an electronic device to be installed in your home that allows you to quickly alert the police or your neighbors. The alarm can also make a loud noise to indicate that there is a problem in your home. We are cooperating with a non-profit organization called MeMeZa that gives out these alarms for free. If you are interested in receiving an alarm, we will pass your contact details on to MeMeZa. However it is not certain that you will receive an alarm, since MeMeZa uses a lottery system to give out the alarms. Depending on the availability of alarms, some of those who are interested will receive an alarm relatively soon, some will receive one later and others may not receive one at all. Would you be interested in receiving an alarm system?'

## A.6 Sampling strategy overview

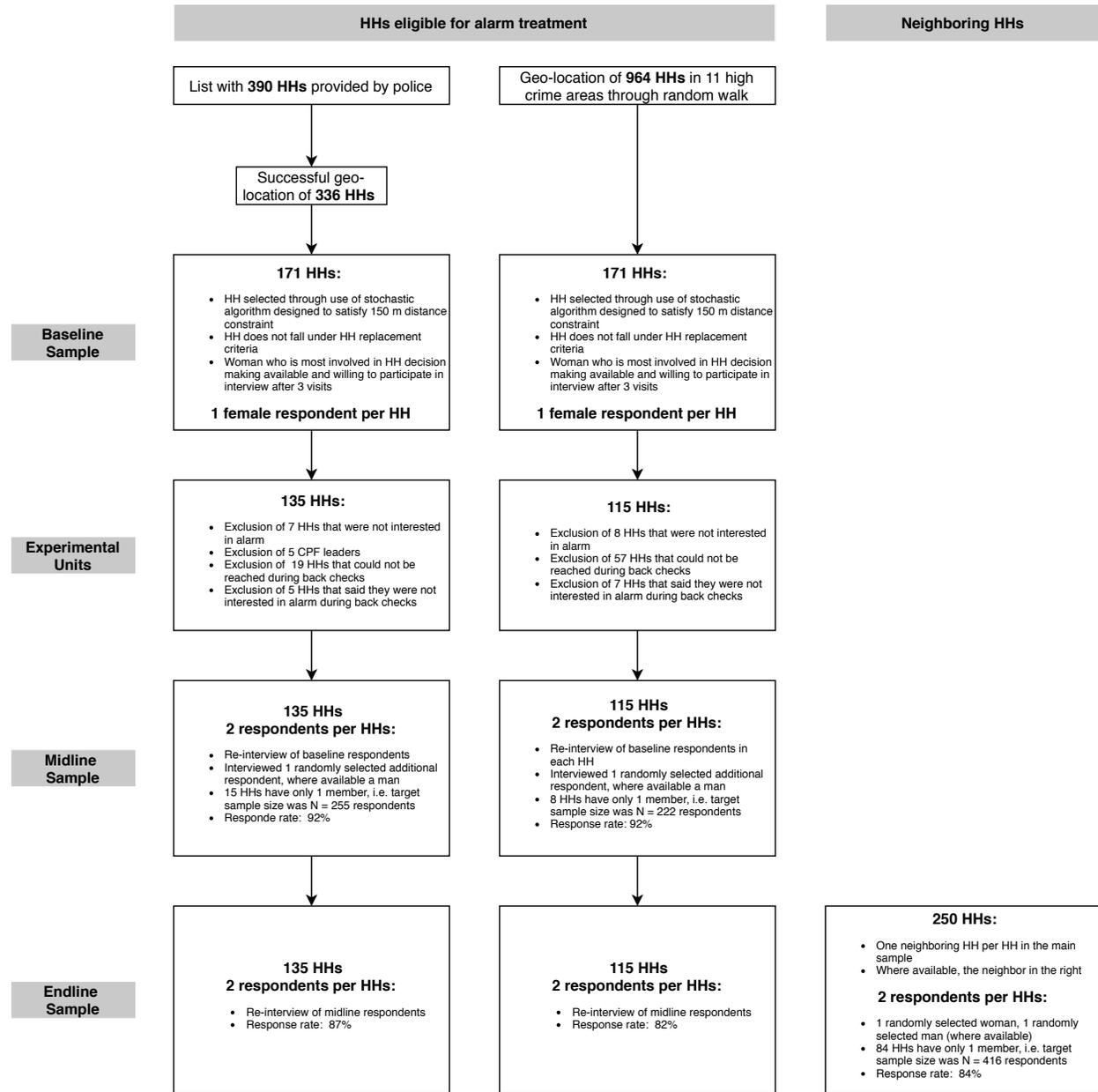


Figure 11: Overview of sampling strategy

## A.7 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
Owens flushing toilet	0.36	0.23
Has tap water in house	0.13	0.07
Owens pay TV	0.59	0.34
Owens electric stove	0.81	0.88
Owens microwave	0.62	0.63
Owens washing machine	0.54	0.45
Owens motor vehicle	0.23	0.20

**Table 8:** Averages of baseline covariates by sampling procedure

	Midline (N = 483)	Endline (N = 448)
Woman	0.63	0.64
Age	42.43	42.88
Married	0.35	0.37
Did not complete secondary education	0.51	0.51
Unemployed	0.37	0.39
Not religious	0.11	0.11

**Table 9:** Sociodemographic characteristics of midline and endline samples

## A.8 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 Siphon 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 Provin-

cial Police Commissioner, Lieutenant General Baile Motswenyane welcomed the hefty sentences. She congratulated the detectives of the Brits polices 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.

Source: <https://kormorant.co.za/41975/rapists-sentenced-13-life-240-years/>

## **Police Fight Mob 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 Commissioners 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.

Source: <https://www.northwestnewspapers.co.za/mafikengmail/community/blogs/editor-s-viewpoint/393-acts-of-vigilantism-a-concern-to-nw-police-commissioner>

## **B Identification**

### **B.1 Covariate balance**

Tables 10 to 14 below examine balance on pre-treatment covariates across experimental conditions among, respectively, baseline, midline and endline respondents. Most covariates are taken from the baseline. Since only one respondent was interviewed at baseline, baseline measurements are treated as household-level measurements when analyzing the midline and endline samples that include two respondents per household. A smaller set of measures which are plausibly unaffected by treatment (e.g. demographic characteristics such as age) are taken from midline and endline surveys.

For each covariate, I test for a significant relationship to treatment using randomization inference. The first column of each table shows the name of the covariate and the following columns show means of the covariate in the treatment and control condition. Treatment and control differences are estimated using a linear model that regressed a given covariate on a treatment assignment indicator. For tables 11 to 14, the model also controls for cluster size (the number of respondents from a given household included in the analysis). The last column in each table shows a two-tailed  $p$ -value that is calculated by comparing the observed coefficient of the treatment assignment indicator to the sampling distribution of the estimator under the sharp null hypothesis of no effect of treatment on the respective covariate for any unit. The sampling distribution under the sharp null hypothesis is simulated by permuting treatment assignment 2,000 times and re-estimating the same

regression model. If tests were independent, we would expect 5% of covariates to show imbalance that is significant at the 5% level.

- Table 10 pertains to the baseline sample. 7/79 (9%) of tests yield a  $p$ -value equal to or less than .05.
- Table 11 pertains to all respondents that were interviewed during the midline survey. 7/103 (7%) of tests yield a  $p$ -value equal to or less than .05.
- Table 12 pertains to the two main respondents sampled in the midline survey, i.e. excluding additional respondents sampled in an ad hoc way. 8/103 (8%) of tests yield a  $p$ -value equal to or less than .05.
- Table 13 pertains to all respondents that were interviewed during the endline survey. 7/103 (7%) of tests yield a  $p$ -value equal to or less than .05.
- Table 14 pertains to the two main respondents sampled in the endline survey, i.e. excluding additional respondents sampled in an ad hoc way. 4/103 (4%) of tests yield a  $p$ -value equal to or less than .05.

	Control	Treatment	$p$ -value
electric_stove_bl	0.80	0.91	0.01
main_income_other_bl	0.19	0.09	0.03
criminals_from_area_bl	0.43	0.30	0.04
prepaid_electricity_bl	0.82	0.90	0.04
electricity_source_other_categories_bl	0.18	0.10	0.04
criminals_from_other_areas_bl	0.21	0.32	0.05
approached_police_bl	0.65	0.52	0.05
number_births_bl	2.77	2.39	0.06
discuss_neighbors_bl	1.71	1.48	0.06
main_income_pensions_bl	0.07	0.15	0.06
water_source_other_categories_bl	0.30	0.20	0.06
rubbish_collection_other_categories_bl	0.20	0.11	0.07
own_refuse_dump_bl	0.80	0.89	0.08
pray_private_bl	7.59	7.74	0.09
microwave_bl	0.59	0.69	0.09
tap_water_in_yard_bl	0.60	0.70	0.10
dishwasher_bl	0.05	0.01	0.15
prisoners_guilty_bl	0.47	0.56	0.16
interest_public_affairs_bl	2.21	2.06	0.20
member_organization_bl	0.78	0.84	0.22
spend_electricity_bl	0.56	0.48	0.22
police_ask_for_bribe_bl	0.89	0.76	0.25
spend_police_1_bl	0.25	0.18	0.25
government_does_enough_bl	0.53	0.60	0.27
experienced_violent_crime_bl	0.11	0.15	0.28
pit_latrine_bl	0.72	0.66	0.28
pay_tv_bl	0.45	0.51	0.29
other_organizations_bl	0.07	0.04	0.40
beat_truck_driver_bl	0.32	0.35	0.44
washing_machine_bl	0.48	0.53	0.44
flush_toilet_tank_bl	0.15	0.18	0.48
mob_violence_police_reaction_bl	1.83	1.74	0.49
trust_neighbor_bl	0.77	0.73	0.50
mob_violence_plausibility_bl	1.65	1.74	0.52
satisfaction_services_bl	0.39	0.37	0.53
motor_vehicle_bl	0.23	0.20	0.53
religious_service_bl	1.38	1.31	0.54
observed_conditions_bl	2.69	2.61	0.58
private_security_bl	0.02	0.01	0.58
discuss_government_bl	2.19	2.28	0.60
flush_toilet_public_bl	0.13	0.15	0.61
guard_dogs_bl	0.22	0.25	0.65
call_police_bl	2.27	2.31	0.66
adequate_force_bl	0.56	0.59	0.69
number_children_bl	1.81	1.89	0.70
response_time_bl	3.11	3.17	0.70
street_committee_connection_bl	0.43	0.39	0.71
able_to_name_bl	1.81	1.84	0.72
shout_community_bl	0.75	0.73	0.72
police_quality_bl	1.59	1.55	0.72
criminals_from_outside_bl	0.36	0.38	0.73
main_income_salary_bl	0.31	0.33	0.73
other_categories_toilet_bl	0.01	0.01	0.74
government_corrupt_bl	0.60	0.62	0.75

blow_whistle_bl	0.14	0.15	0.75
punishment_preferences_bl	0.71	0.73	0.76
report_informal_provider_bl	0.77	0.76	0.76
feel_safe_bl	0.27	0.26	0.78
voice_heard_bl	0.91	0.88	0.79
know_state_official_bl	0.40	0.39	0.80
state_official_bl	0.12	0.13	0.82
discussed_crime_bl	0.90	0.91	0.82
join_mob_bl	0.45	0.44	0.82
number_school_children_bl	1.38	1.41	0.83
government_unresponsive_bl	0.77	0.79	0.84
tap_water_in_house_bl	0.10	0.10	0.84
spend_police_2_bl	0.53	0.54	0.86
number_incidents_bl	0.94	0.91	0.88
attend_meetings_cpf_bl	1.49	1.47	0.92
cpf_connection_bl	1.35	1.33	0.94
know_number_bl	0.81	0.82	0.96
spend_education_bl	0.61	0.62	0.96
hh_size_bl	4.82	4.81	0.97
due_process_bl	0.87	0.87	0.98
courts_punish_not_enough_bl	0.75	0.75	0.98
perceived_crime_risk_bl	1.88	1.88	0.98
other_first_action_bl	0.11	0.12	0.99
main_income_social_grants_bl	0.43	0.43	0.99
attend_meetings_street_committee_bl	0.47	0.46	1.00

Table 10: Balance on covariates among respondents in baseline ( $N = 250$ )

	Control	Treatment	p-value
electric_stove_bl	0.82	0.93	0.01
approached_police_bl	0.67	0.52	0.03
criminals_from_area_bl	0.43	0.30	0.04
spend_police_1_bl	0.26	0.15	0.04
prepaid_electricity_bl	0.83	0.92	0.04
earn_salary	0.56	0.47	0.04
floor_material_missing	0.31	0.21	0.05
number_births_bl	2.84	2.46	0.06
tsonga	0.07	0.14	0.07
microwave_bl	0.61	0.74	0.08
discuss_neighbors_bl	1.68	1.43	0.09
dishwasher_bl	0.05	0.01	0.09
own_refuse_dump_bl	0.81	0.89	0.12
prisoners_guilty_bl	0.47	0.57	0.14
retired	0.13	0.19	0.14
spend_electricity_bl	0.56	0.49	0.16
age	41.67	43.52	0.16
completed_secondary_education	0.37	0.31	0.18
work_full_time	0.18	0.14	0.19
pray_private_bl	7.60	7.73	0.20
member_organization_bl	0.79	0.86	0.20
main_income_pensions_bl	0.09	0.16	0.20
sepedi	0.19	0.13	0.22
other_organizations_bl	0.08	0.04	0.23
observed_conditions_bl	2.78	2.64	0.24
interest_public_affairs_bl	2.18	2.01	0.25
mob_violence_police_reaction_bl	1.87	1.71	0.25
experienced_violent_crime_bl	0.10	0.16	0.26
child_hh_head	0.30	0.27	0.26
no_religion	0.10	0.14	0.27
home_language_sepedi	0.12	0.08	0.28
government_does_enough_bl	0.53	0.60	0.30
tiled_floor	0.25	0.32	0.31
police_ask_for_bribe_bl	0.87	0.75	0.32
pay_tv_bl	0.45	0.53	0.32
tap_water_in_yard_bl	0.63	0.69	0.34
pit_latrine_bl	0.72	0.66	0.34
hh_head	0.38	0.39	0.34
interview_tswana	0.96	0.94	0.34
in_a_relationship	0.17	0.15	0.35
lutheran	0.23	0.28	0.38
others_present	0.21	0.26	0.38
state_official_bl	0.10	0.14	0.39
secondary_education_incomplete	0.39	0.43	0.39
adequate_force_bl	0.54	0.59	0.40
flush_toilet_tank_bl	0.14	0.18	0.40
hh_size_bl	5.08	4.98	0.41
satisfaction_services_bl	0.39	0.36	0.42
mob_violence_plausibility_bl	1.64	1.74	0.42
number_incidents_bl	0.97	0.82	0.47
guard_dogs_bl	0.22	0.27	0.49
unemployed	0.38	0.35	0.49
work_part_time	0.17	0.15	0.50
single	0.31	0.34	0.51
concrete_floor	0.36	0.40	0.51
private_security_bl	0.02	0.01	0.54
police_quality_bl	1.56	1.66	0.56
motor_vehicle_bl	0.24	0.23	0.57

voice_heard_bl	0.91	0.84	0.58
criminals_from_outside_bl	0.35	0.38	0.61
religious_service_bl	1.37	1.31	0.63
know_state_official_bl	0.40	0.37	0.63
flush_toilet_public_bl	0.13	0.16	0.63
main_income_social_grants_bl	0.41	0.42	0.64
discussed_crime_bl	0.89	0.91	0.65
beat_truck_driver_bl	0.31	0.34	0.66
main_income_salary_bl	0.31	0.34	0.66
length_stay	4.07	4.15	0.67
washing_machine_bl	0.51	0.56	0.68
courts_punish_not_enough_bl	0.76	0.73	0.69
discuss_government_bl	2.22	2.31	0.70
spouse_hh_head	0.23	0.24	0.70
report_informal_provider_bl	0.77	0.75	0.71
able_to_name_bl	1.82	1.78	0.72
zcc	0.18	0.19	0.73
due_process_bl	0.86	0.88	0.75
call_police_bl	2.27	2.30	0.75
home_language_tswana	0.70	0.69	0.75
trust_neighbor_bl	0.76	0.75	0.77
attend_meetings_cpf_bl	1.49	1.60	0.78
apostolic	0.21	0.20	0.78
street_committee_connection_bl	0.43	0.39	0.80
spend_education_bl	0.60	0.61	0.80
number_school_children_bl	1.46	1.45	0.81
kind_day	1.55	1.58	0.81
punishment_preferences_bl	0.72	0.74	0.82
feel_safe_bl	0.28	0.28	0.82
tswana	0.51	0.51	0.82
blow_whistle_bl	0.14	0.14	0.84
know_number_bl	0.82	0.84	0.84
response_time_bl	3.08	3.11	0.85
female	0.64	0.62	0.85
cpf_connection_bl	1.33	1.35	0.88
government_corrupt_bl	0.62	0.62	0.90
spend_police_2_bl	0.53	0.54	0.90
perceived_crime_risk_bl	1.87	1.89	0.91
shout_community_bl	0.75	0.74	0.91
married	0.35	0.35	0.91
number_children_bl	1.88	1.94	0.92
attend_meetings_street_committee_bl	0.47	0.45	0.94
join_mob_bl	0.45	0.46	0.96
government_unresponsive_bl	0.78	0.78	0.97
tap_water_in_house_bl	0.11	0.12	0.97

Table 11: Balance on covariates among all respondents in midline ( $N = 483$ )

	Control	Treatment	<i>p</i> -value
electric_stove_bl	0.80	0.93	0.01
approached_police_bl	0.67	0.52	0.03
prepaid_electricity_bl	0.83	0.92	0.03
microwave_bl	0.59	0.72	0.04
earn_salary	0.54	0.44	0.04
criminals_from_area_bl	0.43	0.30	0.05
dishwasher_bl	0.05	0.01	0.05
floor_material_missing	0.31	0.23	0.05
main_income_pensions_bl	0.08	0.15	0.07
number_births_bl	2.81	2.45	0.08
discuss_neighbors_bl	1.69	1.46	0.08
spend_police_1_bl	0.25	0.16	0.08
own_refuse_dump_bl	0.80	0.89	0.08
prisoners_guilty_bl	0.47	0.57	0.11
pray_private_bl	7.59	7.74	0.12
member_organization_bl	0.78	0.86	0.13
interest_public_affairs_bl	2.21	2.03	0.13
spend_electricity_bl	0.57	0.46	0.13
retired	0.14	0.20	0.16
tap_water_in_yard_bl	0.61	0.69	0.17
age	42.55	44.62	0.18
sepedi	0.19	0.13	0.18
tsonga	0.08	0.13	0.18
pay_tv_bl	0.45	0.53	0.19
beat_truck_driver_bl	0.30	0.34	0.21
tiled_floor	0.24	0.31	0.21
mob_violence_police_reaction_bl	1.88	1.72	0.23
in_a_relationship	0.17	0.13	0.25
work_part_time	0.17	0.13	0.25
completed_secondary_education	0.37	0.32	0.26
work_full_time	0.20	0.15	0.26
others_present	0.20	0.25	0.26
home_language_sepedi	0.13	0.08	0.26
length_stay	4.05	4.19	0.27
government_does_enough_bl	0.53	0.60	0.28
police_ask_for_bribe_bl	0.88	0.76	0.28
pit_latrine_bl	0.72	0.66	0.33
lutheran	0.22	0.27	0.33

kind_day	1.55	1.59	0.38
single	0.29	0.33	0.38
mob_violence_plausibility_bl	1.64	1.77	0.39
flush_toilet_public_bl	0.13	0.17	0.39
satisfaction_services_bl	0.39	0.36	0.40
other_organizations_bl	0.07	0.05	0.40
guard_dogs_bl	0.21	0.26	0.41
number_incidents_bl	0.98	0.83	0.42
private_security_bl	0.02	0.01	0.44
observed_conditions_bl	2.74	2.63	0.47
religious_service_bl	1.39	1.31	0.47
experienced_violent_crime_bl	0.11	0.14	0.49
washing_machine_bl	0.50	0.55	0.50
adequate_force_bl	0.56	0.60	0.52
voice_heard_bl	0.92	0.85	0.54
punishment_preferences_bl	0.71	0.75	0.54
child_hh_head	0.27	0.25	0.54
zcc	0.18	0.20	0.55
concrete_floor	0.36	0.40	0.56
street_committee_connection_bl	0.46	0.40	0.57
join_mob_bl	0.44	0.46	0.57
secondary_education_incomplete	0.39	0.41	0.59
number_children_bl	1.85	1.97	0.61
call_police_bl	2.27	2.32	0.61
no_religion	0.11	0.12	0.61
flush_toilet_tank_bl	0.15	0.17	0.64
discussed_crime_bl	0.89	0.91	0.66
due_process_bl	0.85	0.88	0.66
criminals_from_outside_bl	0.35	0.38	0.68
spouse_hh_head	0.25	0.26	0.68
state_official_bl	0.11	0.13	0.69
discuss_government_bl	2.21	2.28	0.69
tswana	0.49	0.51	0.70
interview_tswana	0.96	0.95	0.71
home_language_tswana	0.68	0.70	0.71
know_number_bl	0.82	0.84	0.72
unemployed	0.37	0.36	0.73
tap_water_in_house_bl	0.10	0.11	0.74
attend_meetings_street_committee_bl	0.50	0.45	0.75
perceived_crime_risk_bl	1.86	1.90	0.76
spend_education_bl	0.60	0.62	0.76
main_income_salary_bl	0.31	0.33	0.76
response_time_bl	3.07	3.12	0.77
trust_neighbor_bl	0.77	0.75	0.80
number_school_children_bl	1.43	1.46	0.82
blow_whistle_bl	0.15	0.14	0.82
know_state_official_bl	0.39	0.37	0.83
motor_vehicle_bl	0.23	0.21	0.83
attend_meetings_cpf_bl	1.49	1.52	0.85
spend_police_2_bl	0.54	0.53	0.85
police_quality_bl	1.60	1.63	0.86
female	0.63	0.63	0.86
hh_head	0.41	0.41	0.86
married	0.36	0.37	0.86
apostolic	0.20	0.20	0.86
able_to_name_bl	1.80	1.78	0.88
hh_size_bl	4.90	4.94	0.89
cpf_connection_bl	1.33	1.35	0.89
main_income_social_grants_bl	0.42	0.43	0.90
shout_community_bl	0.74	0.75	0.91
report_informal_provider_bl	0.76	0.77	0.92
government_corrupt_bl	0.61	0.60	0.93
courts_punish_not_enough_bl	0.75	0.75	0.94
government_unresponsive_bl	0.78	0.77	0.95
feel_safe_bl	0.27	0.27	0.98

**Table 12:** Balance on covariates among main respondents in midline ( $N = 438$ )

	Control	Treatment	<i>p</i> -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 13:** Balance on covariates among all respondents in endline ( $N = 448$ )

	Control	Treatment	p-value
electric_stove_bl	0.79	0.93	0.00
microwave_bl	0.56	0.74	0.00
prepaid_electricity_bl	0.82	0.92	0.01
approached_police_bl	0.67	0.53	0.04
floor_material_missing_el_fu	0.30	0.21	0.06
number_births_bl	2.81	2.42	0.06
pray_private_bl	7.57	7.75	0.06
spend_police_1_bl	0.26	0.15	0.07
dishwasher_bl	0.05	0.01	0.07
prisoners_guilty_bl	0.46	0.58	0.08
sepedi_el_fu	0.18	0.11	0.09
own_refuse_dump_bl	0.80	0.89	0.09
earn_salary_el_fu	0.53	0.45	0.11
tap_water_in_yard_bl	0.60	0.70	0.12
retired_el_fu	0.13	0.20	0.13
main_income_pensions_bl	0.08	0.14	0.13
criminals_from_area_bl	0.40	0.30	0.14
interest_public_affairs_bl	2.21	2.02	0.16
work_full_time_el_fu	0.20	0.14	0.17
length_stay_el_fu	4.04	4.21	0.17
discuss_neighbors_bl	1.68	1.49	0.17
age_el_fu	42.26	44.20	0.20
washing_machine_bl	0.48	0.57	0.20
government_does_enough_bl	0.53	0.62	0.21
experienced_violent_crime_bl	0.09	0.15	0.21
member_organization_bl	0.79	0.86	0.23
spend_electricity_bl	0.56	0.48	0.23
pay_tv_bl	0.45	0.53	0.25
kind_day_el_fu	1.56	1.63	0.26
tsonga_el_fu	0.09	0.14	0.26
mob_violence_police_reaction_bl	1.88	1.75	0.28
tiled_floor_el_fu	0.24	0.30	0.29
lutheran_el_fu	0.23	0.28	0.32
others_present_el_fu	0.19	0.24	0.32
home_language_sepedi_el_fu	0.11	0.07	0.32
pit_latrine_bl	0.73	0.68	0.32
no_religion_el_fu	0.10	0.14	0.34
completed_secondary_education_el_fu	0.37	0.33	0.35
religious_service_bl	1.38	1.28	0.36
punishment_preferences_bl	0.70	0.76	0.36
satisfaction_services_bl	0.38	0.35	0.38
know_number_bl	0.80	0.86	0.38
police_ask_for_bribe_bl	0.85	0.73	0.38
other_organizations_bl	0.08	0.05	0.38
work_part_time_el_fu	0.16	0.13	0.39
discuss_government_bl	2.16	2.29	0.40
flush_toilet_tank_bl	0.13	0.17	0.43
single_el_fu	0.30	0.33	0.44
main_income_salary_bl	0.29	0.33	0.44
state_official_bl	0.11	0.14	0.45
interview_tswana_el_fu	0.96	0.95	0.46
in_a_relationship_el_fu	0.18	0.15	0.47
tswana_el_fu	0.48	0.51	0.50
due_process_bl	0.85	0.88	0.51
private_security_bl	0.02	0.01	0.51
motor_vehicle_bl	0.24	0.21	0.53
courts_punish_not_enough_bl	0.73	0.76	0.55
flush_toilet_public_bl	0.12	0.15	0.55
secondary_education_incomplete_el_fu	0.42	0.39	0.56
mob_violence_plausibility_bl	1.67	1.75	0.58
guard_dogs_bl	0.22	0.25	0.59
attend_meetings_cpf_bl	1.50	1.62	0.60
beat_truck_driver_bl	0.30	0.31	0.60
spend_education_bl	0.59	0.62	0.60
discussed_crime_bl	0.89	0.91	0.63
shout_community_bl	0.73	0.76	0.66
street_committee_connection_bl	0.45	0.40	0.66
zcc_el_fu	0.17	0.19	0.67
unemployed_el_fu	0.39	0.38	0.67
call_police_bl	2.29	2.34	0.67
number_incidents_bl	0.93	0.85	0.67
cpf_connection_bl	1.35	1.41	0.67
hh_head_el_fu	0.40	0.41	0.69
observed_conditions_bl	2.72	2.66	0.69
main_income_social_grants_bl	0.46	0.44	0.69
spouse_hh_head_el_fu	0.25	0.24	0.70
concrete_floor_el_fu	0.38	0.40	0.72
join_mob_bl	0.42	0.44	0.72
voice_heard_bl	0.91	0.88	0.73
spend_police_2_bl	0.52	0.54	0.73
trust_neighbor_bl	0.78	0.77	0.77
government_corrupt_bl	0.60	0.62	0.77
police_quality_bl	1.59	1.56	0.77
home_language_tswana_el_fu	0.68	0.69	0.78
blow_whistle_bl	0.15	0.13	0.78
child_hh_head_el_fu	0.26	0.25	0.80
feel_safe_bl	0.28	0.28	0.81
able_to_name_bl	1.83	1.80	0.82
number_school_children_bl	1.45	1.46	0.83
report_informal_provider_bl	0.77	0.76	0.84

response_time_bl	3.10	3.11	0.86
tap_water_in_house_bl	0.11	0.12	0.87
apostolic_el_fu	0.20	0.21	0.90
attend_meetings_street_committee_bl	0.48	0.46	0.90
married_el_fu	0.35	0.37	0.93
government_unresponsive_bl	0.77	0.76	0.94
hh_size_bl	4.90	4.99	0.95
number_children_bl	1.90	1.93	0.95
criminals_from_outside_bl	0.38	0.38	0.95
perceived_crime_risk_bl	1.88	1.88	0.95
female_el_fu	0.64	0.63	0.96
know_state_official_bl	0.38	0.37	0.96
adequate_force_bl	0.57	0.58	0.96

**Table 14:** Balance on covariates among main respondents in endline ( $N = 409$ )

## 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 15:** Reported household size and rates of attrition across experimental conditions

The outcome in row 1 is an indicator for whether enumerators reported that there is only one person living in a particular household. The unit of analysis is the household. The outcomes in row 2 and 3 are, respectively, an indicator for whether a respondent attrited in the midline or endline survey. The unit of analysis is the respondent. Rows 2 and 3 are based on a “completed” dataset, which assumes 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 stem from an unequal variance  $t$ -test that was 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 16:**  $F$ -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 (0.048)	0.066 (0.049)	0.068 (0.059)	0.077 (0.051)
Control Mean	0.136	0.134	0.156	0.134
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 <sup>2</sup>	0.002	0.004	0.001	0.005

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

**Table 17:** Additional respondents sampled across experimental conditions

The unit of analysis is the household for all analyses. 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 in the respective survey wave. The outcome in columns 2 and 4 is the number of additional respondents interviewed. Outcomes are regressed on an indicator for treatment assignment and  $p$ -values are calculated using randomization inference.

## C Additional Analyses

### C.1 Additional outcomes

	Spoken to police			
	Midline (1)	Endline (2)	Endline (3)	Endline (4)
Alarm	-0.017 (0.035)	0.041 (0.050)	0.042 (0.071)	0.054 (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
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	-	-	-	-
Number HFs	245	237	237	237
Observations	483	448	448	448

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

**Table 18:** Effects of the alarm treatment on whether respondents have recently spoken to police

Section A.2 of the appendix contains more details on model specification. See section D.7 for question wording and coding of the outcome.

	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	-0.041*	-0.031	-0.003	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	0.58
RI p-value Main	0.085	0.23	0.446	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&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

**Table 19:** Effects of the alarm treatment on respondents' support for mob vigilantism and willingness to call the community.

All outcome measures range from 0 to 1. The results in columns 5 to 8 are based on specifications that 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 5 and 6) 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 7 and 8) 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.6 for 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 and testing. See section D.8 for question wording and coding of outcomes.

## C.2 Robustness

### Excluding additional respondents

	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*** (0.030)	0.059** (0.033)	-0.061** (0.032)	-0.012 (0.030)	0.104** (0.045)	-0.101** (0.045)	0.109*** (0.043)	-0.043 (0.039)
Alarm × High Prior Punishment					-0.085* (0.065)	0.161*** (0.060)		
Alarm × High Prior Service							-0.106* (0.064)	0.066 (0.061)
Control Mean	0.61	0.65	0.22	0.18	0.65	0.18	0.65	0.18
RI <i>p</i> -value Main	0	0.032	0.026	0.346	0.012	0.012	0.004	0.146
Hypothesis Main	upr	upr	lwr	lwr	upr	lwr	upr	lwr
RI <i>p</i> -value Diff.	-	-	-	-	0.066	0.002	0.059	0.206
Hypothesis Diff	-	-	-	-	lwr	upr	lwr	upr
Number HHs	243	233	243	233	233	233	233	233
Observations	438	409	438	409	409	409	409	409

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

**Table 20:** Effects of alarm treatment on respondents' willingness to rely on police and participate in mob vigilantism among initial respondents.

All analyses are run on samples that are subset to the two initially sampled respondents. All outcome measures range from 0 to 1. The results in columns 5 to 8 are based on specifications that 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 5 and 6) 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 7 and 8) 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.6 for 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 and testing. See section D.1 for question wording and coding of outcomes.

## 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	0.067*** (0.026)	0.048** (0.029)	-0.071** (0.029)	-0.005 (0.026)	0.070* (0.042)	-0.070 (0.038)	0.093*** (0.038)	-0.026 (0.036)
Alarm × High Prior Punishment					-0.047 (0.056)	0.124** (0.053)		
Alarm × High Prior Service							-0.105** (0.055)	0.048 (0.054)
Control Mean	0.6	0.64	0.24	0.17	0.64	0.17	0.64	0.17
RI p-value Main	0.005	0.035	0.011	0.438	0.054	0.104	0.007	0.289
Hypothesis Main	upr	upr	lwr	lwr	upr	lwr	upr	lwr
RI <i>p</i> -value Diff.	-	-	-	-	0.203	0.021	0.034	0.144
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 21:** Effects of alarm treatment on respondents’ willingness to rely on police and participate in mob vigilantism estimated with covariate adjustment.

All outcome measures range from 0 to 1. The results in columns 5 to 8 are based on specifications that 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. In addition, all specifications control for a set of covariates selected through a pre-specified LASSO regression procedure. 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 5 and 6) 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 7 and 8) 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.6 for 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 and testing. See section D.1 for question wording and coding of outcomes.

## Listwise deletion

	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	0.096*** (0.028)	0.076*** (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 × High Prior Punishment					-0.106** (0.062)	0.158*** (0.057)		
Alarm × High Prior Service							-0.101* (0.061)	0.066 (0.057)
Control Mean		0.64	0.24	0.17	0.64	0.17	0.64	0.17
RI p-value Main	0	0.002	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.	-	-	-	-	0.037	0.002	0.066	0.206
Hypothesis Diff	-	-	-	-	lwr	upr	lwr	upr
Number HHs	245	237	245	237	237	237	237	237
Observations	476	447	483	448	447	448	447	448

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

**Table 22:** Effects of alarm treatment on respondents' willingness to rely on police and participate in mob vigilantism without imputations.

All outcome measures range from 0 to 1. The results in columns 5 to 8 are based on specifications that 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 5 and 6) 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 7 and 8) 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.6 for 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 and testing. See section D.1 for question wording and coding of outcomes.

### 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*** (0.038)	0.079** (0.037)	0.091*** (0.030)	0.070** (0.035)	0.140*** (0.051)	0.124** (0.054)	0.132*** (0.051)	0.120*** (0.047)
Alarm × High Prior Punishment					-0.114* (0.074)	-0.103* (0.071)		
Alarm × High Prior Service							-0.110* (0.072)	-0.098 (0.071)
Control Mean	0.65	0.7	0.55	0.58	0.7	0.58	0.7	0.58
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 23:** Effects of the alarm treatment on the willingness to alert and to cooperate with police

All outcome measures range from 0 to 1. The results in columns 5 to 8 are based on specifications that 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 5 and 6) 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 7 and 8) 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.6 for 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 and testing. See section D.1 for question wording and coding of outcomes.

Support MV Index (ML)					
	Not arrest mob	Beat known thief	Beat petty thief	Beat driver	Comm. deal crime
Alarm	-0.063 (0.048)	-0.033 (0.044)	-0.028 (0.039)	-0.042 (0.035)	-0.039 (0.045)
Control Mean	0.46	0.35	0.22	0.18	0.3
RI p-value	0.193	0.438	0.464	0.234	0.358
Hypothesis	two	two	two	two	two
Number HFs	245	245	245	245	245
Observations	483	483	483	483	483

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

**Table 24:** The effect of the alarm treatment on individual items used to create the index “Support MV” at midline. The unit of analysis is the respondent in all analyses. Section A.2 of the appendix contains details on model specification. See section D.8 for question wording and coding of outcomes.

	Join MV Index (EL)		Support MV Index (EL)	
	Join beating	Join mob	Not arrest mob	Beat known thief
Alarm	-0.039 (0.035)	0.016 (0.033)	0.049 (0.049)	-0.013 (0.048)
Control Mean	0.21	0.14	0.58	0.33
RI p-value	0.136	0.684	0.857	0.39
Hypothesis	lwr	lwr	lwr	lwr
Number HFs	237	237	237	237
Observations	448	448	448	448

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

**Table 25:** The effect of the alarm treatment on individual items used to create the indices “Join MV” and “Support MV” at endline. The unit of analysis is the respondent in all analyses. Section A.2 of the appendix contains details on model specification. See sections D.1 and D.8 for question wording and coding of outcomes.

	Service Quality Index ML					
	Take problem seriously	Appear competent	Send guilty to prison	Make effort	Not pay to escape	Care about community
	(1)	(2)	(3)	(4)	(5)	(6)
Alarm	0.057* (0.031)	0.055* (0.031)	0.097** (0.045)	0.069 (0.047)	0.051* (0.028)	0.042 (0.028)
Control Mean	0.6	0.68	0.35	0.42	0.25	0.44
RI <i>p</i> -value	0.072	0.099	0.042	0.144	0.052	0.399
Hypothesis	two	two	two	two	two	two
Number HHs	245	245	245	245	245	245
Observations	483	483	483	483	483	483

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

**Table 26:** The effect of the alarm treatment on individual items used to create the index “Service quality” at midline. The unit of analysis is the respondent in all analyses. Section A.2 of the appendix contains details on model specification. See section D.2 for question wording and coding of outcomes.

79

	Service Quality Index EL		
	Take problem seriously	Send guilty to prison	Make effort
	(1)	(2)	(3)
Alarm	-0.030 (0.029)	0.040 (0.048)	0.114** (0.048)
Control Mean	0.74	0.46	0.47
RI <i>p</i> -value	0.846	0.208	0.014
Hypothesis	upr	upr	upr
Number HHs	237	237	237
Observations	448	448	448

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

**Table 27:** The effect of the alarm treatment on individual items used to create the index “Service quality” at endline. The unit of analysis is the respondent in all analyses. Section A.2 of the appendix contains details on model specification. See section D.2 for question wording and coding of outcomes.

#### C.4 Additional results information treatments

	Believes police fight crime	Believes police fight MV
Police fight Crime	0.020 (0.035)	
Police fight MV		-0.009 (0.028)
Control Mean	0.46	0.65
RI p-value	0.263	0.63
Hypothesis	upr	upr
Observations	815	815

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

**Table 28:** Effect of information treatments among all endline respondents

Analyses are run among all endline respondents (from main *and* neighboring households). Column 1 regresses the outcome on an indicator for assignment to ‘Police fight crime’ treatment. Column 2 regressed the outcome on an indicator for assignment to the ‘Police fight MV’ treatment. See section A.2 of the appendix for information on model specification and section D.4 of the appendix for outcome question wording and coding.

	Believes police fight crime		Believes police fight MV		Would participate MV	
Police fight Crime	0.068** (0.040)		0.009 (0.039)		0.005 (0.035)	
Police fight MV		-0.024 (0.040)		0.027 (0.038)		-0.023 (0.035)
Control Mean	0.2	0.25	0.56	0.56	0.35	0.36
RI p-value	0.046	0.724	0.402	0.238	0.549	0.262
Hypothesis	upr	upr	upr	upr	lwr	lwr
Observations	451	451	451	451	451	451

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

**Table 29:** Effect of information treatments among respondents with low priors about police service

Analyses are run on a subset of all respondents (from main *and* neighboring households) that has low priors about the quality of police service as measured during the endline survey, prior to the administration of the information treatments. See section D.6 for the endline item used to measure prior beliefs. Columns 1, 3 and 5 regress the outcome on an indicator for assignment to ‘Police fight crime’ treatment. Columns 2, 4 and 6 regress the outcome on an indicator for assignment to the ‘Police fight MV’ treatment. See section A.2 of the appendix for information on model specification and section D.4 of the appendix for outcome question wording and coding.

## C.5 Ruling out alternative explanations

### C.5.1 Social desirability bias

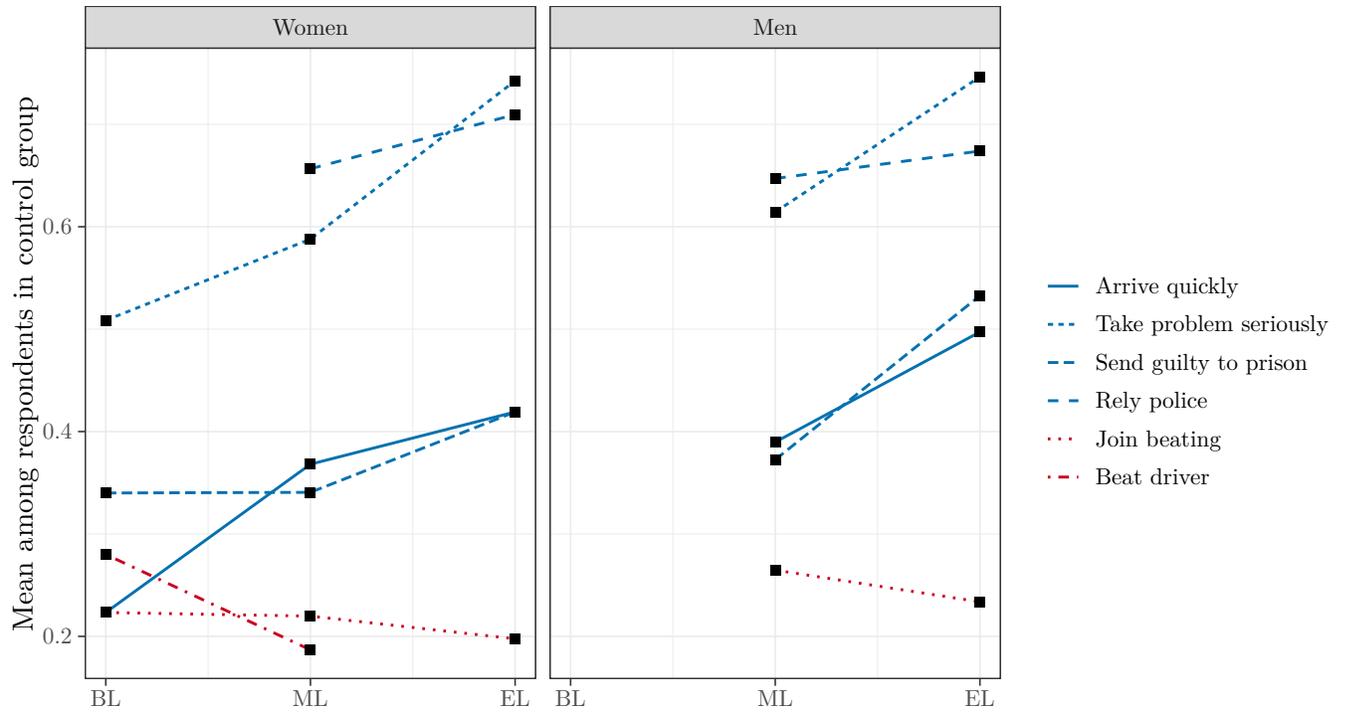
	All endline respondents			Low priors legal repercussions MV		
	Any MV incidents (1)	Number MV incidents (2)	Witnessed any (3)	Any MV incidents (4)	Number MV incidents (5)	Witnessed any (6)
Alarm	0.042 (0.050)	0.285 (0.175)	0.053 (0.046)	0.048 (0.075)	0.342 (0.225)	0.040 (0.066)
Control Mean	0.31	0.75	0.22	0.33	0.67	0.22
RI p-value	0.802	0.96	0.884	0.74	0.94	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 30:** 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.5.2 Changes among control group



**Figure 12:** Change in outcomes in control group across survey waves by gender

Only women respondents were interviewed in the baseline. Outcomes in blue relate to the police; outcomes in red relate to mob vigilantism. BL stands for baseline, ML stands for midline and EL stands for endline. See sections [D.1](#), [D.2](#), and [D.8](#) for question wording.

### C.5.3 Change in punishment preferences due to improved safety

	Feel Safe	HH experienced crime	Punish more	Quick justice
	(1)	(2)	(3)	(4)
Alarm	0.099*** (0.028)	-0.018 (0.049)	-0.009 (0.047)	-0.020 (0.050)
Control Mean	0.59	0.18	0.46	0.51
RI p-value	0	0.368	0.86	0.684
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 31:** Effect of the alarm treatment on safety and punishment preferences

All analyses draw on data from the endline survey. The question wording of outcome measures is as follows. *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 youve 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 dont 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

This section provides the question wording and coding of all outcome measures except for those that are explained in table captions, broken down by tables. The responses “don’t know” and “refuse to answer” have been coded as missing and imputed using multivariate imputation via chained equations for all measurements. Indices have been created by averaging across items.

### D.1 Table 22

*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

*Column 1: 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 3: 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 dont 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

*Column 4: Service Quality Midline.* This outcome measure is an index of the following measures:

- *Customer service:* This measure is a sub-index of two 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

- *Appear competent:* When you or someone like you takes a problem to the police, how likely is it that the police appear to know what they are doing?

- \* 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 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.
- *Make effort:* Which of the following statements comes closer 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.
- *No collusion with criminals:* This measure is a sub-index of two items:
  - *Not pay to escape:* Suppose the police catches a criminal. Do you think it is likely that the criminal could escape punishment by offering money to the police officers? If yes: Do you think it is just likely or very likely? If no: Do you think it is unlikely or very unlikely?
    - \* 0 = Very likely
    - \* 0.25 = Likely
    - \* 0.5 = Unlikely
    - \* 1 = Very unlikely
  - *Care about community:* Which of the following statements is closest to your view:
    - \* 1 = Statement 1: The police have the community's best interest at heart.
    - \* 0 = Statement 2: The police do not care about the community and are colluding with the criminals.

*Column 5: Service Quality Endline.* 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.
- *Make effort:* Which of the following statements comes closer 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.

*Column 6: 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

*Column 7: 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 while it is happening and arrive while the community members are still beating the criminal.

*Column 8: 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.

### D.3 Table 4

Outcomes in columns 1, 2, 5 and 6 are the same as those in table 3 and have been described in section D.2.

- *Columns 3 and 6: Service index.* In line with the pre-analysis plan, this is the same index as the one called *Service Quality Endline* in column 6 of table 3 and described in section D.2. The only difference is that the item *Arrive quickly* was added to the index (see explanation of column 4 of table 3 in section D.2.) The rationale behind this pre-registered change was to limit the number of outcomes in the subgroup analyses.

### D.4 Tables 6 and 29

*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 Table 7

The outcome in all columns is the same as that in column 5 and 6 in table 6 described in section D.4.

## D.6 Measures of prior beliefs

### Alarm Treatment

Tables 22 and 4 in the main text provide evidence of how the effect of the alarm treatment varies across subgroups defined by prior beliefs about, respectively, the risk of facing state punishment for involvement in mob vigilantism ( $q$ ) and the expected quality of services provided by police ( $p_s$ ).

Measures of prior beliefs are taken from the baseline survey. Since only one woman respondent was interviewed during the baseline survey, baseline measures of prior beliefs are treated as household level measures. For example, if the baseline respondent in a given household thought it unlikely that those who participate in vigilante violence will be arrested, then all respondents interviewed in this household at midline and endline will be treated as having had a low prior belief about  $q$  in all analyses of effect heterogeneity.

*Prior beliefs about legal repercussions for MV ( $q$ ):*

- 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  $q$  (1 =) “Very likely” or “Somewhat likely”
  - Low prior  $q$  (0 =) “Not very likely” or “Not likely at all”

*Prior beliefs about police service quality ( $p_s$ ):* 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  $p_s$  (1 =) respondent's score falls strictly above baseline sample median of index
  - Low prior  $p_s$  (0 =) respondent's score falls below baseline sample median of index

### Information Treatments

Analyses that draw on information treatments only (see tables 6 and 29) use questions that were asked during the endline survey prior to the administration of the information treatments to measure prior beliefs about the risk of legal repercussions for mob vigilantism and police service quality. The advantage is that these measures, contrast to the above described baseline measures of priors are available for every respondent and not only for the one respondent per household that has been interviewed during the baseline. Moreover, these are the only measures of priors that are also available for neighbors. Since these measures are post-treatment when it comes to the alarm treatment, I do not rely on them when analyzing the alarm treatment and the information treatments together.

*Endline beliefs (prior to information treatment) about legal repercussions for MV (q):*

- 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 ( $p_s$ ):*

- 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.7 Table 18

*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.8 Table 19

*Column 1,2,5,7: Support MV* This measure is 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.