Online Appendix to: "How the State Discourages Vigilantism - Evidence from a Field Experiment in South Africa"

Contents

A Additional Information

A.1 Pre-registration

Pre-analysis plans (PAPs) can be found at [REDACTED]. Here, I describe divergences from the PAPs. Most divergences arise because of inconsistencies between the midline PAP (registered prior to the midline) and the endline PAP (registered in between the mid- and endline survey). Generally, I follow the endline PAP.

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

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

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

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

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

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

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

Registered analyses not reported in this paper:

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

A.2 Explanatory note for regression tables

Alarm treatment. Unless otherwise indicated, the unit of analysis is the respondent. Standard errors allow for clustering on the household level unless the dataset is collapsed to the household level. As pre-specified, I control for cluster size, i.e. the number of respondents interviewed per household. Unless indicated otherwise, no additional covariates are included. p-values are calculated, as pre-specified, using randomization inference by permuting treatment assignment 2,000 times to simulate the sampling distribution under the sharp null hypothesis of no (positive/negative) treatment effect for any unit. The row labeled "hypothesis" in each table indicates the direction of hypotheses tests. Heterogeneous effects analyses make use of the pre-registered interaction specification, regressing the outcome on an indicator for treatment assignment, the moderator and the interaction between the two as well as the cluster size control. Randomization inference p-values for hypotheses involving conditional intent-to-treat effects (ITTs) are calculated by sub-setting to the respective group and using the same procedure of permuting treatment assignment 2,000 times. Tests of hypotheses involving the difference between conditional ITTs pertain to the sharp null hypothesis that the treatment effect for each unit is equal to the estimated ITT in the sample as a whole. The testing procedure is as follows: First, I adjust outcomes in the treatment group as if the estimated ITT for the sample as a whole were the true unit-level effect. Second, I permute treatment assignment 2,000 times. Third, I

estimate the ITT in each subgroup and the difference between the two ITTs for every permutation. Finally, I compare the observed difference in conditional ITT estimates to the simulated sampling distribution to calculate a p-value. The resulting p-values can differ from parametric p-values based on clustered standard errors but the differences tend to be minor and can go in either direction (larger/smaller p- values).

Information treatments. Information treatments were randomized across the entire endline sample including 448 respondents from main and 376 from neighboring households. Analyses in Tables 5, A18 and A19 pertain to information treatments only, marginalize across the alarm treatment, include respondents from main and neighboring households, and estimate the effect of one factor of the full factorial design while marginalizing over the other. The pre-specified regression is

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

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

Alarm and information treatments. Table 6 relies on the following pre-registered specification:

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

where $\boldsymbol{Y}$ is a vector of outcomes; α an intercept; τ_1 the ITT of the alarm treatment among respondents not assigned to the respective information treatment; $\boldsymbol{z}_{alarm}$ is a vector of indicators of assignment to the alarm; τ_2 is the ITT of the information treatment among those who were not assigned to the alarm; $\boldsymbol{z}_{info}$ is a vector of indicators of assignments to the information treatment; τ_3 is the difference in the effect of the alarm across those who were and were not assigned to the information treatment; $\boldsymbol{n}$ is a vector of cluster sizes and δ the associated coefficient; and $\boldsymbol{\epsilon}$ a vector of error terms that allows for clustering at the household level. p-values that pertain to hypotheses about τ_1 and τ_2 are calculated by sub-setting to the respective group and using randomization inference based on the random assignment function that was used, respectively, to assign households to the alarm or respondents to the information treatment. The p-value for τ_3 pertains to the sharp null hypothesis that the treatment effect for each unit is equal to the estimated ITT in the sample as a whole and is calculated using the procedure described above, this time permuting both assignment of the alarm and of the information treatments.

A.3 Formal framework

I here formalize my argument about how police capacity may affect a citizen's choice between the police and vigilantism. Let p_s and p_v be the probabilities that turning to, respectively, the police or vigilantism would result in punishment of the perpetrator, x_v the utility that the citizen would derive from vigilante and x_s her utility from state punishment. Finally, let q be the probability that the citizen would be punished by the state for resorting to vigilantism and y her resulting utility loss. Police capacity, θ, can be thought of as an input into the production functions for p_s and q, both of which are outputs of police activity. E.g., the more officers a police chief has at her disposal, the more effectively she will be able to respond to both non-vigilante and vigilante crimes. There likely exist other inputs specific to each police output. E.g., whether vigilantes are arrested will also depend on how much effort police dedicate to vigilantism as opposed to other crimes. For concreteness, let

e_s be the effort that police would make to convict the perpetrator whom the citizen reports and e_v the effort police would make to convict the citizen if she resorts to vigilantism. We can then write the two outputs as $q(e_v, \theta)$ and $p_s(e_s, \theta)$, where both functions are assumed to increase in their arguments. I abstract here from other inputs like the functioning of the court system that are unlikely to be affected by the treatment under study. Taken together, a citizen will resort to vigilantism if the expected utility of community punishment net of the expected legal costs exceeds the expected utility of reporting the crime to police:

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

Hence, a shock to police capacity may affect the citizen's behavior through a change in both, the expected costs of vigilantism and the expected utility of reliance on the police.

A priori pessimistic citizens should be most affected. Citizens are likely uncertain about police capacity and what matters for citizens' decision-making is the perceived rather than de facto capacity of police. To the degree that citizens are heterogeneous in terms of their prior views about the police, a positive capacity shock like the alarm treatment should have a greater effect on citizens who are a priori pessimistic. Suppose the citizen's prior belief about θ is given by $\theta \sim \mathcal{N}(m, \sigma_\theta^2)$, where m is the prior mean and σ_θ^2 the prior variance. Furthermore, presume a citizen who is subject to the capacity shock receives an imperfect signal $s \sim \mathcal{N}(\theta, \sigma_\epsilon^2)$ of police capacity that is drawn from a normal likelihood with variance σ_ϵ^2 and centered on the true parameter θ. It is straightforward to calculate the effect of the capacity shock on the citizen's belief about θ:

$$\tau_\theta = m' - m = (s - m) \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2},$$

where m' is the citizen's posterior. The shift in the citizen's beliefs is greater the larger the distance between the signal and her prior belief m. Citizens in the study context are a priori pessimistic about the police and the alarm was designed to increase police capacity. Hence, it seems plausible that $s > m$. The alarm treatment should thus cause citizens to revert their beliefs about police capacity upwards and the shift in their beliefs should be larger the more pessimistic they were at the outset. Regardless of whether one or both of the hypothesized mechanisms are at play, citizens who see a greater upward shift in their perception of police capacity should also see greater shifts in their perceptions of police outputs, p_s and q, and be ultimately most likely to shift their behavior.

To the degree that beliefs about capacity θ are an important determinant of variation in beliefs about police service quality p_s and punishment risks q, citizens who are pessimistic about p_s and q will also be pessimistic about θ and hence strongly affected by a positive capacity shock. E.g., assume police effort and capacity are complements such that $p_s = e_s\theta$ and $q = e_v\theta$. Then, the citizen's prior beliefs about police service quality and punishment are $\tilde{p}_s = e_s m$ and $\tilde{q} = e_v m$, and the effects of receiving the signal s about police capacity on citizens' beliefs about police outputs are given by $e_s\tau_\theta$ and $e_v\tau_\theta$. In this model, citizens can have low prior beliefs about a given police output for two reasons. First, citizens may be pessimistic about p_s or q, because they expect police to make little effort to convict perpetrators of vigilante or non-vigilante crimes (small e_s or e_v). If so, then a positive signal about police capacity will have little effect among pessimistic citizens ($e_s\tau_\theta$ and $e_v\tau_\theta$ are small). If, however, citizens are pessimistic about p_s or q, because they perceive police capacity to be low (small m), then the effects of a capacity shock should be strongest among pessimistic citizens (τ_θ is large if m is small). If heterogeneity in prior beliefs is large in the study context, treatment effects may be difficult to detect in the entire sample. I thus pre-registered to analyze effects within sub-groups defined by prior beliefs about the two police outputs.

Effect heterogeneity does not help distinguish mechanisms. One may be tempted to interpret a finding that treatment effects on citizens' behavior are concentrated among citizens who are pessimistic about a particular police output as evidence that changes in citizens' beliefs about this output must be the mechanism through which the treatment affects behavior. E.g., if the treatment discourages vigilantism only among citizens who a priori judge the risk of state punishment for vigilantism to be small, one may conclude that changes in perceived punishment risks must be the mechanism driving the effect on citizens' behavior. This interpretation reflects the intuition that low prior beliefs about punishment risks "switch on" the punishment risk mechanism, i.e., serve as what Fu and Slough (2023) call a mechanism indicator variable. Yet, as Fu and Slough (2023) show, heterogeneous effects with respect to a mechanism indicator variable can be informative about the relevance of mechanisms only if the indicator variable switches on a single as opposed to multiple mechanisms. This condition does not hold here. To the degree that citizens' pessimistic prior beliefs about police outputs are driven by their perception that police capacity is low, such low prior beliefs plausibly switch on *both* hypothesized mechanisms. E.g., if citizens are pessimistic about punishment risks because they believe police have little capacity, then a positive capacity shock should produce large upward shifts in citizens' capacity perception which should translate into greater expectations of *both* police service quality *and* punishment risks. A similar problem of observational equivalence would persist if we analyzed effects among respondents with low prior beliefs about one but high prior beliefs about another police output. Since it is unknown how perceptions of other inputs like effort correlate with prior beliefs about capacity, any difference in effects across such groups can plausibly be driven by a difference in the distribution of prior beliefs about capacity.

How to distinguish mechanisms. I rely on three other strategies to probe mechanism relevance. First, I test whether the alarm affects citizens' perceptions of police service quality p_s and punishment risks q. Second, I study treatments which individually manipulate perceptions about police effort levels e_s and e_v and as such should would trigger only one of the two mechanisms. Third, I test theoretical expectations about how shocks to perceptions of police efforts and capacity should interact to affect citizens' behavior if a given mechanism is at play. In the context of the alarm treatment, police capacity and efforts are likely to be complements, i.e., $\frac{\partial^2 q(e_v,\theta)}{\partial\theta\partial e_v} > 0$ and $\frac{\partial^2 p_s(e_s,\theta)}{\partial\theta\partial e_s} > 0$. Then, if the discouraging effect of the alarm system on vigilantism is mediated by perceptions of one of the police outputs, this effect should be stronger the more effort citizens expect police to invest into this output. As an example, note that the multiplicative production function assumed above implies that effects of a capacity shock on citizen's perceptions of police outputs are $e_s\tau_\theta$ and $e_v\tau_\theta$.

A.4 Ethical considerations

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

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

Finally, several measures were taken to protect study staff. Enumerators were residents of the study community but worked in sections other than their own. Approvals were obtained from community leaders and police were aware of all survey activities. Where communities seemed hostile, enumeration was stopped until problems were solved with local authorities. Enumerators worked in pairs and were picked up from households by car. Walking was kept to a minimum and enumerators received bags to avoid carrying equipment like tablets in the open. Wherever possible, enumeration stopped before nightfall. If enumerators worked after dark, a car was kept close and they were brought home afterwards.

A.5 General equilibrium effects

How should the alarm affect the likelihood that a household experiences crime? If the alarm's primary effect is to improve law enforcement services to such an extent that citizens voluntarily substitutes reliance on the state for vigilantism, we may expect alarm owners to become safer from crime as a result. If the alarm operates primarily by increasing the risk of state punishment for vigilantism, however, matters are less clear. On the one hand, the alarm is likely to yield some kind of improved police protection that citizens desire – e.g., against heavily armed perpetrators that communities cannot face by themselves. Otherwise, it seems unclear why study participants were willing to take up the alarm in the first place. Yet, the increased legal risk for vigilantism – if perceived by potential attackers of a household – may weaken the deterrence effect of community punishment which may increase alarm owners' risk of experiencing a crime. The direction of the overall impact on alarm owners' safety depends on the relative magnitude of these two effects. Theoretical expectations for the alarm's effect on citizens' actual frequency of contact with police are equally ambiguous. On the one hand, the arguments in the main text suggest citizens should become more willing to reach out to police. Yet, if the alarm makes citizens safer, they may have less of a need to actually do so.

Empirically, the evidence is mixed. Table A21 suggests alarm owners feel safer but their households did not experience fewer crimes. One possibility is that the noisiness of my crime victimization measure masks an increase in safety. Crime victimization surveys usually ask detailed questions about specific crimes. Due to time constraints and to avoid re-traumatization, I only asked the following general question: "Since last Christmas, did any crime happen in your house or yard?" This interpretation would go together with the lack of evidence for an effect on whether respondents recently spoke to police in Table A9. A more complicated interpretation is that alarm owners report feeling safer because the alarm offers better police protection against heavily armed perpetrators but have become less safe from other crime because of a reduction in deterrence through vigilantism. The net effect may be no apparent change in the victimization rate. This interpretation requires citizens to weigh some crimes more heavily than others when forming safety perceptions. How to explain the apparent absence of an effect on contact with police? Perhaps alarm owners did not report crimes they would have dealt with through vigilantism to police and the increase in their willingness to report crimes by highly armed perpetrators is countered by the decrease in exposure to such crimes.

A.6 Sampling strategy for households

At baseline, I sampled households through two strategies. In line with the implementing partner's standard way of selecting alarm recipients, 150 households were sampled from a list provided by police. The police listed 390 households, of which 336 could be geo-located. To limit spillovers, a stochastic algorithm was used to identify sets of households such that every household is located at least 150 m from all other households in the set. Among the identified sets, the set with the overall maximum distance between units and the largest share of units with a distance of at least 200 m from all other units in the set was chosen. Second, 150 households were selected from areas with particularly high crime rates and low trust in the police as judged by local police. Enumerators walked through these areas and geo-located every tenth house. 946 households were geo-located in 11

areas. The same algorithm was used to identify sets of households that satisfy the distance constraint. Both sub-samples cover a similar area, since the police identified households in the same high-crime areas from which the second sub-sample was drawn. Sampled households were replaced if there was no adult woman residing there permanently, if the respondent could not be found after 3 visits, or if the respondent refused to be interviewed. 15 respondents refused and 16 could not be found. Towards the end of the survey, surveying became impossible in two areas due to opposition from the community who did not trust the surveyors. Thus, more households from the other high crime areas in the sample were added. Moreover, the existence of multiple houses with the same numbers led to inaccuracies in the geo-coordinates. Hence, not all households in the sample satisfy the 150 m distance constraint. Additional households were sampled to alleviate this problem. In total, 358 households were interviewed at baseline, 171 from the police list and 187 sampled from high crime areas. 250 were chosen as experimental units so as to minimize non-compliance and attrition:

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

A.7 Sampling strategy for neighbors

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

A.8 Additional figures

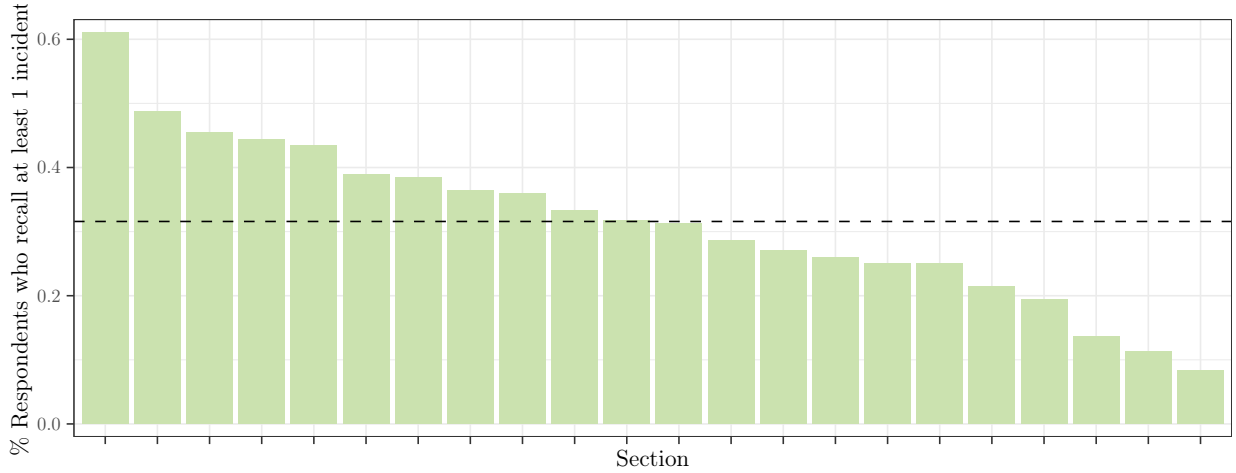

Figure A1: How many mob violence incidents can you recall that happened in your section?
Based on all endline responses ($N = 815$). Question: "I would like you to think back to last year last winter, meaning May, June and July last year (2018). Can you recall any mob justice incidents that happened in your section during last winter?" If yes: "How many mob justice incidents can you recall from last winter?" Bars show share of respondents recalling at least one incident.

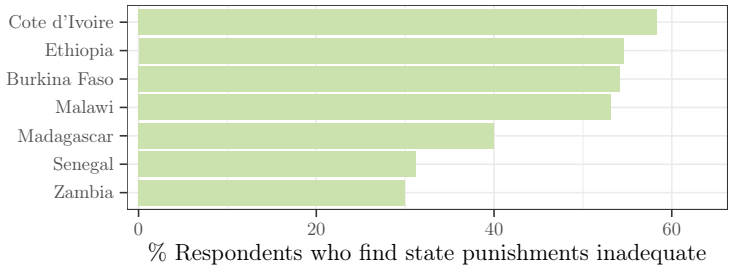

Figure A2: Views on state punishment in Sub-Saharan Africa
Data from a 2017 survey by the World Justice Project with around 1,000 respondents per country from urban centers. Question: "Please tell us how confident you are that the criminal justice system gives punishments which fit the crime?" Figure shows % respondents who chose "Not very confident" or "Not at all confident."

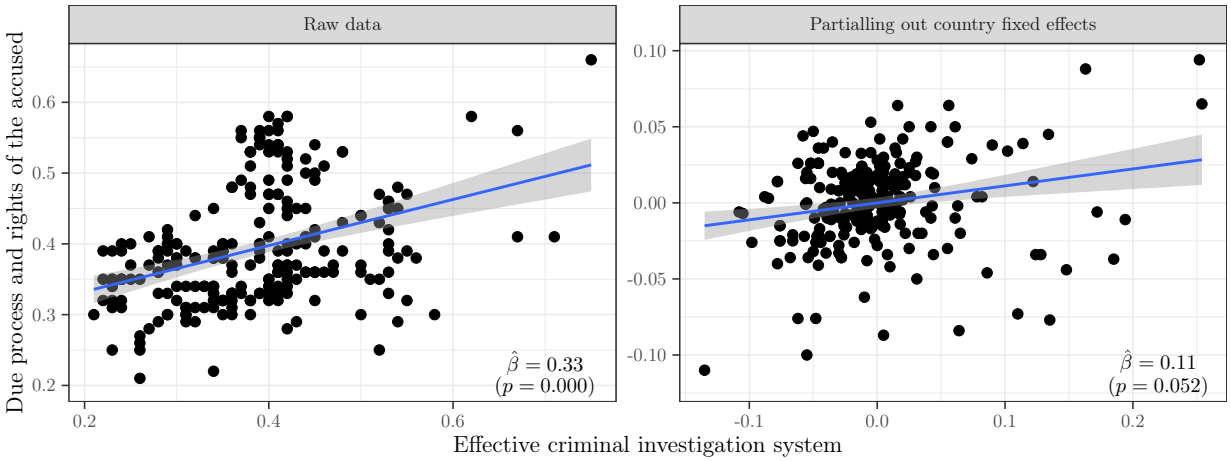

Figure A3: Investigative capacity and due process protections across country-years in Sub-Saharan Africa
Data stem from the World Justice Project Rule of Law Index (Factors 8.1 and 8.7) and cover 34 Sub-Saharan countries for varying years between 2012 and 2023. Dots are country-years. Plots show OLS regression line with 95% confidence intervals.

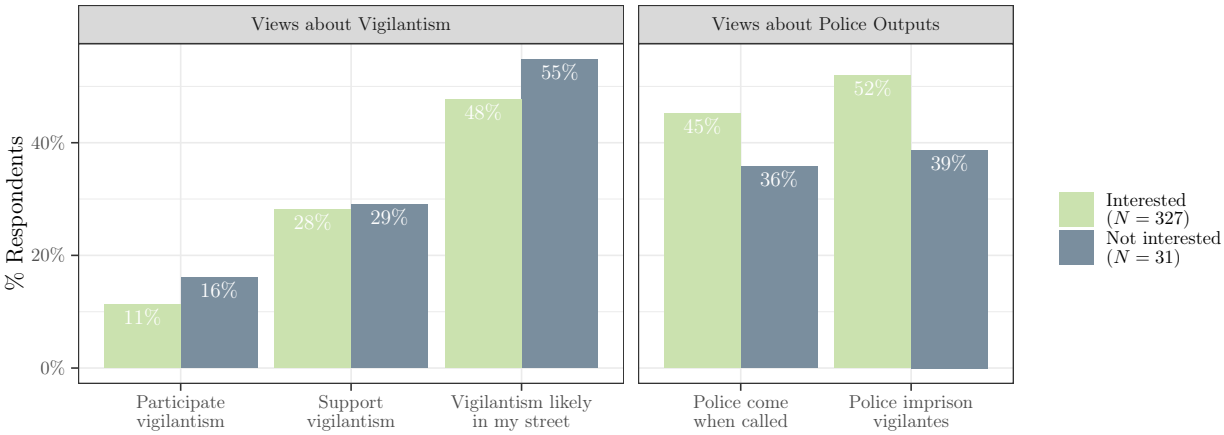

Figure A4: Baseline views by interest in police alarm.
Sample includes 358 baseline respondents. 15 households showed no interest at baseline. 12 were interested at baseline, but had changed their mind when contacted during back-checks. Four refused the alarm after random assignment. Note that it is not known whether households assigned to control might have refused the alarm had they been assigned to treatment.

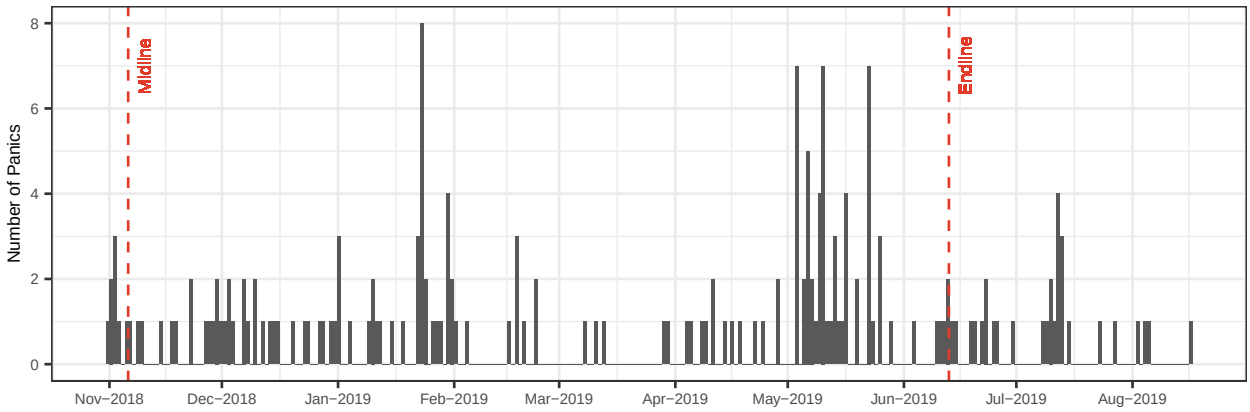

Figure A5: Alarm panics over time.
Data are taken from the back end system of the implementing partner.

A.9 Information treatments

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

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

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

Table A1: Number of respondents across information treatment conditions.
First number in each cell pertains to respondents from all (main and neighboring) and number in parentheses to respondents from main households.

A.10 Descriptive statistics

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

Table A2: Averages of baseline covariates by sampling procedure

Table A3: Averages of baseline covariates by prior beliefs (among $N = 250$ women interviewed at baseline)

			Prior Service/Prior Punishment				Prior Service		Prior Punishment	
	Min	Max	Low/Low	Low/High	High/Low	High/High	Low	High	Low	High
Police										
Was sampled through police	0	1	0.37	0.59	0.56	0.68	0.47	0.63	0.44	0.63
Knows police number	0	1	0.74	0.82	0.82	0.89	0.78	0.87	0.77	0.86
Likelihood would call police	0	3	2.22	2.18	2.28	2.49	2.20	2.41	2.24	2.33
Has approached police	0	1	0.59	0.60	0.51	0.65	0.60	0.60	0.56	0.62
Connection to Community Policing Forum	0	3	1.14	1.22	1.46	1.62	1.18	1.56	1.25	1.41
Vigilantism										
Join MV	0	2	0.54	0.47	0.38	0.32	0.51	0.35	0.49	0.40
Support beat driver	0	1	0.33	0.34	0.21	0.20	0.34	0.20	0.29	0.27
Crime and Safety										
HH experienced crime	0	1	0.46	0.51	0.33	0.42	0.49	0.38	0.42	0.47
Feel safe	0	1	0.21	0.34	0.28	0.26	0.27	0.27	0.23	0.30
Justice										
Courts punish not enough	0	1	0.76	0.85	0.69	0.68	0.80	0.68	0.74	0.77
Punish more	0	1	0.72	0.75	0.72	0.68	0.73	0.69	0.72	0.71
Government										
Satisfaction government services	0	1	0.35	0.33	0.45	0.42	0.34	0.43	0.38	0.37
Government officials corrupt	0	1	0.63	0.65	0.46	0.63	0.64	0.57	0.57	0.64
Frequency make voice heard to government	0	3	0.88	0.91	0.85	0.92	0.90	0.89	0.87	0.92
Knows state official	0	1	0.33	0.44	0.46	0.38	0.38	0.41	0.38	0.41
Community										
Lives in neighborhood for less than 10 yrs	0	1	0.33	0.18	0.18	0.18	0.26	0.18	0.28	0.18
Trusts neighbor	0	1	0.79	0.71	0.79	0.72	0.75	0.75	0.79	0.71
Frequency discuss with neighbors	0	3	1.63	1.59	1.82	1.51	1.61	1.62	1.69	1.55
Member neighborhood organization	0	1	0.81	0.76	0.79	0.85	0.79	0.83	0.80	0.80
Demographics										
Age	20	95	46.55	45.00	45.33	44.74	45.83	44.96	46.15	44.87
Married	0	1	0.40	0.53	0.26	0.35	0.46	0.32	0.35	0.44
HH size	1	16	4.53	4.90	4.77	5.11	4.70	4.98	4.61	5.00
Asset ownership	0	1	0.42	0.47	0.42	0.44	0.44	0.43	0.42	0.46
Has finished secondary school	0	1	0.41	0.47	0.46	0.45	0.44	0.45	0.43	0.46
Frequency pray in private	3	8	7.65	7.63	7.74	7.60	7.64	7.65	7.68	7.62
Observations										
N			78.00	68.00	39.00	65.00	146.00	104.00	117.00	133.00

Table A4: Share women and control group means at endline by prior beliefs

			Prior Service/Prior Punishment				Prior Service		Prior Punishment	
	Min	Max	Low/Low	Low/High	High/Low	High/High	Low	High	Low	High
Share women (N = 448)										
Women	0	1	0.64	0.63	0.66	0.63	0.64	0.64	0.65	0.63
Control group means (N = 259)										
Rely police	0	1	0.58	0.59	0.66	0.71	0.59	0.70	0.61	0.66
Join MV	0	1	0.24	0.17	0.24	0.11	0.20	0.15	0.24	0.13
Service index	0	1	0.49	0.47	0.57	0.65	0.48	0.62	0.52	0.57
Imprison MV	0	1	0.60	0.70	0.72	0.79	0.65	0.77	0.64	0.75
Police know HH	0	1	0.37	0.46	0.42	0.50	0.41	0.48	0.39	0.48
Police are motivated	0	1	0.32	0.39	0.53	0.62	0.36	0.59	0.39	0.52

B Identification

B.1 Covariate balance

Table A5 examines balance among endline respondents. Most covariates are from the endline. Some measures plausibly unaffected by treatment (e.g. age) are from the endline. Columns show covariate means across conditions. To calculate the two-tailed p-value in the last column, I regress each covariate on a treatment assignment indicator and the cluster size control. I simulate the sampling distribution under the sharp null hypothesis of no effect of treatment on a covariate for any unit by permuting treatment assignment 2,000 times and re-estimating the same model. Then, I compare the observed coefficient of the treatment assignment indicator to the sampling distribution. If tests were independent, we would expect 5% of covariates to show imbalance significant at the 5% level. Here, 7/103 (7%) of tests yield a p-value equal to or less than .05.

	Control	Treatment	p-value
prepaid_electricity_bl	0.83	0.93	0.01
electric_stove_bl	0.80	0.94	0.01
microwave_bl	0.58	0.75	0.01
approached_police_bl	0.68	0.53	0.02
spend_police_1_bl	0.27	0.15	0.03
floor_material_missing_el_fu	0.29	0.20	0.04
number_births_bl	2.84	2.45	0.05
earn_salary_el_fu	0.55	0.48	0.09
prisoners_guilty_bl	0.46	0.58	0.09
dishwasher_bl	0.05	0.01	0.10
tsonga_el_fu	0.08	0.15	0.11
interest_public_affairs_bl	2.21	1.99	0.11
own_refuse_dump_bl	0.81	0.89	0.11
discuss_neighbors_bl	1.69	1.47	0.12
pray_private_bl	7.58	7.74	0.12
criminals_from_area_bl	0.40	0.29	0.13
mob_violence_police_reaction_bl	1.91	1.71	0.13
sepedi_el_fu	0.18	0.11	0.14
experienced_violent_crime_bl	0.09	0.17	0.16
tap_water_in_yard_bl	0.61	0.69	0.16
in_a_relationship_el_fu	0.19	0.15	0.19
spend_electricity_bl	0.56	0.49	0.20
other_organizations_bl	0.08	0.04	0.21
retired_el_fu	0.13	0.18	0.22
interview_tswana_el_fu	0.97	0.94	0.22
work_full_time_el_fu	0.19	0.14	0.23
length_stay_el_fu	4.02	4.16	0.23
age_el_fu	41.62	43.08	0.24
pay_tv_bl	0.45	0.53	0.26
no_religion_el_fu	0.10	0.13	0.29
washing_machine_bl	0.49	0.58	0.29
home_language_sepedi_el_fu	0.11	0.07	0.31
hh_head_el_fu	0.38	0.39	0.32
government_does_enough_bl	0.54	0.61	0.32
main_income_salary_bl	0.29	0.35	0.32
pit_latrine_bl	0.73	0.68	0.32
private_security_bl	0.02	0.01	0.33
single_el_fu	0.31	0.35	0.34
flush_toilet_tank_bl	0.13	0.16	0.37
unemployed_el_fu	0.41	0.38	0.38
member_organization_bl	0.80	0.86	0.38
satisfaction_services_bl	0.38	0.35	0.39
kind_day_el_fu	1.58	1.63	0.40
state_official_bl	0.10	0.14	0.40
religious_service_bl	1.38	1.30	0.41
punishment_preferences_bl	0.69	0.74	0.41
know_number_bl	0.80	0.86	0.41
main_income_pensions_bl	0.09	0.14	0.42
observed_conditions_bl	2.75	2.65	0.44
join_mob_bl	0.41	0.43	0.45
motor_vehicle_bl	0.25	0.23	0.45
others_present_el_fu	0.21	0.24	0.46
police_ask_for_bribe_bl	0.85	0.74	0.47
completed_secondary_education_el_fu	0.36	0.32	0.50
tiled_floor_el_fu	0.25	0.30	0.50
trust_neighbor_bl	0.79	0.76	0.50
guard_dogs_bl	0.22	0.28	0.50
discuss_government_bl	2.17	2.30	0.52
child_hh_head_el_fu	0.28	0.27	0.53
lutheran_el_fu	0.25	0.29	0.54
concrete_floor_el_fu	0.38	0.42	0.56
mob_violence_plausibility_bl	1.67	1.75	0.56
beat_truck_driver_bl	0.30	0.31	0.57
due_process_bl	0.85	0.88	0.59
blow_whistle_bl	0.15	0.13	0.60
voice_heard_bl	0.92	0.86	0.61
spend_education_bl	0.59	0.60	0.61
number_incidents_bl	0.93	0.86	0.63

flush_toilet_public_bl	0.13	0.16	0.64
know_state_official_bl	0.39	0.37	0.65
spouse_hh_head_el_fu	0.24	0.22	0.66
work_part_time_el_fu	0.15	0.14	0.67
able_to_name_bl	1.84	1.79	0.68
discussed_crime_bl	0.89	0.92	0.68
secondary_education_incomplete_el_fu	0.43	0.41	0.71
report_informal_provider_bl	0.78	0.76	0.71
spend_police_2_bl	0.52	0.55	0.71
shout_community_bl	0.73	0.76	0.74
street_committee_connection_bl	0.44	0.37	0.76
police_quality_bl	1.57	1.62	0.77
home_language_tswana_el_fu	0.69	0.68	0.81
government_unresponsive_bl	0.76	0.76	0.82
call_police_bl	2.30	2.33	0.84
hh_size_bl	4.98	5.09	0.85
response_time_bl	3.09	3.07	0.87
tap_water_in_house_bl	0.12	0.12	0.87
zcc_el_fu	0.18	0.19	0.88
perceived_crime_risk_bl	1.90	1.87	0.88
number_school_children_bl	1.46	1.50	0.89
main_income_social_grants_bl	0.44	0.43	0.90
apostolic_el_fu	0.21	0.21	0.91
cpf_connection_bl	1.38	1.39	0.91
government_corrupt_bl	0.61	0.62	0.94
number_children_bl	1.91	1.98	0.95
feel_safe_bl	0.28	0.29	0.97
attend_meetings_street_committee_bl	0.47	0.42	0.97
female_el_fu	0.64	0.63	0.98
criminals_from_outside_bl	0.38	0.38	0.98
courts_punish_not_enough_bl	0.75	0.73	0.98
tswana_el_fu	0.49	0.51	0.99
adequate_force_bl	0.57	0.57	0.99
attend_meetings_cpf_bl	1.54	1.63	0.99
married_el_fu	0.34	0.36	1.00

Table A5: Balance on covariates among all respondents in endline ($N = 448$)

B.2 Attrition

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

Table A6: Reported household size and rates of attrition across experimental conditions
The outcome in row 1 is an indicator for whether a household has only one member. The unit of analysis is the household. The outcomes in rows 2 and 3 are indicators for whether a respondent attrited in the midline or endline survey, respectively. The unit of analysis is the respondent. Rows 2 and 3 assume that for the response rate to be 100%, 477 respondents should have been interviewed, two for each household other than the 23 single-member households. p-values stem from an unequal variance t-test conducted via randomization inference by permuting treatment assignment 2,000 times to generate the distribution of the test statistic under the sharp null hypothesis of no effect of treatment on reported household size or attrition for any unit.

	p-value	N
1	0.222	477
2	0.818	477

Table A7: F-test of treatment-by-covariate interactions in models of attrition
P-values come from an F-test that compares the following two models. The full model regresses an indicator for whether a respondent attrited on an indicator for treatment assignment and all treatment-by-covariate interactions using eight pre-registered baseline covariates. The nested model restricts all interaction terms to be zero. Row 1 pertains to the midline survey and row 2 pertains to the endline survey. The unit of analysis is the respondent and the analysis is based on two "completed" datasets which assume that, for the response rate to be 100%, 477 respondents should have been interviewed, two respondents per household other than the 23 households that have only one household member. p-values have been calculated using randomization inference by permuting treatment assignment 2,000 times.

B.3 Additional respondents

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

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

Table A8: Additional respondents sampled across experimental conditions
The unit of analysis is the household. The sample contains all main households in which at least one respondent was interviewed at, respectively, midline and endline. The outcome in columns 1 and 3 is an indicator for whether an additional respondent was interviewed in a given household. The outcome in columns 2 and 4 is the number of additional respondents interviewed. Outcomes are regressed on an indicator for treatment assignment. p-values are calculated using randomization inference.

C Additional Analyses

C.1 Additional outcomes

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

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

Table A9: Effects of the alarm on whether respondents have recently spoken to police
Outcome is binary. Appendix section D.5 contains details on measures of prior beliefs and table 1 shows their distribution. See appendix section A.2 for model specification, and appendix section D.6 on outcome question wording.

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

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

Table A10: Effects of the alarm treatment on respondents' support for mob vigilantism and willingness to call the community.
Outcomes range from 0 to 1. Analyses in columns 5 to 8 regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction, and the cluster size control. One respondent was interviewed per household at baseline and their response is interpreted as a household-level measure of prior beliefs. The measure of priors about punishment (columns 5 and 6) asks whether it is likely (unlikely) that vigilantism perpetrators would be arrested. The measure of service quality priors (columns 7 and 8) indicates whether respondents fall above the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See appendix section D.5 for question wording and Table 1 for the distribution of prior beliefs. The table displays randomization inference *p*-values and directions of hypothesis tests. Appendix section A.2 provides details on model specification and testing, and appendix section D.7 on outcome question wording and coding.

	Join group		Allow continue		Join group	Allow continue	Join group	Allow continue
	Midline	Endline	Midline	Endline	Endline	Endline	Endline	Endline
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Alarm	−0.084***	−0.045*	0.012	0.012	−0.119***	−0.018	−0.079**	−0.010
	(0.029)	(0.030)	(0.034)	(0.035)	(0.046)	(0.052)	(0.041)	(0.048)
Alarm × High Prior Punishment					0.139**	0.052		
					(0.061)	(0.074)		
Alarm × High Prior Service							0.080	0.042
							(0.061)	(0.072)
Control Mean	0.15	0.13	0.18	0.17	0.13	0.17	0.13	0.17
Control SD	0.36	0.33	0.38	0.37	0.33	0.37	0.33	0.37
RI p-value Main	0.002	0.07	0.641	0.621	0.006	0.373	0.029	0.475
Hypothesis Main	lwr	lwr	lwr	lwr	lwr	lwr	lwr	lwr
RI *p*-value Diff.	-	-	-	-	0.02	0.244	0.114	0.352
Hypothesis Diff	-	-	-	-	upr	upr	upr	upr
Number HHs	245	237	245	237	237	237	237	237
Observations	483	448	483	448	448	448	448	448

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

Table A11: Effects of alarm on willingness to participate in and let mob vigilantism continue (alternative coding of Join beating)
Join group and *Allow continue* are both constructed using the outcome *Join beating* which asks: "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?" *Join group* takes the value 1 if the respondent answered "I would join the group in beating up the thief" and the value 0 if the respondent answered "I would not join the group but allow the men to continue with the beating" or "I would try to calm the group down and tell them we should wait for the police.". *Allow continue* takes the value 1 if the respondent answered "I would not join the group but allow the men to continue with the beating" and the value 0 otherwise. Appendix section D.5 contains details on measures of prior beliefs and table 1 shows their distribution. Appendix section A.2 provides details on model specification and testing, and appendix section D.1 on outcome question wording.

C.2 Adjusting for covariates

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

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

Table A12: Effects of alarm treatment on respondents' willingness to rely on police and participate in mob vigilantism estimated with covariate adjustment.

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

C.3 Disaggregating indices

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

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

Table A13: Effects of alarm on components of index "Rely Police" (see Table 2)

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

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

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

Table A14: Effects of alarm on individual items used to create the index "Join MV" at endline.
Outcomes range from 0 to 1. Analyses in columns 3 to 6 regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction, and the cluster size control. One respondent was interviewed per household at baseline and their response is interpreted as a household-level measure of prior beliefs. The measure of priors about punishment (columns 3 and 4) asks whether it is likely (unlikely) that vigilantism perpetrators would be arrested. The measure of service quality priors (columns 5 and 6) indicates whether respondents fall above the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See appendix section D.5 for question wording and Table 1 for the distribution of prior beliefs. The table displays randomization inference *p*-values and directions of hypothesis tests. Appendix section A.2 provides details on model specification and testing, and appendix section D.1 on outcome question wording and coding.

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

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

Table A15: Effects of alarm on individual items used to create the index "Police know HH".
Outcomes range from 0 to 1. All analyses regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction, and the cluster size control. One respondent was interviewed per household at baseline and their response is interpreted as a household-level measure of prior beliefs. The measure of priors about punishment (columns 1 and 2) asks whether it is likely (unlikely) that vigilantism perpetrators would be arrested. The measure of service quality priors (columns 3 and 4) indicates whether respondents fall above the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See appendix section D.5 for question wording and Table 1 for the distribution of prior beliefs. The table displays randomization inference p-values and directions of hypothesis tests. Appendix section A.2 provides details on model specification and testing, and appendix section D.2 on outcome question wording and coding.

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

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

Table A16: Effects of alarm on on individual items used to create the "Service index."
All outcome measures range from 0 to 1. All specifications regress the outcome on an indicator for treatment assignment, an indicator for high prior beliefs at baseline, the interaction between the two, and the cluster size control. Dichotomous baseline measures of prior beliefs are treated as household-level measurements, since only one respondent was interviewed per household at baseline. Prior beliefs about punishment (columns 1 to 3) are measured through an item that asks whether it is likely (unlikely) that participants in a hypothetical incident of vigilantism would be arrested. The measure of prior beliefs about service quality (columns 4 to 6) captures whether respondents fall above or below the median of an index of three items: *Arrive quickly*, *Send guilty to prison* and *Customer service*. See section D.5 for details on question wording and Table 1 for the joint distribution of prior beliefs. Randomization inference p-values and directions of hypothesis tests are displayed in the table. Section A.2 of the appendix contains more details on model specification. See section D.3 for question wording and coding of outcomes.

C.4 Unconditional effects on intermediate outcomes

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

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

Table A17: Effects of alarm on perceptions of police.
All outcome measures range from zero to one. Randomization inference p-values and directions of hypothesis tests are displayed in the table. Section A.2 of the appendix contains details on model specification. See appendix sections D.2, D.3, and D.8 for question wording and coding of outcomes.

C.5 Additional results information treatments

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

Table A18: Effect of information treatments among all endline respondents
Sample includes all endline respondents from main and neighboring households. See appendix sections A.2 and D.4 on model specification and question wording.

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

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

Table A19: Effect of information treatments among respondents with low priors about police service
Sample includes respondents from main and neighboring households with low priors about police service quality as measured at endline. See appendix section D.5 for prior belief measures, section A.2 on model specification and section D.4 on outcome question wording.

C.6 Ruling out alternative explanations

C.6.1 Social desirability bias

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

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

Table A20: Effect of the alarm treatment on recollection of incidents of mob vigilantism that happened *prior* to treatment
During the endline survey, respondents were asked "I would like you to think back to last year last winter, meaning May, June and July last year (2018). Can you recall any mob justice incidents that happened in your section during last winter?" If they answered "yes," they were asked "How many mob justice incidents can you recall from last winter?" as well as "Did you personally witness any of these mob justice incidents?" The outcome in columns 1 and 4 is an indicator variable for whether respondents can recall any incidents. The outcome in columns 2 and 5 is the number of incidents that a respondent can recall. The outcome in columns 3 and 6 is an indicator for whether a respondent reports having witnessed any incidents of vigilantism. Those who cannot recall an incident are coded as zero. Analyses in columns 1 to 3 are based on the entire sample. Analyses in columns 4 to 6 are subset to respondents from households with low priors about the likelihood of state punishment for vigilante violence. This subgroup is of relevance, because it sees the largest (in absolute value) treatment effects on the willingness to participate in vigilante violence. See section A.2 of the appendix for information on model specification.

C.6.2 Changes among control group

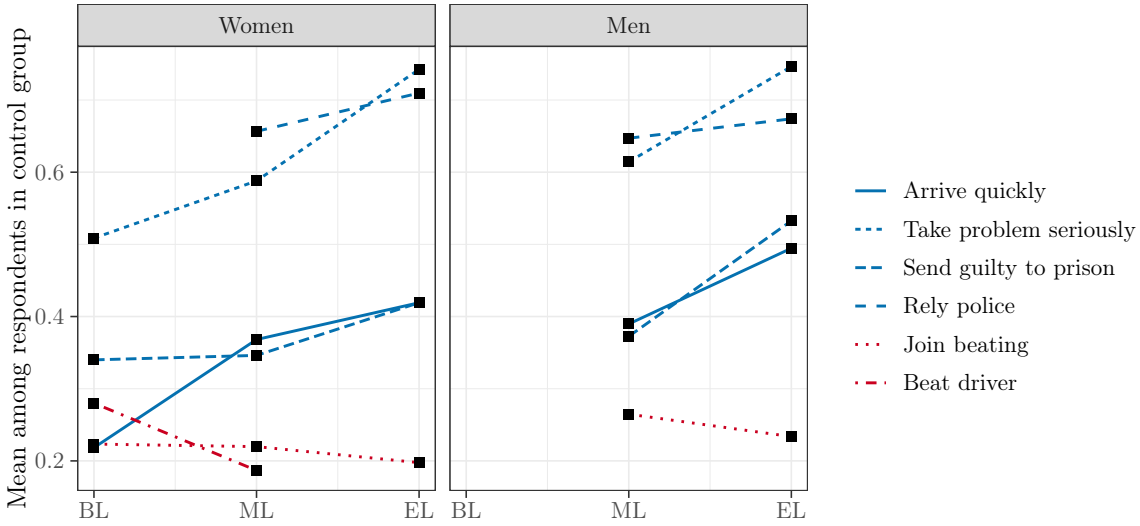

Figure A6: Change in outcomes in control group across survey waves by gender

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

C.6.3 Change in punishment preferences due to improved safety

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

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

Table A21: Effect of the alarm treatment on safety and punishment preferences

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

C.7 Statistical power

	Estimate	MDE	Power for MDE (RI)
Column 1	0.097^{***} ($p = 0.000$)	0.07	80%
Column 2	0.075^{***} ($p = 0.003$)	0.076	77%
Column 3	-0.078^{***} ($p = 0.006$)	-0.079	80%
Column 4	-0.012 ($p = 0.344$)	-0.07	81%
Column 5 Main	0.132^{***} ($p = 0.002$)	0.11	77%
Column 5 Diff.	-0.108^{**} ($p = 0.034$)	-0.153	82%
Column 6 Main	-0.1^{**} ($p = 0.011$)	-0.109	81%
Column 6 Diff.	0.158^{***} ($p = 0.002$)	0.142	78%
Column 7 Main	0.126^{***} ($p = 0.001$)	0.103	80%
Column 7 Diff.	-0.104^{*} ($p = 0.061$)	-0.151	77%
Column 8 Main	-0.042 ($p = 0.148$)	-0.091	82%
Column 8 Diff.	0.066 ($p = 0.206$)	0.141	75%

Table A22: Minimum detectable effects and statistical power (main results)
The column labelled "Estimate" shows the estimates and RI p-values shown in Table 2 in the main text. Estimates labeled *Main* are of conditional effects among low prior groups. Those labeled *Diff* are of interaction terms. The column labelled *MDE* shows estimated minimal detectable effect sizes calculated as 2.486 times the estimated parametric standard error for the respective coefficient. This formula corresponds to that suggested by Gelman and Hill (2007, chap. 20), but using a critical value of 1.64 to account for the fact that I conduct one-tailed tests as pre-registered. The column labelled *Power for MDE (RI)* shows the estimated statistical power of the design to detect a constant treatment effect equal to the estimated MDE using RI. To estimate power, I subtract the estimated MDE from the observed outcome for all units in the treatment group and add the estimated MDE to observed outcomes in the control group to create a complete schedule of potential outcomes. I then permute treatment assignment to simulate 1,000 experiments and calculate an RI p-value for each of the simulations by comparing the observed estimate to a reference distribution of 1,000 estimates simulated under the sharp null hypothesis of no treatment effect for any unit. I report the proportion of experiments that yield a p-value smaller than 0.05.

Outcome	ICC	Lower Bound	Upper Bound
Rely Police Midline	0.23	0.10	0.34
Rely Police Endline	0.30	0.17	0.41
Join MV Midline	0.08	-0.04	0.21
Join MV Endline	0.22	0.09	0.35

Table A23: Estimated intra-cluster correlation in main outcomes and 95% confidence intervals
Estimates are obtained using the *ICC* package in R which relies on the variance components of a one-way ANOVA. Confidence limits are calculated using the *THD* method for exact limits. Note that the degrees of freedom correction in the calculations can lead to estimates of the ICC or confidence bounds to be negative.

C.8 Effects by joint distribution of prior beliefs

	Rely police				Join MV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Alarm	0.154***	0.126*	0.097*	−0.029	−0.107**	−0.127*	0.045	0.082
	(0.055)	(0.073)	(0.068)	(0.058)	(0.053)	(0.088)	(0.053)	(0.061)
Prior Punishment	low	low	high	high	low	low	high	high
Prior Service	low	high	low	high	low	high	low	high
Control Mean	0.58	0.66	0.59	0.71	0.24	0.24	0.17	0.11
Control SD	0.33	0.28	0.3	0.28	0.33	0.34	0.26	0.25
RI p-value Main	0.003	0.064	0.064	0.699	0.026	0.059	0.835	0.907
Hypothesis Main	upr	upr	upr	upr	lwr	lwr	lwr	lwr
Number HHs	73	37	67	60	73	37	67	60
Observations	131	71	123	123	131	71	123	123

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

Table A24: Effects of alarm on willingness to rely on police and participate in mob vigilantism by joint distribution of priors.

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

	Police know HH				Police are motivated			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Alarm	0.094	0.110	0.087	0.212**	0.252***	0.154	0.084	0.028
	(0.082)	(0.124)	(0.097)	(0.082)	(0.081)	(0.119)	(0.101)	(0.098)
Prior Punishment	low	low	high	high	low	low	high	high
Prior Service	low	high	low	high	low	high	low	high
Control Mean	0.37	0.42	0.46	0.5	0.32	0.53	0.39	0.62
Control SD	0.44	0.46	0.46	0.44	0.47	0.51	0.49	0.49
RI p-value Main	0.132	0.16	0.142	0.011	0.002	0.106	0.211	0.402
Hypothesis Main	upr	upr	upr	upr	upr	upr	upr	upr
Number HHs	73	37	67	60	73	37	67	60
Observations	131	71	123	123	131	71	123	123

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

Table A25: Effects of alarm on perceptions of legibility and police motivation by joint distribution of priors.

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

	Imprison MV				Service index			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Alarm	0.165**	−0.006	0.053	0.008	0.085*	0.071	0.074*	−0.075
	(0.077)	(0.130)	(0.082)	(0.084)	(0.051)	(0.088)	(0.057)	(0.057)
Prior Punishment	low	low	high	high	low	low	high	high
Prior Service	low	high	low	high	low	high	low	high
Control Mean	0.6	0.72	0.7	0.79	0.49	0.57	0.47	0.65
Control SD	0.49	0.45	0.46	0.41	0.3	0.31	0.3	0.28
RI p-value Main	0.023	0.531	0.278	0.482	0.056	0.202	0.094	0.929
Hypothesis Main	upr	upr	upr	upr	upr	upr	upr	upr
Number HHs	73	37	67	60	73	37	67	60
Observations	131	71	123	123	131	71	123	123

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

Table A26: Effects of alarm on perceptions of risk of state punishment for MV and of police service quality by joint distribution of priors.

Outcomes range from 0 to 1. Appendix section D.5 contains details on measures of prior beliefs and Table 1 shows their distribution. Appendix section A.2 provides details on model specification and testing, and section D.3 on outcome wording.

D Question Wording

D.1 Table 2

Column 1, 2, 5 and 7: Rely Police. 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. A single item at midline and an index of two items at endline:

- *Join beating:* In this same situation, suppose some men from your community do get hold of the burglar who stole from you and that they want to beat him up. Which action are you most likely to take? 0 = I would try to calm the group down and tell them we should wait for the police, 0.5 = I would not join the group but allow the men to continue with the beating, 1 = I would join the group in beating up the thief.

- *Join mob (only endline):* Suppose you are on your way home. In your street, you encounter a group of [at random: 10/50] community members. The community members are beating a man who has been caught stealing from your neighbor's yard. Would you join the group? 1 = Yes, 0 = No

D.2 Table 3

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

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

- *Know House:* Do you think someone from the police knows your house? 0 = No, 1 = Yes

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

D.3 Table 4

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

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

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

- *Take problem seriously:* When you or someone like you takes a problem to the police, how likely is it that the police take your problem seriously? 1 = Very likely, 0.5 = Somewhat likely, 0.25 = Not very likely, 0 = Not likely at all

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

- *Arrive quickly:* Imagine you are at home and alert the police in an emergency. Do you think the police would come to your help? If Yes or Maybe: Do you think the police would take more or less than an hour to come to your help? If you don't know, please give your best guess. If More than an hour: Do you believe the police would take more than two hours or less than that? If Less than an hour: Do you believe the police would take less than 30 minutes or more than that? 0 = The police would not come, 0.25 = The police take more than two hours, 0.5 = The police would take more than one hour but less than two, 0.75 = The police would take less than one hour but more than 30 minutes, 1 = The police would take less than 30 minutes

D.4 Table 5

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

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

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

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

D.5 Measures of prior beliefs

Alarm Treatment

Priors are measured at baseline for one woman per household and treated as household measures.

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

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

- *Customer service:* When you or someone like you takes a problem to the police, how likely is it that the police [at random: take your problem seriously/ appear to know what they are doing]? 1 = Very likely, 0.5 = Somewhat likely, 0.25 = Not very likely, 0 = Not likely at all

- *Arrive quickly:* Imagine you are at home and alert the police in an emergency. Do you think the police would come to your help? If Yes or Maybe: Do you think the police would take more or less than an hour to come to your help? If you don't know, please give your best guess. If More than an hour: Do you believe the police would take more than two hours or less than that? If Less than an hour: Do you believe the police would take less than 30 minutes or more than that? 0 = The police would not come, 0.25 = The police take more than two hours, 0.5 = The police would take more than one hour but less than two, 0.75 = The police would take less than one hour but more than 30 minutes, 1 = The police would take less than 30 minutes

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

High prior (1 =): score strictly above baseline sample median. Low prior (0 =): below median score.

Information Treatments

Analyses of information treatments only (Tables 5 and A19) use prior belief measures asked at endline prior to the administration of information treatments. Since these measures were collected after alarm installations, I do not use them for analyses of the alarm treatment.

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

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

D.6 Table A9

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

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

D.7 Table A10

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

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

- *Beat known thief:* Someone in your community is known to be involved in stealing cars and plasma TVs. One day, the community catches him red-handed as he is breaking into a house. Which of the following do you believe the community members should do? 0 = The community should call the police and leave it to them to deal with the thief, 1 = The community members should beat the thief there and then.

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

- *Beat driver (only midline):* Imagine the following situation: A truck driver drove drunk through your neighborhood and knocked over a small girl and the girl died. A group of men from your community got hold of the truck driver. Which of the following do you believe they should do? 0 = The group should leave it to the police to investigate, 1 = The group of men should beat the truck driver to teach him a lesson.

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

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

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

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

D.8 Table A17 column 6

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

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

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

References

Fu, Jiawei, and Tara Slough. 2023. "Heterogeneous Treatment Effects and Causal Mechanisms." *Unpublished Manuscript.* .
URL: *https://taraslough.github.io/assets/pdf/hte.pdf*

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and and Multilevel/Hierarchical Models.* New York: Cambridge University Press.