DOCUMENT DE TRAVAIL DT/2023-01

Measuring insecurity-related experiences and preferences in a fragile State. A list experiment in Mali.

Olivia BERTELLI

Thomas CALVO

Emmanuelle LAVALLEE

Marion MERCIER

Sandrine MESPLE-SOMPS

UMR LEDa
Place du Maréchal de Lattre de Tassigny 75775 • Paris •Tél. (33) 01 44 05 45 42 • Fax (33) 01 44 05 45 45
DIAL • 4, rue d'Enghien • 75010 Paris • Tél. (33) 01 53 24 14 50 • Fax (33) 01 53 24 14 51
E-mail : dialogue.leda@ird.fr • Site : dial.ird.fr

Measuring insecurity-related experiences and preferences in a fragile State. A list experiment in Mali.*

Olivia Bertelli[†‡] Thomas Calvo[§‡] Emmanuelle Lavallée[‡] Marion Mercier[¶‡]

Sandrine Mesplé-Somps[§‡]

March 15, 2023

Abstract

Measuring behaviors and preferences in times of conflict is of great interest for understanding conflict dynamics and designing conflict-resolution interventions. Yet, data users often cast doubts on the reliability of sensitive self-reported measures, especially in fragile contexts. We study sensitive experiences and preferences related to insecurity in a fragile State – Mali – by explicitly addressing potential response biases using a List Experiment (LE) method. We survey 1,500 individuals across the entire country and randomly assign respondents to answer sensitive questions through the LE or direct questions (DQ) techniques to measure response biases. We focus on three experience-related items (physical assault victimization, firearms' possession, willingness to engage in violence) and two preference-related items (support for the military regime and trust in foreign armed forces in Mali). Results show significant biases affecting responses about preference-related items. Our analysis confirms that popular support for the military regime and mistrust in the foreign armed forces are large, but suffer from a substantial overestimation. Misreporting is not uniformly distributed across the population, but varies depending on gender, education and conflict exposure. Further results suggest that such heterogeneity in response bias can yield fake significant correlations between individual characteristics and sensitive items' prevalence rates depending on the survey technique used.

Keywords: Survey method; Measurement bias; Fragile State; Africa.

JEL Classification: C83; D74; O55.

*This work was made possible thanks to the funding of the French National Research Institute for Sustainable Development (IRD), Project GDRI–MOV (*Groupement de recherche international - Sud "Mesurer et Observer les Violences, Violences de genre et de conflits dans les pays en voie de développement"*). We are grateful to Prof. Massa Coulibaly and Moussa Coulibaly of the GREAT team for the survey management in Mali, to Ousmane Z. Traore for the coding of the questionnaire. We sincerely thanks all the enumerators. We thank Olivier Bargain, Tanguy Bernard, Anne Boring, Delphine Boutin, Margherita Comola, José de Sousa, Kai Gehring, Marine Jouvin, Florence Kondylis, Camille Laville, Pierre Mandon for their helpful comments, as well as participants to the 2022 ICDE, HiCN and ASWEDE conferences and to seminars at DIAL, ENS Lyon, Erasmus University of Rotterdam, Bordeaux School of Economics and Paris-Saclay for insightful discussions. We also thank the Paris School of Economics internal review board for approving our experimental protocol. The Pre-analysis plan is registered under: Bertelli, Olivia et al.(2021) "Violence-related experiences and opinions in Mali: a list-experiment based analysis." AEA RCT Registry. The usual disclaimer applies.

[†]Corresponding author. Email: olivia.bertelli@dauphine.psl.eu

[‡]DIAL-LEDa, UMR CNRS 8007 & IRD 260, Université Paris-Dauphine, PSL Research University.

[§]IRD

[¶]CNRS

1 Introduction

When asked directly about sensitive issues, individuals can be reluctant to answer truthfully. They might feel the question to be intrusive, they might perceive their sincere answer as socially unacceptable and/or they might fear repercussions if their answers are not kept confidential (Tourangeau & Yan, 2007).[1] Misreporting is of particular concern in fragile States where privacy protection and legal rights are poorly enforced (Isaqzadeh et al., 2020). In conflict-torn areas, obtaining honest answers about experiences of violence, participation in armed violence or support for conflict actors might be even harder. Yet, disposing of reliable information about experiences and preferences related to insecurity is of prime importance. These are key elements for understanding and anticipating the evolution of conflict-related violence, for providing assistance to victims and for implementing appropriate conflict-resolution and post-conflict policies.

We study the prevalence of sensitive experiences and preferences at times of conflict. We implement five list experiments (LEs) to measure the unbiased prevalence rates of five insecurity-related outcomes of interest. We quantify the response biases in self-reported measures by comparing the LE unbiased prevalence rates to those based on standard direct questions. We conduct our study during the on-going conflict in Mali, short after the 2021 military coup, a context where conflict-related insecurity remains widespread.

We rely on a nation-wide sample of men and women whom we randomly assign to a treatment and a control group. To the former, we administer five items' lists, each one including a different sensitive item: experience of physical assault, firearms' possession, willingness to take up arms to defend the community, support for the military regime in Mali, and trust in the foreign armed forces present in the country. To the latter, we administer the same lists, but without the sensitive items, that are, instead, asked as direct binary questions. Two prevalence rates for each sensitive item can thus be retrieved: one based on the direct responses declared by individuals in the control group, and one based on the comparison between the average responses to the lists across the treated and control groups.

We argue that insecurity-related experiences and preferences in Mali are sensitive. Thus, social desirability and the fear of lack of confidentiality are likely to bias response obtained with the direct questioning (DQ) technique, whereas the LE technique shall eliminate such bias. Therefore, in the absence of response bias, the prevalence rates measured under the two techniques should not be significantly different from each other – on the contrary, a significant difference provides a measure of the response bias. This measure of misreporting hence captures the two strategies that respondents can adopt when asked a direct question they do not want to answer truthfully: lying or refusing to answer.

We document three main results. First, we find substantial misreporting in insecurity-related experiences and preferences. In particular, declared direct responses about the support for the military regime and trust in foreign armed forces suffer from significant biases: the support for

[1] A vast literature highlights the inaccuracy of sensitive self-declared data in standard surveys. See for instance Fendrich & Vaughn (1994) about drug use, Gong (2015) and Minnis et al. (2009) about sexual activity or Kraay & Murrell (2016) about corruption.

the military regime is over-estimated by 10 percentage point with the DQ technique as compared to the LE, and the level of trust in the foreign armed forces under-estimated by 11 percentage points. In turn, no significant response bias is found with respect to the experience of physical assault, firearms' possession and willingness to take up arms to defend the community.

Second, we document that response biases are not homogeneously distributed across the population. We explore heterogeneity in DQ misreporting along respondents' characteristics (gender, education, and exposure to conflict-related violence). Results show that female respondents and those without primary education over-report their support for the military regime, whereas men under-report their trust in the foreign armed forces. We also find that female respondents, those with at least primary education and those living in conflict-affected areas under-report household's firearms possession.

Our third set of results explores how the correlation between sensitive items' prevalence rates and respondents' characteristics may vary across survey techniques. We find that misreporting can translate into fake correlations between individual characteristics and DQ-measured sensitive outcomes. For instance, men appear to be more likely than women to declare owning firearms and to be willing to engage in conflict when directly asked. This gender heterogeneity, however, fades away when the sensitive items are measured with the LE technique. Similarly, respondents with at least primary education are more likely to report being physically assaulted and being willing to engage in conflict when directly asked, but this correlation vanishes under the LE technique. On the contrary, living in conflict-affected zones is associated with a higher willingness to engage in conflict, as measured by the LE technique, something that the DQ-based measure fails to capture, due to misreporting. These results suggest that asymmetric response biases in DQ across sub-groups of respondents might translate into biased standard regression analysis.

This paper contributes to three strands of the literature. First, we add to the sensitive behavior measurement methodology literature, that discusses diverse survey techniques to elicit the prevalence of specific traits. A recent meta-analysis by Blair et al. (2020) illustrates that sensitive items close to the ones we focus on are generally found to suffer from response bias. They show that across 21 studies respondents over-report support for authoritarian regimes by 14 points and under-report opposition to them by 8 points. Closely related to our work and focusing on an insecurity-related outcome in a context of conflict, the work by Blair et al. (2014) conducts a LE and an endorsement experiment about the support for the International Security Assistance Force (ISAF) during the war in Afghanistan.[2] While both indirect survey methods yield consistent estimates of the level of support, the authors do not measure the possible response bias affecting the same question under direct reporting. We differ from their study, as we adopt a systematic comparative approach between LE-based and DQ-based measures on a set of sensitive items about experiences and preferences related to insecurity.

In this respect, we directly relate to the set of papers employing LEs to measure the preva-

[2] In an endorsement experiment, treated respondents rate their support for policies endorsed by socially sensitive actors, whereas control respondents rate the same policies without endorsement by any particular actor. The difference between the two groups is interpreted as the (lack of) support for the actor of interest. See also for instance Bullock et al. (2011) or Lyall et al. (2013).

lence of sensitive behaviors or opinions and comparing LE-based to DQ-based prevalence across respondents to quantify response biases. As such, misreporting is documented on a wide variety of behaviors and contexts, like voting (Holbrook & Krosnick, 2009; Rosenfeld et al., 2016), loan use (Karlan & Zinman, 2012), physical violence (Porter et al., 2021), sexual behavior (Coffman et al., 2017; Jamison et al., 2013), the use of condoms among female sexual workers (Lépine et al., 2020; Treibich & Lépine, 2019) and domestic violence and/or intimate partner violence against women (Agüero & Frisancho, 2022; Bulte & Lensink, 2019; Chuang et al., 2021; Cullen, 2022; Joseph et al., 2017; Peterman et al., 2018; Traunmüller et al., 2019; Treibich & Lépine, 2019). As far as preferences are concerned, Coffman et al. (2017) document a higher prevalence of anti-gay sentiment in the U.S. under LE than directly admitted; Asadullah et al. (2021) a higher rate of acceptance of IPV and child marriage among young girls in Bangladesh; and De Cao & Lutz (2018) and Gibson et al. (2018) a higher support to female-genital mutilation practices in Ethiopia. To our knowledge, our paper is the first to quantify response biases in insecurity-related experiences and preferences in a conflict environment. In addition, we explore the heterogeneity of response biases across sub-groups of the population and investigate the consequences of using biased self-reported data on correlation analyses.

Second, we relate to the literature on data collection and research conducted in sensitive contexts. Both the qualitative and the quantitative literature are very much concerned about suitable investigation protocols when the research object tackles potentially traumatic events and/or when the research setting is prone to violence – for ethical and security reasons, and for data quality. In particular, Wood (2006) discusses research practices and ethical challenges in the frame of a fieldwork during the civil war in El Salvador; Nordås et al. (2016a,b) illustrate specific protocols implemented to interview survivors of sexual violence participating in support programs in the DRC; and Brück et al. (2016) review the standard practices and methodological challenges related to the collection of survey data in conflict settings, and draw some guidelines on the conduct of quantitative empirical work in fragile areas. Our approach contributes to this literature by documenting the reliability of a specific, indirect survey technique to measure insecurity-related experiences and preferences in a violence-prone context.

Last, our paper relates to a broader extent to the strand of the literature that documents behaviours and preferences in conflict-torn or post-conflict settings. In particular, a series of papers highlights specific attitudes among individuals who were exposed to violence. For instance, Rohner et al. (2013) and Cassar et al. (2013) discuss the impact of violence exposure on trust, respectively in Uganda and Tajikistan; and Bellows & Miguel (2009) and Calvo et al. (2020) illustrate the legacies of conflict in terms of social capital, the former in the case of Sierra Leone and the latter in the case of Mali. Other related papers implement experimental games to look over behavioral outcomes, such as altruism, pro-social attitudes, risk-taking and time-preferences among communities that experienced violence – see in particular Voors et al. (2012) on Burundi, Gilligan et al. (2014) on Nepal, Bertelli & Kurdi (2022) on Yemen. We add to this body of literature by measuring insecurity-related experiences and preferences and documenting the magnitude and heterogeneity of reporting biases on self-reported data in a fragile State undergoing conflict.

The rest of the paper is organized as follows. Next section presents the Malian context. In Section 3 we discuss the data collected, as well as external data on violence that we make use of. Section 4 explains the empirical strategy and discusses the experimental validity of our setting. We present the main results in Section 5. The last section draws the main conclusions.

2 The Malian context

Mali has been at war since 2012. The conflict broke out when a number of armed groups challenged the government's authority. The insurgents were mainly Tuareg rebel groups fighting for the autonomy or the independence of northern Mali and jihadist groups keen to impose the Sharia law. Although Mali is not new to Tuareg uprisings,[3] jihadist groups only started thriving in the 2000s, driven by growing insecurity and central government withdrawal from the Sahelian regions (Julien, 2011).

In January 2012, the jihadists together with the Tuaregs launched their first attacks on military positions in the North. After three months of intense fighting, the major cities of the North, including Timbuktu, Gao and Kidal, were under the insurgents' yoke. Meanwhile, a military coup toppled President Amadou Toumani Touré, held responsible for the crisis. As armed groups moved South, the French government launched the *Serval* military operation soon supported by an African-led mission to Mali. By the end of January 2013, most cities had been taken back.

Although the conflict was initially rooted in Northern Mali, it progressively spread and intensified to central Mali, and throughout the country. Attacks took place even on the Ivorian border, the southern edge of the country. Despite two peace agreements signed in 2013 and 2015 (in Ouagadougou and Algiers), jihadist groups continuous attacks and political instability have spawned the emergence of new insurgent groups with ethnic and community demands. This long-lasting conflict is accompanied by growing political instability and a considerable weakening of the Malian State.

The present study was thus carried out in a particularly fragile context. On May 24, 2021, a month before the start of the survey, Mali experienced its second coup in less than a year. After dismissing president Ibrahim Boubacar Keita in August 2020, the military junta, under the lead of Colonel Assimi Goïta, arrested and removed from office the president and prime minister of the transition before delaying the organization of presidential elections to 2026 and then to 2024.[4] In response, the Economic Community of West African States (ECOWAS) decreed the closure of Mali's borders with the other member States, as well as an economic and financial embargo. In this new political context marked by repeated accusations against foreign armed forces of supporting insurgent armed groups and demands for Russian military aid from the government, France and its allies have announced the departure of their troops from Mali. Our survey took place before this announcement.

[3] Four major uprisings in 1963, 1991, 2006 and 2012 were all followed by peace agreements.
[4] The decision to postpone the presidential elections was made after the survey took place, first in December 2021 and second in June 2022.

To date, the Malian conflict has claimed thousands of lives and displaced hundreds of thousands. It appears to be in a deadlock, responses from national and international authorities have fallen short of the mark (Guichaoua & Pellerin, 2017), and the climate of violence spreads in West Africa and destabilizes the entire region.

3 Data

3.1 Items' lists

List experiment, also named unmatched count or item count technique, is a popular tool for eliciting truthful responses to sensitive questions, as it grants a much higher confidentiality level than standard face-to-face techniques. A list of statements is administered to respondents, who are asked to report the number of statements that they agree with or that apply to their situation, without naming which one(s). The statements are either neutral (*i.e.* concern everyday life) or sensitive. The sample of respondents is randomly allocated to answer either a list of only neutral statements (control group), or the same list of neutral statements augmented with one sensitive item (treatment group). In this way, the difference in the average response (*i.e.* the number of statements with which the respondent agrees) between the two groups provides an accurate measure of the prevalence rate of the sensitive item (Blair & Imai, 2012). This survey technique reduces declarative bias insofar as sensitive information is not revealed directly. No direct answer about the sensitive item is provided to the interviewer, granting a much higher confidentiality than direct questioning (Blair et al., 2020).

In our study, respondents in the control group were administered five lists of three neutral items each. In the treatment group, each list was augmented by a fourth item about an insecurity-related experience or political preference. Respondents in the control group were also asked about the same sensitive item in a direct way, with a "Yes/No" question. They had the option of skipping the question and go directly to the next one if they did not want to answer. It is important to mention that respondents in both groups did not have the option of skipping a whole list and go to the next one, but they could interrupt the questionnaire at any time. Enumerators carefully explained the functioning of questions asked under the form of lists, and guided the respondents through the following example before administrating the first true LE:[5]

"*We will first do a test on a first list to see if you have understood how these questions work. Please tell me the number of situations that are true in your life or with which you agree, between 0 and 3:*"

1. *I like to chat with my neighbours*

[5] To introduce the list experiment, enumerators read the following statement (the following example is for the control group): "Please Madam/Sir, among the following statements, tell me the number of situations that are true in your life or with which you agree. Please do not tell me which ones, but just the number between 0 and 3. For each list, you can close your fist and for each situation that occurs in your life or with which you agree, you can raise a new finger. At the end of each list, tell me only the number of fingers that you have raised, between 0 and 3."

2. *Yesterday I went to the market*

3. *I have been sick recently*

Once the respondent answered, the enumerator had to reply in the following way: "*Alright, so you agree with [figure] proposals. This means that you do not agree or you do not feel concerned by [3-figure] proposals. Is that correct?*". The enumerator had to verify the answer and had to go through the explanation and the example again if the exercise went initially wrong, until the respondent understood properly the functioning of the lists.[6]

The enumerator could then move on to the actual lists. The items' lists and direct questions constituting the object of this analysis are reported in Table A1 in Appendix A. Our sensitive items focus on two main aspects: personal experience of insecurity and political preferences in a fragile State. The former dimension is captured by the first three items:

1. Over the last 12 months, me or a member of my household was physically assaulted outside home;

2. Me or a member of my household owns a firearm;

3. Me or a member of my household is ready to take up arms to defend our community;

and the latter by the last two items:

4. Today's Mali should be ruled by the army;

5. I trust the foreign armed forces in Mali.

Our first two items could be administered anywhere in the world, and do not relate to a context of insecurity that is specific to Mali only. Nevertheless, collecting unbiased data on these dimensions is especially important in countries with widespread insecurity conditions. The last three items relate to context-specific attitudes, particularly related to the fragile context of Mali. The third item assumes a threat to one's own community. Among the possible threats communities may face, jihadist groups are the main one in Mali but we can also count other armed rebel groups (including Tuareg rebel groups), local and ethnic militias and, depending on which side one stands on, the Malian army or the foreign armed forces. Given the widespread insecurity context, respondents might be afraid of openly reporting victimization, firearm possession and involvement in violence.

The fourth and fifth items question the popular support for two important actors in the country's current political situation – the Malian army and the foreign armed forces. In the case of Mali, several foreign armed forces, including France, the former colonial power, are deeply involved in the conflict. Anecdotal evidence suggests that the support for the Malian army and mistrust toward the foreign military forces present over the territory were dominant by the time

[6] Unfortunately, the data from this exercise were not collected.

of our survey (Haidara, 2020, 2021). Besides, as our survey was run after the coup that brought the army into power, the fear for possible reprisal likely increased declared loyalty toward the national army. We thus expect an over-reporting of the support for a military government and an under-reporting of the trust in the foreign armed forces due to social desirability bias and the fear of limited confidentiality.

The choice of the baseline items aims at limiting the design, ceiling and floor effects (Blair & Imai, 2012; Glynn, 2013).[7] We include items that are neither too common nor too uncommon and that are likely negatively correlated with the sensitive items and among themselves.Moreover, we chose items that are neither too innocuous nor too sensitive, following Chuang et al. (2021).[8] In addition, we avoided too short lists, as these tend to increase the likelihood of ceiling effects. Finally, we simplified as much as possible the phrasing of each item to avoid both memory load and to ease translation to Malian languages (Bambara, Soninké, Tamacheq, Peul and Sonhrai). We provide supportive evidence of the absence of any design, ceiling and floor effects in subsection 4.2.

Nevertheless, our approach has some limitations. First, we cannot test for the negative correlation between the baseline items as we did not administrated them in a DQ form. Nevertheless, we devoted a considerable amount of work and reflection to identify the baseline items following the guidelines of Chuang et al. (2021). Besides, whenever possible, we relied on alternative sources of data to verify the negative correlation among the baseline items. Second, the lists' order was not randomized across respondents, implying a higher risk of fatigue for the last lists. To mitigate as much as possible this risk, we made the survey as short as possible (interviews were, in most cases, between 15 to 20 minutes long). Moreover, to encourage respondents to complete the questionnaire, we provided them with 1000FCFA of air time at the end of the interview. This was announced just after the consent to answer our survey was obtained.

3.2 Sample selection and randomization

We select our sample from a national household survey, the harmonized survey on households' living standards (*Enquête harmonisée sur les conditions de vie des ménages*, hereafter EHCVM), conducted in 2018-2019 by the Malian Institute of Statistics (INSTAT).[9] This survey is representative at national, regional and urban/rural levels.[10] It provides contact information of the surveyed households in the form of a mobile phone number of the household head and another household member (in general, the head's spouse). Contact information is available for 5,835 households out of 6,602 and for 6,415 individuals out of 8000. Our sample is representative

[7] So-called ceiling effects appear when control items are very likely to be answered positively. On the contrary, floor effects correspond to the case where control items are very likely to be answered negatively.

[8] We could not implement a Double List Experiment, as recommended by Chuang et al. (2021) due to limited sample size.

[9] The sampling and survey protocol are further discussed in Bertelli et al. (2023).

[10] The EHCVM households sample is drawn from a traditional two-way stratified random sampling design. First, enumeration areas are randomly drawn with selection proportional to regions' population size, taking into account urban/rural distribution.Second, 6,602 households are then drawn from the sample of enumeration areas and were interviewed in one of the two rounds of data collection (between October and December 2018 and between April and June 2019).

of households in Mali owning a mobile phone at national and regional levels (see Figure 1a for the survey geographic coverage). Note that 89.9% of households own a mobile phone in Mali in 2018 (INSTAT, 2019).

Among those reporting a telephone number, we randomly selected 2,000 individuals, one per household, aged 18 or more. Our selected sample has the same regional share and gender share as the original one made of individuals with phone numbers.

We assigned individuals to the treatment group and to the control group through a randomization process, stratified by region. Among the 1,000 men initially selected, 500 were assigned to the treatment group and 500 to the control group. The same was done separately for women. We divided the sample into a main list (1,400 individuals) and a replacement list (600 individuals) to address issues of erroneous phone number, absence of consent or ineligibility of the respondent.[11] In the end, interviewers tried to contact 1,719 individuals (874 in the control group, 845 in the treated group), and among them 1,509 were reached by phone and gave consent to answer the questionnaire (764 in the control group, 745 in the treated group). Section 3.3 further discusses the issue of non-response.

The survey was conducted by Malian enumerators hired by the *Groupe de Recherche en Economie Théorique et Appliquée* (GREAT) and the authors animated the training sessions. Data collection was done using tablets in July and August 2021 and was administered by phone due to the sanitary and security context at the time of our experiment. Twelve enumerators (men and women in equal proportion) were selected according to their spoken language(s), in order to conduct interviews in the respondents' preferred language. Each enumerator conducted both LE and DQ surveys, following a random assignment.

3.3 Randomization balance check and non-response

The test of randomization balance is conducted using the (few) respondents' characteristics we collected with our survey: age, gender, primary school completion, marital status, ethnic group, migration status, region of residence and exposure to conflict violence (measured by ACLED).[12] Table 1 shows that respondents are aged 42 on average. Around 40% of them completed primary education, most of them (88%) are married, half of them speak Bambara as their mother-tongue, and roughly a third was not born in the current place of residence. Most of the respondents (86%) were, however, living in the current place of residence before the start of the conflict in 2012, suggesting that change of residence was mostly unrelated to the start of the conflict. Moreover, two-thirds of respondents live in communes having witnessed at least one violent conflict-related event since 2012, for an average number of 88 events between 2012 and 2021, as measured with the ACLED data. Most often, governmental forces were involved in these events, whereas the involvement of foreign armed forces is rarer.

[11] For more details on the replacement protocol, see Appendix B.

[12] Since we interviewed adults answering the telephone, regardless of whether they were part of the original EHCVM sample, it is not possible to exploit information from the EHCVM survey.

The results of the balance check and of the joint F-test on the set of observable characteristics show that treatment and control groups are statistically similar. Nevertheless, we observe two main significant differences. The share of respondents whose mother tongue is Bambara is 3.9 percentage points higher in the treatment group, while the share of respondents living in communes with at least one violent conflict-related event (or with at least one violent event in which governmental forces were involved) since 2012 is lower in the treatment group. In Table E1 in Appendix E (Panel 1) we show that our benchmark results are robust to controlling for this set of covariates.

Table 1: Randomization balance check

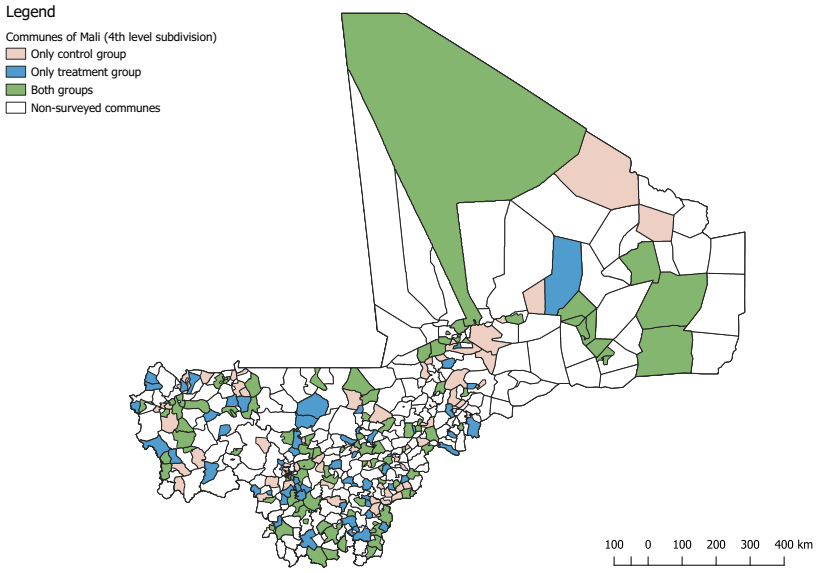
	Control			Treatment			
	n	mean	sd	n	mean	sd	Diff
Age	764	42.00	13.72	745	42.53	14.34	0.505
Primary education completed	764	0.41	0.49	745	0.39	0.49	-0.018
Married	764	0.88	0.33	745	0.88	0.33	-0.003
Mother tongue Bambara	764	0.47	0.50	745	0.51	0.50	0.039*
Mother tongue Sonrhai	764	0.19	0.40	745	0.18	0.39	-0.007
Mother tongue Other	764	0.15	0.35	745	0.13	0.34	-0.013
First year in place of residence	764	1986.83	18.92	745	1987.60	19.42	0.278
Migrant	764	0.30	0.46	745	0.35	0.48	0.020
Lived in current residence place before 2012	764	0.87	0.33	745	0.86	0.35	-0.002
Violent event	764	0.74	0.44	745	0.69	0.46	-0.043**
N. violent events	764	88.32	126.85	745	88.94	128.04	-0.973
Govt. involved in violent events	764	0.68	0.47	745	0.62	0.49	-0.050**
Foreign A.F. involved in violent events	764	0.39	0.49	745	0.38	0.49	-0.013
Joint F-test							1.11
p-value							0.346

Notes: Control variables include regions fixed-effects, respondent's gender and enumerator fixed-effects. "Violent event" is a dummy equal to one if the respondent lives in a commune exposed at least once to a conflict-related violent event since 2012 (ACLED); "N. violent events" is the number of those events (ACLED); "Govt. involved in violent events" is a dummy equal to one if respondent lives in a commune where the Government of Mali was involved at least once in a conflict-related violent event and zero otherwise (ACLED). "Foreign A.F. involved in violent events" is a dummy equal to one if respondent lives in a commune where the foreign Armed Forces were involved at least once in a conflict-related violent event and zero otherwise (ACLED).

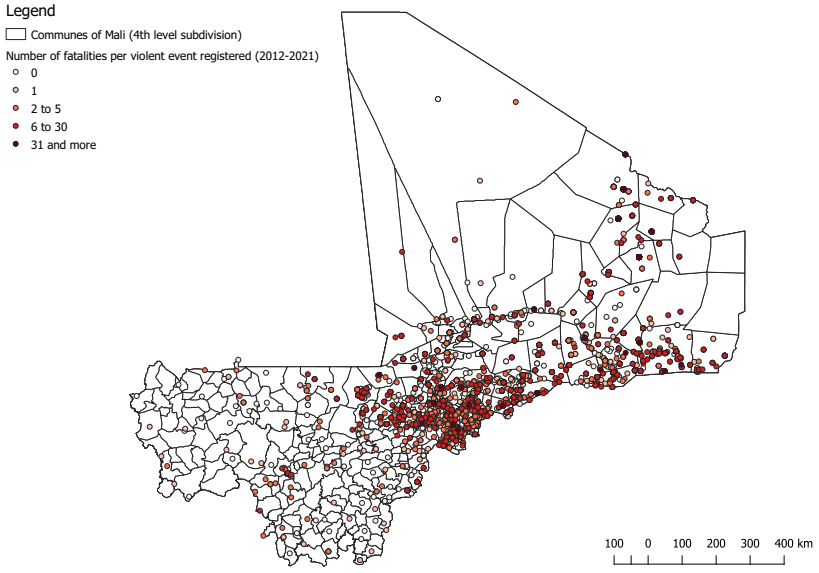
The non-response rate, defined as the ratio between the number of respondents who were not reached or did not consent to participate to the survey, over the number of individuals that we tried to reach, is equal to 12.2% (11.8% in the treated group and 12.6% in the control group).[13] The main source of non-response is the inability to reach people (see Table B1 in Appendix B). Note that all those who were not reached by phone or who did not consent to the survey were replaced by another person living in the same region. We describe the replacement protocol and the set of non-respondents, and test whether non-response is correlated with individual characteristics in Appendix B.

[13] Respondents who refused to answer to one (or more) question(s) and went directly to the next question, but eventually completed the questionnaire, are here ignored and will be dealt with later in the analysis. There are no instances of interrupted questionnaires (beyond the consent question) in our sample.

Figure 1: Survey coverage and commune exposition to violent events since 2012



(a) Survey coverage per treatment status



(b) Violent events registered since 2012

Notes: Map (a) represents communes where individuals of the sample live, depending if they includes respondents belonging to the control, the treatment group or both. Map (b) shows communes and the number per commune of fatalities per violent event registreted (source ACLED, authors' computation).

3.4 Violence data

To explore heterogeneity in response biases according to conflict-related violence exposure, we use the Armed Conflict and Location and Event Data (ACLED).[14] We merge ACLED measures of violence computed at commune level (equivalent to administration level 4) with our survey data based on respondents' communes of residence.

[14] For additional details on ACLED, see Appendix C.

Conflict-related violence concerns half (119) of the 234 communes covered by our survey (see Figure 1). All surveyed communes located in Northern Mali (regions of Gao, Kidal and Tombouctou), as well as those in the regions of Mopti and Bamako, were exposed to violence at least once. Violent events kept intensifying over time but remained concentrated in the centre and the north of the country (see Figure 1b). The distribution of violent events in communes exposed at least once since 2012 is highly right-skewed both at the extensive and intensive margins. The median number of violent events and associated fatalities by exposed commune is 3 and 5, respectively.

4 Empirical strategy

4.1 Specifications

Step 1. Detecting response bias. The first step of our empirical analysis consists in measuring response bias affecting our five insecurity-related outcomes, by comparing the prevalence rates computed under the LE-technique with those computed under the DQ-technique.

The prevalence rate of sensitive item k (for $k \in \{1;2;3;4;5\}$) from the DQ, $\hat{g}_k$ is obtained from a regression of the direct answer on a constant:

$$Z_{i,k} = g_k + \epsilon_{i,k}, \tag{1}$$

where $Z_{i,k}$ equals 1 if individual i answers 'Yes' to sensitive question k, and 0 otherwise.

The prevalence rate of sensitive item k from the LE, $\hat{\gamma}_k$, is measured based on the comparison of the control and treatment groups. In particular, we estimate the following equation:

$$Y_{i,k} = \alpha_k + \gamma_k T_i + \lambda_k W_i^{'} + \epsilon_{i,k} \tag{2}$$

where $Y_{i,k}$ denotes the response given by individual i to the list of statements including item k ($Y_{i,k}$ ranges from 0 to 3 for respondents in the control group and from 0 to 4 for respondents in the treatment group), T_i is a dummy variable that takes the value of 1 if i is in the treatment group and 0 if i is in the control group, and $\epsilon_{i,k}$ is the error term robust to heteroskedasticity. Moreover, we introduce a vector of control variables W_i that includes the randomization strata, namely gender and the region of residence, as well as enumerator fixed effects to account for potential systematic differences in the quality of the data collected.[15]

Given the proposed notations, the first empirical step is to compare $\hat{g}_k$ and $\hat{\gamma}_k$, as estimated by Equations 1 and 2, for all five k sensitive items.

Step 2. Detecting heterogeneity in response bias. The second step of our empirical analysis explores whether response biases are heterogeneous across sub-groups of the population. Response biases, due to either social desirability and/or fear of lack of confidentiality,

[15] Results are robust to excluding the full set of control variables, as shown in Table E1 in Appendix E.

might vary across individuals, depending on their gender, education and violence exposure. Social desirability bias pushes respondents to declare what they believe is expected from them. For instance, depending on the individual's perception of the norm, men might feel compelled to over-report weapons' possession, while respondents with higher education might prefer to under-report it. Individual characteristics may determine with which norm respondents identify themselves, how much they stick to it and the feeling of shame they associate with deviating from it. Similarly, the fear of lack of confidentiality may vary across individual characteristics. Gender, education and violence exposure may affect respondents' trust in the survey firm, willingness to keep their personal information private, or capacity to anticipate possible consequences of disclosing their answers to third parties, in particular in terms of future victimization and possible reprisal.[16]

Moreover, list experiments results were found to be inconsistent among less educated respondents (Kramon & Weghorst, 2019). This is mostly due to the cognitive load required by list experiments, which may lead people with cognitive difficulties or low levels of education to give a random answer to the enumerator. We tried to limit this by guiding the respondents first through an example of a list experiment, the answer to which was explicitly verified by the enumerator. Verifying that response biases are not solely driven by uneducated respondents is, hence, a useful indirect quality check.

We follow Blair & Imai (2012) to test whether the response bias (still measured by comparing prevalence rates across the two survey methods, LE and DQ), varies across individual characteristics. The DQ prevalence rate of each of the five sensitive questions is estimated with the following equation:

$$Z_{i,k} = g_k + \mu_k X_i + \epsilon_{i,k} \tag{3}$$

where $Z_{i,k}$ denotes the answer given by individual i to sensitive question k and X_i is one of the three dummies for gender, primary education, and violence exposure.[17] The DQ prevalence rate of sensitive question k is then given by $\hat{g}_k$ for the omitted category of variable X_i, and by $\hat{g}_k + \hat{\mu}_k$ for the category under study.

The LE prevalence rate of each of the five sensitive questions is estimated with the following equation:

$$Y_{i,k} = \alpha_k + \beta_k X_i + \gamma_k T_i + \delta_k X_i \times T_i + \lambda_k W_i^{'} + \epsilon_{i,k} \tag{4}$$

where we interact the treatment status T_i with one of the three dummies for gender, primary education, and violence exposure. The estimated term $\hat{\gamma_k} + \hat{\delta_k}$ captures the prevalence rate as measured by LE among individuals with $X_i = 1$ while $\hat{\gamma_k}$ measures the prevalence for those with $X_i = 0$. As before, W_i is the vector of control variables at the individual level.

Given the proposed notations, the second empirical step consists in comparing $\hat{g}_k$ and $\hat{\gamma}_k$ on

[16] Existing literature finds heterogeneous response biases depending on education (Agüero & Frisancho, 2022; Joseph et al., 2017), age (Joseph et al., 2017), exposure to UN peacekeeping missions (Blair et al., 2014), being at higher risk of HIV infection (Treibich & Lépine, 2019), gender, marital status, urban/rural area and occupation (Koehler et al., 2022).

[17] We measure violence exposure with a dummy that takes the value 1 for individuals living in a commune with at least one violent event since 2012, as documented by the ACLED data, and zero otherwise.

the one hand, and $\hat{g}_k + \hat{\mu}_k$ and $\hat{\gamma}_k + \hat{\delta}_k$ on the other hand, based on the parameters estimated in Equations 3 and 4.

Step 3. Detecting heterogeneity in prevalence rates. Last, the third part of our analysis explores the correlation between individual characteristics and prevalence rates measured with DQ or LE. In fact, individual characteristics might significantly correlate with prevalence rates measured with one technique but not the other, especially if response biases are heterogeneous across sub-groups of the population, as the analysis in Step 2 will show. In other words, the analysis in Step 3 will provide suggestive evidence about possible consequences of heterogeneous response biases in regression analysis when using self-reported data.[18]

To run this analysis we consider the terms estimated separately for each survey technique: $\hat{\mu}_k$ (Eq. 3) for the DQ and $\hat{\delta}_k$ (Eq. 4) for the LE. While $\hat{\mu}_k$ captures the correlation between respondents' characteristics and the DQ-measured prevalence rates, $\hat{\delta}_k$ captures the same correlation but for the LE prevalence rates.

Contrary to the two previous steps, this time we include the full set of controls also in the DQ-based analysis. In previous steps, we were assessing response biases by comparing answers *across techniques* - i.e., for instance, whether support for the military regime is higher under DQ than LE (step 1) or whether women report higher support for the military regime under DQ than under LE (step 2). In step 3, we are first interested in the response rates heterogeneity *within the same technique* - i.e. whether support for the military regime is higher among women than men in the DQ and whether the same pattern is found with the sensitive item is measured with the LE.

Caution should be applied when interpreting the results from Step 3 for two main reasons. First, omitted variables or other sources of bias may be driving the correlations. Second, the LE technique has the main drawback of being much less precise than the DQ technique. Hence, standard errors are larger and rejecting the null hypothesis of no correlation is harder.

4.2 Design, ceiling and floor effects

The validity of a list experiment is based on the absence of any design effect, meaning that adding the sensitive item does not change the sum of affirmative answers to the baseline items; and on the assumption of no liars, meaning that respondents answer the questions honestly.

Following Blair & Imai (2012), let $S_{i,j}(t)$ be a binary variable denoting respondent i's preference for the j^{th} baseline item for *j=1, ..., J* under the treatment status *t = 0, 1*. Similarly, we use $S_{i,J+1}(t)$ to represent respondent i's answer to the sensitive item under the treatment condition. We further define the potential answer respondent i would give under the treatment and control conditions as $Y_i(1) = \sum_{j=1}^{J+1} S_{i,j}(1)$ and $Y_i(0) = \sum_{j=1}^{J} S_{i,j}(0)$. The observed response is given by $Y_i = Y_i(T_i)$, where $Y_i(1)$ takes a non-negative integer not greater than $J + 1$, while the range of $Y_i(0)$ is given by $0, 1, .., J$.

[18] Cullen (2022) finds that several respondent characteristics significantly correlate with IPV when measured with the DQ-technique, but not when measured with the LE-technique.

Given this notation, the no design effect assumption can be written as:

$$\sum_{j=1}^{J} S_{i,j}(0) = \sum_{j=1}^{J} S_{i,j}(1) \quad \text{or equivalently} \quad Y_i(1) = Y_i(0) + S_{i,J+1}(1) \tag{5}$$

The second main assumption behind list experiments is that respondents truthfully answer the survey and do not hide their answer to the sensitive item. This assumption is violated if respondents lie about their true preferences and under-declare the number of items they agree with. As a consequence, the LE prevalence estimates are underestimated.

The ceiling effect takes place if respondents agree with all items, including the sensitive one. As such, their answer will indicate their (positive) response to each single item. Respondents will then give an answer $Y_i = J$ rather than $Y_i = J + 1$, hiding their true affirmative preference for the sensitive item. In this case, they agree with all the list's items, baseline and sensitive, but prefer to under-declare the number of items.

Some other respondents might agree only with the sensitive item, but not with the baseline items. If they fear that their true preference for the sensitive item would be revealed if they report $Y_i = 1$, then they would rather report $Y_i = 0$. This is the problem of a "floor effect", that would translate in an abnormal high share of $Y_i = 0$ answers. The floor effect may apply to all our items' lists, except list 4, where the sensitive answer is to respond "No" to the support for the military regime.

List 4 may be subject to another type of floor effect. Respondents disagreeing with the sensitive item, and with all the baseline ones, would like to declare $Y_i = 0$. But if they fear that reporting $Y_i = 0$ would reveal their true preference about their support for the regime, then they would rather report $Y_i = 1$, resulting in an abnormal high share of $Y_i = 1$ answers.

Under the two assumptions of no design effect and true responses, the estimate of γ_k in Equation 2 is unbiased and we can identify the joint distribution of $(Y_i(0), S^*_{i,J+1})$, where $S^*_{i,j}$ denotes respondent's i truthful answer to the j^{th} item.

Looking at the proportions of each response across treatment groups is particularly useful as it allows to gauge the plausibility of the assumptions of no design, ceiling or floor effects. To do so, we follow Blair & Imai (2012) and Glynn (2013) and run two empirical exercises. First, we look at the shares of respondents agreeing with j items (with j=0, 1, 2, 3, 4) for each list. Figure 2 provides a graphic representation of these shares for each list, for both the treatment and the control groups (see Table D2 in Appendix D for the exact proportions). We are particularly interested in the distributions' tails as we want to avoid the two extreme cases of a large share of respondents disagreeing with all the items or agreeing with all of them. The share of respondents in the control group who do not conform with any item (zero item) or who conform with all items (three items) is always substantially lower than the share reporting one or two items and fairly similar across lists. In particular, the share of respondents in the control group not reporting any item or agreeing with all of them is quite small (around 10%)

and in line with that of other studies, suggesting the absence of any ceiling or floor effects. Only for the second list we observe a smaller share of zero items (3.93%) and a larger share of three items (21.3%) in the control group, suggesting a possible ceiling effect and, hence, a risk of underestimating the sensitive item.

A special case concerns list 4. Respondents who are not in favor of a military government and who disagree with all baseline items would declare zero. However, they may answer $Y_i = 1$ instead of $Y_i = 0$ if they fear a lack of confidentiality. A large share of treated individuals reporting $Y_i = 1$ for list 4, relative to treated individuals reporting $Y_i = 1$ for the other lists, could then signal the presence of a floor effect. Reassuringly, the share of respondents in the treatment group who report only one item is very similar across lists and not larger in list 4 (Figure 2 and Table D2 in Appendix D).

Second, to assuage the concern of a possible design effect, we compare the proportions of respondents in the treatment and control groups reporting at least J ($J = 1, 2, 3$) items. The List Experiment relies on the assumption of a normal distribution of the baseline items in the population and the inclusion of the sensitive item should not affect it. The only difference in the proportions estimated for the two groups should be due to the sensitive item's prevalence. We expect a higher share of respondents declaring at least one/two/three items in the treatment group, given the longer items' list compared to the control group. The difference between the two groups' proportions should, hence, be positive. A negative difference might signal the presence of a design effect, meaning that the answers to the three baseline items are impacted by the inclusion of the sensitive item (Glynn, 2013).

Table D1 in Appendix D shows a negative difference of the proportions only for the first list. The share of respondents in the treatment group reporting zero items is larger than in the control group, which would downward-bias the LE prevalence rates. The p-value of the t-test of the difference of the two proportions is 0.336, meaning that we cannot reject the null hypothesis that this negative difference is due to luck.[19] This is reassuring about the absence of any design effect.

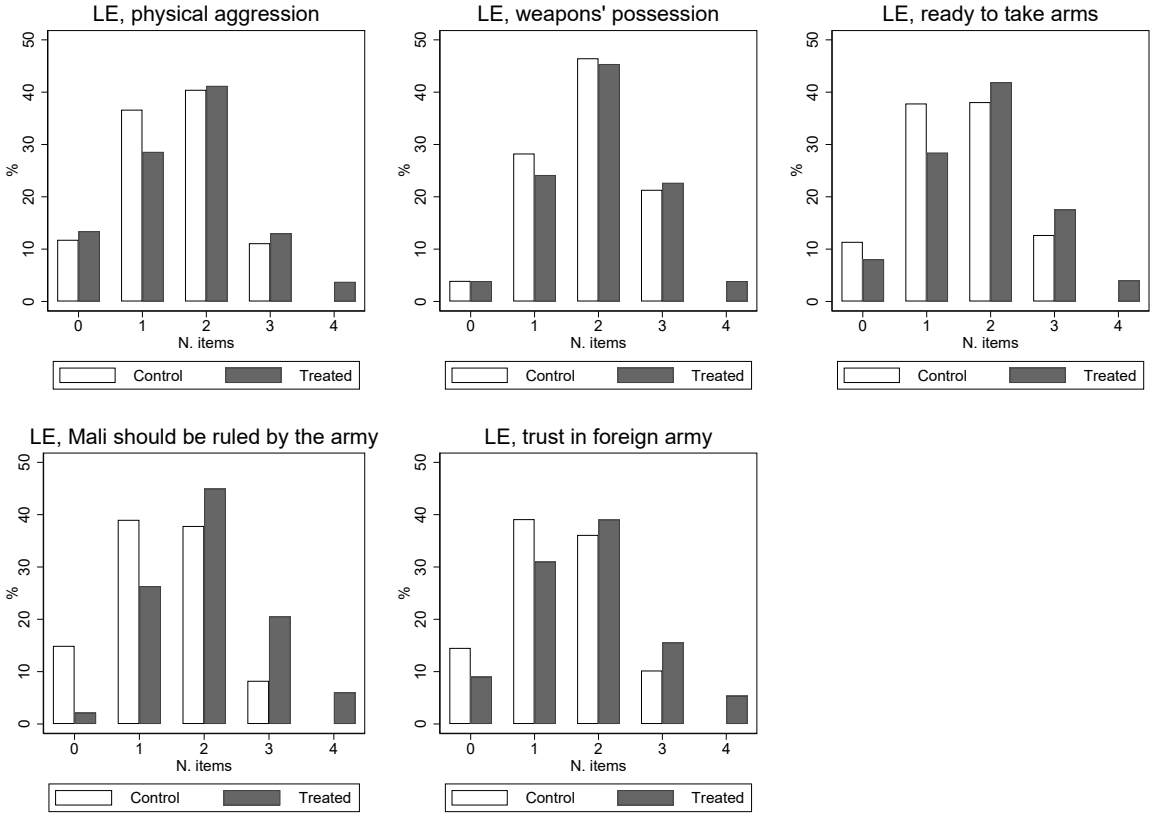
5 Results

5.1 Prevalence and response biases of insecurity-related experiences and preferences

Table 2 and Figure 3 present the average prevalence rates of each sensitive item measured with the two techniques (*i.e.*, $\hat{g}_k$ estimated from Equation 1 and $\hat{\gamma}_k$ estimated from Equation 2), along with the difference between them. As explained above, the information declared under DQ suffers from a reporting bias if the difference between LE and DQ is significantly different from zero.

[19] The Bonferroni-corrected minimum p-values obtained using the R List package designed by Blair & Imai (2012) show the same result.

Figure 2: Average number of items and share of respondents agreeing with j items, by list and treatment status.



We find a significant response bias affecting the two items related to political preferences: support for the army (Q4) and trust in the foreign armed forces (Q5). When asked directly, 74.0% of the respondents support the current military regime and 24.7% trust the foreign armed forces. In contrast, under the LE technique, 63.2% of respondents support the current military regime and 36% trust the foreign armed forces. These differences are statistically significant. They remain significant (at 10% level) when adjusting q-values for false discovery rate following Anderson (2008); Benjamini et al. (2006) approach to account for multiple hypothesis testing.

Respondents over-report their support for the military regime by 10.8 percentage points (pp) when asked directly, while they under-report their trust in foreign armed forces by 11.6pp. The bias size of item 4 corresponds to 14.6% of the DQ prevalence rate, while the bias size of item 5 corresponds to 45.7% of the DQ prevalence prate.

In turn, questions about personal insecurity-related experience do not appear to be affected by response biases in the overall sample. Nevertheless, the prevalence rates are of interest *per se*. Experience of assault outside the home concerns 17% of respondents (or members of their families). Prevalence is larger under the LE technique than under the DQ technique (12.2%), though the difference is not statistically significantly. Similarly, the prevalence of firearms possession is slightly larger when measured with the LE (16.9%), than with the DQ

Table 2: Differences in prevalence rates, whole sample

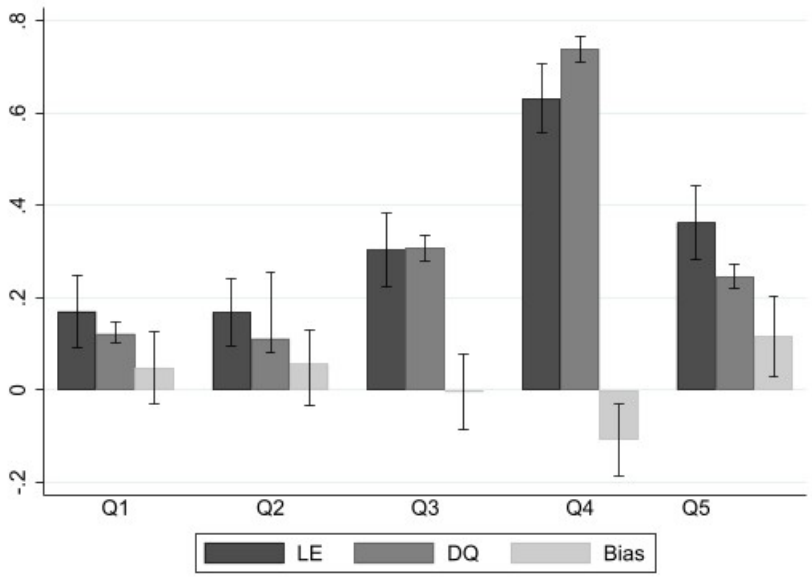
	Q1 Physical assault	**Q2** Firearm possession	**Q3** Ready to take up arms	**Q4** Support for military regime	**Q5** Trust in foreign A.F.
LE	0.171***	0.169***	0.304***	0.632***	0.363***
	(0.047)	(0.044)	(0.048)	(0.045)	(0.050)
DQ	0.122***	0.112***	0.309***	0.740***	0.247***
	(0.012)	(0.011)	(0.017)	(0.017)	(0.016)
LE − DQ	0.049	0.057	-0.005	-0.108	0.116
χ_2	1.013	1.571	.009	5.031	4.689
p-value	0.314	0.210	0.924	0.025	0.030
Sharpened q-value	0.309	0.266	0.587	0.083	0.083
N LE	1509	1509	1509	1509	1509
N DQ	761	760	739	665	693

Notes: OLS estimations; A.F.: Armed Forced; LE equations include strata dummies (sex, and nine region fixed effects), and enumerator fixed effects as control variables.
Robust standard errors in brackets; *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

technique (11.2%). Finally, almost one third of the sample declares being ready to engage in conflict, under both interview techniques.

It is interesting to note that refusal to answer in DQ is very rare for the first two questions (respectively, 3 and 4 respondents, hence 0.4% and 0.5% of the sample), but more frequent for the next three (respectively, 25, 99 and 71 respondents, or 3.3%, 13% and 9.3% of the sample). Still, since some respondents declined to answer some of the sensitive items in the direct form, we are not able to disentangle the part of the observed measurement bias due to respondents answering untruthfully in DQ from the part due to respondents refusing to answer in DQ. Moreover, refusal to answer in DQ could have incidence on our measure of misreporting if refusals are not randomly distributed. The individual determinants of the refusal to answer in DQ, as well as its possible repercussions on our measure of misreporting, are explored in Appendix F. In particular, we find that a negligible share of the variance in the probability of refusal to answer is explained by observed variables, and that our measure of the response bias is robust to disregarding respondents who refused to answer a direct question from the control group when we compute the LE-based prevalence rate of the corresponding item.

Figure 3: Prevalence rates and reporting bias



5.2 Heterogeneity in misreporting

We now turn to investigating whether response bias is homogeneous or not over the population, depending on respondent's gender, education level and past exposure to violence. The analysis is built on the comparison between the results of the estimation of Equations 3 and 4, as explained in Section 4.1.

We present the results grouped by item and declined by the three heterogeneity dimensions. Table 3 displays the results on the first three items, documenting successively heterogeneity across gender (columns 1, 4 and 7), education (columns 2, 5 and 8) and violence exposure (columns 3, 6 and 9). For the sake of space saving, we report only the difference of the prevalence rates between the two survey techniques. In each column, the first row reports the difference along with its p-value for the sub-group of the omitted category (respectively, females, respondents with no primary education, and respondents not exposed to violence); while the second row reports the difference and its p-value for the reference sub-group (respectively, males, respondents with primary education at least, and exposed respondents). Going back to the notations adopted in Section 4.1, the first line thus reports the difference $\hat{\gamma}_k - \hat{g}_k$, and the second line the difference $(\hat{\gamma}_k + \hat{\delta}_k) - (\hat{g}_k + \hat{\mu}_k)$. Table 4 then displays the results on the last two items about the support for the military regime and trust in the foreign armed forces.[20]

Most of the response biases are not significantly different in the statistical sense across sub-groups of respondents. Figure E1 in Appendix E allows to appreciate this point. Still, the

[20] Tables E2, E3 and E4 in Appendix E display the sub-group-specific prevalence rates and number of observations for both the DQ and the LE technique, for each question. Tables E2 displays results on heterogeneity across gender, Table E3 on the level of education, and Table E4 on violence exposure.

differences in the point estimates are informative, as discussed below.[21]

First, although no significant response bias by sub-group is identified regarding physical assault, it is worth noticing that the bias appears to go in opposite directions for respondents with and without primary education (Table 3, column 2). Uneducated respondents tend to under-report their victimization, while the opposite holds for those who completed at least primary education. We observe similar patterns for respondents' exposure to violent events (column 3), whereby those in conflict-affected communes tend to under-report victimization, while those in non-affected communes over-report it.

Table 3: Response biases by sub-groups, Q1–Q3

	Q1 Physical assault			**Q2** Firearm possession			**Q3** Ready to take up arms		
	Male (1)	Primary (2)	VE (3)	Male (4)	Primary (5)	VE (6)	Male (7)	Primary (8)	VE (9)
0	0.101	0.099	-0.024	0.126	0.010	-0.129	0.030	0.032	-0.146
p-value	(0.134)	(0.102)	(0.788)	(0.046)	(0.856)	(0.135)	(0.668)	(0.604)	(0.122)
1	0.004	-0.027	0.072	-0.003	0.125	0.115	-0.035	-0.066	0.046
p-value	(0.955)	(0.728)	(0.200)	(0.962)	(0.079)	(0.028)	(0.609)	(0.400)	(0.427)

Notes: In each column, the first line reports the difference between the LE-based and the DQ-based prevalence rate for the sub-group of the omitted category ($\hat{\gamma}_k - \hat{g}_k$), while the second line reports the same difference for the non-omitted sub-group ($(\hat{\gamma}_k + \hat{\delta}_k) - (\hat{g}_k + \hat{\mu}_k)$); Primary measures the achievement of the primary level; VE equals one if the respondent lives in a commune with at least one violent event since 2012; LE equations include strata variables (sex and nine region fixed effects), and enumerator fixed effects as control variables.

We find starker differences in response biases across groups with respect to firearms possession (columns 4-6). Women, respondents with at least primary education and those living in conflict-affected areas significantly under-report firearms possession – the bias size ranging between 11.5pp and 12.6pp. On the contrary, very little response bias is found among men and respondents without primary education. Last, those living in commune that were not exposed to violence are characterized by a relatively large, though not statistically significant, over-reporting of firearm possession.

Interestingly enough, response biases related to the willingness to engage in conflict are systematically of opposite directions across the sub-groups of analysis, though never significantly different from zero (columns 7-9). Women, respondents without primary education, and those living in conflict-affected communes appear to under-report their willingness to take up the arms, while men, educated respondents, as well as those who were not exposed to violence tend to over-report it.

As far as preferences are concerned (Table 4), all sub-groups over-declare their support for the military regime when asked directly (columns 1-3). The bias is larger among women

[21] To address the issue of multiple hypothesis testing, we compute p-values controlling for the family-wise error rate, following Romano & Wolf (2005a,b) and using Clarke et al. (2020) *rwolf* Stata command. The response biases remain significant only for the gender dimension. Results for other sub-groups should be interpreted with caution.

(13.4pp) than men (8.4pp and not significant), and among uneducated respondents than among those with primary education or more. In particular, uneducated respondents over-declare their support for the military regime by 18pp (with a prevalence rate of 76% under DQ, as compared to 58% under LE), while the bias is close to zero for respondents who have completed primary education.

Table 4: Response biases by sub-groups, Q4–Q5

	Q4 Support for mil. regime			**Q5** Trust in foreign A.F.		
	Male (1)	Primary (2)	VE (3)	Male (4)	Primary (5)	VE (6)
0	-0.134	-0.180	-0.122	0.054	0.105	0.120
p-value	(0.056)	(0.002)	(0.203)	(0.494)	(0.125)	(0.246)
1	-0.084	-0.004	-0.105	0.169	0.131	0.120
p-value	(0.197)	(0.964)	(0.057)	(0.018)	(0.108)	(0.051)

Notes: A.F.: Armed Forced; in each column, the first line reports the difference between the LE-based and the DQ-based prevalence rate for the sub-group of the omitted category ($\hat{\gamma}_k - \hat{g}_k$), while the second line reports the same difference for the non omitted sub-groups ($(\hat{\gamma}_k + \hat{\delta}_k) - (\hat{g}_k + \hat{\mu}_k)$); Primary measures the achievement of the primary level; VE equals one if the respondent lives in a commune with at least one violent event since 2012; LE equations include strata variables (sex and nine region fixed effects), and enumerator fixed effects as control variables.

Conversely, all sub-groups under-report their trust in the foreign armed forces, but some do more than others (columns 4-6). In particular, response bias is larger for men (16.9pp) than for women (5pp and not significant). When asked directly, 30% of women and 20% of men declare trusting the foreign armed forces, whereas 35% of women and 37% of men reveal their trust in the foreign armed forces in the LE (see Table E2 in Appendix E).

On the other hand, it appears that there is no difference in response bias about the support for the military regime or trust in foreign armed forces depending on violence exposure (columns 3 and 6). We do find a significant response bias for the two items among those who live in conflict-affected communes, while the bias is not statistically significant for those who live in conflict-free communes, but this is likely due to sample size differences – the number of respondents in exposed areas being twice as large as the number of respondents in areas with no conflict event.[22]

We investigate whether the latter results differ according to the type of actors involved in violent events, namely the government or foreign armed forces (Tables 5 and E5 in Appendix E). For this purpose, we distinguish individuals living in communes where the government was involved in at least one violent event since 2012, to individuals living in communes where foreign armed forces were involved in at least one conflict event since 2012. For both categories, respondents living in conflict-free zones constitute the omitted category. Results in Table 5 show

[22] In the treatment group, out of 1,509 respondents, 1,084 live in areas exposed to violence. In the control group, out of 665 respondents, 484 live in exposed areas (see Table E4 in Appendix E).

that response biases of respondents living in communes with at least one violent event since 2012 (Columns 1 and 4) are similar to those of respondents who experienced violent events involving the government forces (Columns 2 and 5). In both cases, respondents over-report their support for the military regime and under-report their trust in the foreign armed forces by, respectively, around 10pp and 12pp. In turn, results are different for respondents exposed to violent events involving foreign armed forces. In particular, the response bias characterizing trust in the foreign armed forces is 8pp larger than the bias measured for those living in areas exposed to violent events in general. DQ respondents living in areas where foreign armed forces intervened under-declare their trust in foreign armed forces by 20pp. In these areas, respondents are, in fact, almost twice as much trustful in foreign armed forces as they declare they are under DQ (see Table E5 in Appendix E). This results corroborates the sensitivity of such matter in conflict-affected areas in Mali. In turn, the response bias regarding the support for the military regime is not significantly different from zero.

Table 5: Reporting biases by actors involved in VE, Q4–Q5

	Q4 Support for mil. regime			**Q5** Trust in foreign A.F.		
	VE (1)	VE Govt (2)	VE foreign A.F. (3)	VE (4)	VE Govt (5)	VE foreign A.F. (6)
0	-0.122	-0.122	-0.128	0.120	0.121	0.136
p-value	(0.203)	(0.201)	(0.184)	(0.246)	(0.246)	(0.192)
1	-0.105	-0.107	-0.027	0.120	0.128	0.203
p-value	(0.057)	(0.062)	(0.724)	(0.051)	(0.048)	(0.016)

Notes: A.F.: Armed Forced; In each column, the first line reports the difference between the LE-based and the DQ-based prevalence rate for the sub-group of the omitted category, defined below $(\hat{\gamma}_k - \hat{g}_k)$; while the second line reports the same difference for the non omitted sub-groups $((\hat{\gamma}_k + \hat{\delta}_k) - (\hat{g}_k + \hat{\mu}_k))$; VE equals 1 if the respondent lives in a commune with at least one violent event that has occurred since 2012; 0 otherwise; VE Govt equals 1 for respondents in communes where the government was involved in at least one violent event since 2012; whereas VE foreign A.F. equals 1 if foreign armed forces were involved; in both cases the omitted category is no violent event; LE equations include strata variables (sex and region fixed effects), and enumerator fixed effects as control variables.

Overall, these results reveal that response bias is not uniformly distributed across the population and show important heterogeneity across respondents' gender, level of education, and past violence exposure. Importantly, we find a widespread fear of revealing one's own true political preferences. All sub-groups of analysis tend to over-report (respectively, under-report) their support for the Malian army and their trust in the foreign armed forces, even though in different proportions. Moreover, insecurity experience-related items appear to be misreported by some segments of the population. This is possibly explained by differences in the norms to which different groups relate, and by differences in the level of fear of a lack of confidentiality. For instance, men and women seem to relate to diverging norms about the acceptability of firearm possession.

These results can have important implications for data collection and policy evaluation.

Administrating surveys with the DQ-technique in conflict-affected zones might significantly mis-measure the insecurity-related experiences and preferences of the local population. As a consequence, interventions aiming at improving the local security conditions might yield misleading results if the outcomes are measured with the wrong survey method. Misreporting seem to be particularly common among women, uneducated respondents and those exposed to violent events, plausibly the most vulnerable segments of the population. The over-stated support for the military regime might, indeed, be due to a fear of retaliation or of lack of confidentiality. This result suggests that development programs working with these segments of vulnerable population should pay special attention to grant their freedom of speech.

5.3 Characteristics associated with insecurity-related experiences and preferences

The last part of the empirical analysis investigates the correlation between respondents' gender, education or violence exposure and insecurity-related outcomes, depending on the survey technique used. We report the results in Table 6. Columns 1, 3 and 5 display $\hat{\mu_k}$ (with the associated p-value) from Equation 3, which reflects the estimated correlates of the DQ-based outcomes, and columns 2, 4 and 6 display $\hat{\delta_k}$ (with the associated p-value) from Equation 4, which reflects the estimated correlates of the LE-based outcomes.

As expected, we observe important differences when estimating the (conditional) correlations between respondent characteristics and insecurity-related experiences and preferences as declared through direct questioning or as revealed by the list experiment. Most often, respondents' characteristics (in particular, gender and education) significantly correlate with prevalence rates when measured with the DQ technique, but much more rarely when measured with the LE technique. This could be partly explained by the fact that the LE limits response bias at the price of a higher variation and a lower precision. Still, in several cases, we find correlation coefficients of opposite signs across the survey techniques, suggesting that this divergence is not only due to a lack of precision in the LE but that it might also reflect correlations driven by the response biases in the DQ. However, let us emphasize that in most cases, the correlations are not significantly different across survey techniques, hence the results should be interpreted with caution.

Table 6: Correlation of respondent's characteristics with sensitive items' prevalence rates measured under DQ *vs.* LE.

	Male			**Primary**			**Violence Exp.**		
	DQ	LE	Diff. (p-value)	DQ	LE	Diff. (p-value)	DQ	LE	Diff. (p-value)
Item:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Physical assault	0.02	-0.07		0.06	-0.07		0.03	0.13	
p-value	(0.40)	(0.43)	0.33	(0.03)	(0.48)	0.21	(0.27)	(0.20)	0.37
Obs.	761	1509		761	1509		761	1509	
Firearm possession	0.07	-0.06		0.01	0.12		-0.04	0.15	
p-value	(0.00)	(0.51)	0.16	(0.68)	(0.18)	0.24	(0.23)	(0.11)	0.05
Obs.	760	1509		760	1509		760	1509	
Ready to take up arms	0.06	0.01		0.11	-0.00		0.00	0.21	
p-value	(0.07)	(0.90)	0.59	(0.00)	(1.00)	0.26	(0.94)	(0.04)	0.07
Obs.	739	1509		739	1509		739	1509	
Support for military regime	0.04	0.08		-0.05	0.12		0.06	0.02	
p-value	(0.26)	(0.39)	0.70	(0.19)	(0.17)	0.08	(0.19)	(0.83)	0.74
Obs.	665	1509		665	1509		665	1509	
Trust in foreign A.F.	-0.11	0.01		0.03	0.06		0.07	0.02	
p-value	(0.00)	(0.88)	0.23	(0.37)	(0.55)	0.79	(0.13)	(0.83)	0.69
Obs.	693	1509		693	1509		693	1509	

OLS estimations. All coefficients refer to a different estimation. All estimations control for region fixed-effects and enumerator fixed-effects.

Col. 1–4–7: Dep. var. = 1 if answer = 'Yes'. Display the coefficients associated with, respectively, *Male*, *Primary* and *Violence exposure*.

Col. 2–5–8: Dep. var. = number of 'Yes' to each list. Display the coefficients associated with, respectively, *Treated* × *Male*, *Treated* × *Primary* and *Treated* × *Violence exposure*. The three columns also control for *Male* and *Treated*, Col. 5 for *Primary* and Col. 8 for *Violence exposure*.

Starting with physical assault, it appears that education is positively correlated with victimization under DQ (column 4), but not under LE (column 5). This can be explained by the fact that educated respondents tend to slightly over-report physical assault experience, while uneducated respondents tend to under-report it (see Table 3). The two correlation coefficients are, however, not significantly different from each other.

Second, we uncover a positive correlation between the male dummy and firearm possession measured under DQ (column 1), but not under LE (column 2). This is due to the fact that women under-report this item, while men do not, as shown in Table 3. Moreover, although not significantly different from zero, the correlation between education and firearm possession measured under LE (column 5) is much larger than its counterpart under DQ (column 4), which echoes the fact that educated respondents under-report firearm possession while uneducated respondents do not. Finally, violence exposure correlates nearly significantly with weapons possession when measured under LE (column 8) but not under DQ (column 7), the two correlation coefficients being, in this case, significantly different from each other. This can be related to the fact that respondents in conflict zones under-report firearm possession, while respondents in conflict-free zones over-report it, as shown in Table 3.

Turning to the willingness to engage in conflict, results show that men are more likely to report it when directly asked (column 1), but not in the list experiment (column 2). This is in line with the fact that men slightly over-report it while women slightly under-report it in

DQ, as shown in Table 3. The two response biases are not significantly different from zero, but the asymmetry between them likely explains the observed difference between the coefficients reported in Table 6 that is, however, no significantly different from zero. In a similar fashion, education is positively correlated with the willingness to engage in conflict under DQ (column 4), but not under LE (column 5) – a result likely explained by the fact that educated respondents tend to slightly over-report their willingness to take up arms while uneducated respondents tend to under-report it (see Table 3).

On the contrary, past conflict exposure is positively correlated with the willingness to engage in conflict measured under LE (column 8), but not under DQ (column 7). Importantly, this difference is statistically significantly different from zero. This is in line with the results in Table 3 showing that respondents in conflict-affected zones under-report their willingness to take up arms (though not significantly), while respondents in conflict-free zones over-report it (almost significantly).

As shown in Table 4, our two preference-related items are less subject to heterogeneity in misreporting insofar as all sub-groups tend to over-report their support for the military regime and to under-report their trust in the foreign armed forces. Yet, since they do so in different proportions, misreporting can still affect the analysis of the individual characteristics that correlate with preferences. In particular, Table 6 shows that the male dummy is significantly and negatively correlated with trust in the foreign armed forces under DQ but not under LE. This discrepancy mirrors the fact that, although both men and women under-report their trust in the foreign armed forces as compared to what the LE reveals, men do so in a much larger proportion than women.

Taken together, the results from Table 6 suggest that asymmetries in reporting biases between sub-groups (for instance, between men and women, or between educated and uneducated persons) can have important consequences when estimating the correlations between individual characteristics and sensitive experiences or preferences. This may have important implications with respect to the choice of the survey method for measuring sensitive traits, as it can yield biased descriptive statistics and causal regression results.

Conclusion

Existing literature shows that directly asking respondents about sensitive topics can lead to misreporting, due to desirability bias and the fear of limited confidentiality. Reporting biases are likely to be even larger in fragile States, as many developing countries are. Widespread insecurity, conflict and/or the limited freedom of speech might affect respondents' trust in the enumerators, biasing their responses to sensitive topics. In addition, an unsafe context may shape the sensitivity of certain topics, in particular those related to the country's political situation, as well as the prevailing norms defining social desirability. All this can have important consequences for data collection and treatment effect estimates.

This is the first study to implement an experimental survey protocol to explore insecurity-related experiences and preferences in a fragile State – Mali. We find substantial average response biases when individuals are directly asked sensitive questions, notably concerning their political preferences. Respondents appear to over-report their support for the military regime, and under-report their trust in the foreign armed forces. We show that response biases are not uniformly distributed across the population and document significant heterogeneity across individual characteristics. Gender, education and conflict exposure drive response bias in different directions.

As a consequence, we provide suggestive evidence that asymmetric misreporting across population sub-groups (e.g. men and women or low-educated and educated individuals) can translate into biased correlations between individual characteristics and sensitive items' prevalence rates when the latter are measured with direct questions. This result is of interest beyond the specific focus of this survey. If individual outcomes are affected by misreporting, studies exploring their determinants could end up with biased correlations or biased heterogeneity treatment effects. Significant results might simply be "statistical artifacts" due to asymmetric response biases across sub-groups. Response bias would hence yield "fake" estimated heterogeneity and biased regression estimates. The targeting of development programs based on beneficiaries' gender or education should be particularly aware of these risks.

References

Agüero, J. M. & Frisancho, V. (2022). Measuring violence against women with experimental methods. *Economic Development and Cultural Change*, 70(4), 1565–1590.

Anderson, M. L. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), 1481–1495.

Asadullah, M. N., De Cao, E., Khatoon, F. Z., & Siddique, Z. (2021). Measuring gender attitudes using list experiments. *Journal of Population Economics*, 34(2), 367–400.

Bellows, J. & Miguel, E. (2009). War and local collective action in sierra leone. *Journal of public Economics*, 93(11-12), 1144–1157.

Benjamini, Y., Krieger, A. M., & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3), 491–507.

Bertelli, O., Calvo, T., Lavallée, E., Mercier, M., & Mesplé-Somps, S. (2023). Expériences et opinions liées à l'insécurité: une analyse des biais déclaratifs au mali. *Revue d'Economie du Développement*.

Bertelli, O. & Kurdi, S. (2022). Inter-group cooperation and conflict: Experimental evidence from yemen. *Mimeo*.

Blair, G., Coppock, A., & Moore, M. (2020). When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments. *American Political Science Review*, 114(4), 1297–1315.

Blair, G. & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis*, 20(1), 47–77.

Blair, G., Imai, K., & Lyall, J. (2014). Comparing and combining list and endorsement experiments: evidence from Afghanistan. *American Journal of Political Science*, 58(4), 1043–1063.

Brück, T., Justino, P., Verwimp, P., Avdeenko, A., & Tedesco, A. (2016). Measuring violent conflict in micro-level surveys: current practices and methodological challenges. *The World Bank Research Observer*, 31(1), 29–58.

Bullock, W., Imai, K., & Shapiro, J. N. (2011). Statistical analysis of endorsement experiments: Measuring support for militant groups in pakistan. *Political Analysis*, 19(4), 363–384.

Bulte, E. & Lensink, R. (2019). Women's empowerment and domestic abuse: Experimental evidence from vietnam. *European economic review*, 115, 172–191.

Calvo, T., Lavallée, E., Razafindrakoto, M., & Roubaud, F. (2020). Fear not for man? armed conflict and social capital in mali. *Journal of Comparative Economics*, 48(2), 251–276.

Cassar, A., Grosjean, P., & Whitt, S. (2013). Legacies of violence: trust and market development. *Journal of Economic Growth*, 18(3), 285–318.

Chuang, E., Dupas, P., Huillery, E., & Seban, J. (2021). Sex, lies, and measurement: Consistency tests for indirect response survey methods. *Journal of Development Economics*, 148, 102582.

Clarke, D., Romano, J. P., & Wolf, M. (2020). The romano–wolf multiple-hypothesis correction in stata. *The Stata Journal*, 20(4), 812–843.

Coffman, K. B., Coffman, L. C., & Ericson, K. M. M. (2017). The size of the lgbt population and the magnitude of antigay sentiment are substantially underestimated. *Management Science*, 63(10), 3168–3186.

Cullen, C. (2022). Method matters: The underreporting of intimate partner violence. *The World Bank Economic Review*.

De Cao, E. & Lutz, C. (2018). Sensitive survey questions: Measuring attitudes regarding female genital cutting through a list experiment. *Oxford Bulletin of Economics and Statistics*, 80(5), 871–892.

Fendrich, M. & Vaughn, C. M. (1994). Diminished lifetime substance use over time: An inquiry into differential underreporting. *Public Opinion Quarterly*, 58, 96–123.

Gibson, M. A., Gurmu, E., Cobo, B., Rueda, M. M., & Scott, I. M. (2018). Indirect questioning method reveals hidden support for female genital cutting in south central ethiopia. *PloS one*, 13(5), e0193985.

Gilligan, M. J., Pasquale, B. J., & Samii, C. (2014). Civil war and social cohesion: Lab-in-the-field evidence from nepal. *American Journal of Political Science*, 58(3), 604–619.

Glynn, A. N. (2013). What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly*, 77(S1), 159–172.

Gong, E. (2015). Hiv testing and risky sexual behaviour. *The Economic Journal*, 125(582), 32–60.

Guichaoua, Y. & Pellerin, M. (2017). *Faire la paix et construire l'Etat : Les relations entre pouvoir central et périphéries sahéliennes au Niger et au Mali*. Technical Report 51, IRSEM.

Haidara, B. S. (2020). Pourquoi l'opinion publique malienne a une vision négative de l'opération barkhane. *The Conversation*, 10/02/2020.

Haidara, B. S. (2021). L'opération barkhane, une déception vue du mali. *La Croix*, 13/01/2021.

Holbrook, A. L. & Krosnick, J. A. (2009). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, 74(1), 37–67.

INSTAT (2019). *Consommation, pauvreté, bien-être des ménages, Avril 2018 Mars 2019*. Technical report, Institut National de la Statistique du Mali.

Isaqzadeh, M., Gulzar, S., & Shapiro, J. (2020). *Studying Sensitive Topics in Fragile Contexts*, (pp. 173–192). Springer International Publishing: Cham.

Jamison, J. C., Karlan, D., & Raffler, P. (2013). *Mixed method evaluation of a passive mHealth sexual information texting service in Uganda*. Technical report, National Bureau of Economic Research.

Joseph, G., Javaid, S. U., Andres, L. A., Chellaraj, G., Solotaroff, J. L., & Rajan, S. I. (2017). Underreporting of Gender-Based Violence in Kerala, India. *World Bank Policy Research Working Paper*, (8044).

Julien, S. (2011). Le Sahel comme espace de transit des stupéfiants. Acteurs et conséquences politiques. *Hérodote*, (3), 125–142.

Karlan, D. S. & Zinman, J. (2012). List randomization for sensitive behavior: An application for measuring use of loan proceeds. *Journal of Development Economics*, 98(1), 71–75.

Koehler, K., Grewal, S., & Albrecht, H. (2022). Who fakes support for the military? experimental evidence from tunisia. *Democratization*, 29(6), 1055–1076.

Kraay, A. & Murrell, P. (2016). Misunderestimating corruption. *The Review of Economics and Statistics*, 98(3), 455–466.

Kramon, E. & Weghorst, K. (2019). (mis) measuring sensitive attitudes with the list experiment: Solutions to list experiment breakdown in kenya. *Public Opinion Quarterly*, 83(S1), 236–263.

Lépine, A., Treibich, C., & d'Exelle, B. (2020). Nothing but the truth: Consistency and efficiency of the list experiment method for the measurement of sensitive health behaviours. *Social Science & Medicine*, 266, 113326.

Lyall, J., Blair, G., & Imai, K. (2013). Explaining support for combatants during wartime: A survey experiment in afghanistan. *American political science review*, 107(4), 679–705.

Minnis, A. M., Steiner, M. J., Gallo, M. F., Warner, L., Hobbs, M. M., van der Straten, A., Chipato, T., Macaluso, M., & Padian, N. S. (2009). Biomarker validation of reports of recent sexual activity: results of a randomized controlled study in zimbabwe. *American Journal of Epidemiology*, 170(7), 918–24.

Nordås, R., Østby, G., Amisi, C., Biaba, R., Quattrocchi, J., Alldén, S., & Cikara, A. (2016a). Social inclusion of survivors of sexual violence: How do support programs work?

Nordås, R., Østby, G., Biaba, R., Amisi, C., Quattrochi, J., Alldén, S., Namegabe, E., & Cikara, A. (2016b). From pain to empowerment?: Impact of an empowerment program for survivors of sexual violence in eastern democratic republic of the congo (drc) on attitudes towards female leadership.

Peterman, A., Palermo, T. M., Handa, S., Seidenfeld, D., & Team, Z. C. G. P. E. (2018). List randomization for soliciting experience of intimate partner violence: application to the evaluation of zambia's unconditional child grant program. *Health economics*, 27(3), 622–628.

Porter, C., Favara, M., Sánchez, A., & Scott, D. (2021). The impact of covid-19 lockdowns on physical domestic violence: Evidence from a list randomization experiment. *SSM-population health*, 14, 100792.

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research*, 47(5), 651–660.

Rohner, D., Thoenig, M., & Zilibotti, F. (2013). Seeds of distrust: Conflict in uganda. *Journal of Economic Growth*, 18(3), 217–252.

Romano, J. P. & Wolf, M. (2005a). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469), 94–108.

Romano, J. P. & Wolf, M. (2005b). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4), 1237–1282.

Rosenfeld, B., Imai, K., & Shapiro, J. N. (2016). An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science*, 60(3), 783–802.

Tourangeau, R. & Yan, T. (2007). Sensitive questions in surveys. *Psychological bulletin*, 133(5), 859.

Traunmüller, R., Kijewski, S., & Freitag, M. (2019). The silent victims of sexual violence during war: Evidence from a list experiment in sri lanka. *Journal of Conflict Resolution*, 63(9), 2015–2042.

Treibich, C. & Lépine, A. (2019). Estimating misreporting in condom use and its determinants among sex workers: Evidence from the list randomisation method. *Health Economics*, 28(1), 144–160.

Voors, M. J., Nillesen, E. E., Verwimp, P., Bulte, E. H., Lensink, R., & Van Soest, D. P. (2012). Violent conflict and behavior: a field experiment in burundi. *American Economic Review*, 102(2), 941–64.

Wood, E. J. (2006). The ethical challenges of field research in conflict zones. *Qualitative sociology*, 29(3), 373–386.

Appendix

A The module on insecurity

Table A1: Lists and questions about insecurity-related experiences and preferences.

	List experiments (LEs)	Direct questioning (DQ)
1.	- Over the last 12 months, I was asked a bribe. - Over the last 12 months, the police was generally efficient in my locality. - Over the last 12 months, the number of road accidents increased in my locality. - *Over the last 12 months, me or a member of my household was physically assaulted outside home.*	Over the last 12 months, were you or a member of your household physically assaulted outside home?
2.	- Lately, local markets have been well supplied. - There are often arguments among the neighborhood residents. - I feel safe at home. - *Me or a member of my household owns a firearm.*	Do you or a member of your household own a firearm?
3.	- Roads are well maintained in my area of residence. - Generally speaking, I trust strangers. - Lately, tensions have calmed down in my area of residence. - *Me or a member of my household is ready to take up arms to defend our community.*	Are you or a member of your household ready to take up arms to defend your community?
4.	- The electrical network is of poor quality. - The political class properly takes care of Malians' problems. - Lately, the public health system has improved. - *Today's Mali should be ruled by the army.*	Do you think that today's Mali should be ruled by the army?
5.	- The police. - The public health system. - The Malian political class. - *The foreign armed forces in Mali.*	Do you trust the foreign armed forces in Mali?

B Non-response and replacement of respondents

All the individuals from the original database that had not been (randomly) selected in our sample constituted a replacement list to pick from in case an individual in the primary one did not respond. All individuals in the replacement list were randomly assigned to either the control or the treatment group – so that replacements were done within the same group – randomization being stratified on gender and regions, as for the primary list. Notice that enumerators were instructed to conduct the interview with the person answering the phone, regardless of whether the person was the initially sampled one or not, but provided that he/she was above eighteen years old. If not, he/she was replaced by an adult member of the same household.

We replaced the initially sampled person with someone from the replacement list coming from the same region and assigned to the same treatment group in the following cases:

(i) if a phone number in the primary list is non-existent,

(ii) if the person from the primary list does not answer the phone, after 3 call backs per day (morning, afternoon, early evening) during three consecutive days, or

(iii) if a person answers the phone but does not consent to participate to the survey

Overall, non-responses were few and mostly due to wrong telephone numbers or failure to answer (Table B1). Reassuringly, treatment assignment is not correlated with the likelihood of non-response (Table B2). In terms of observable characteristics, men are more likely to answer the survey than women, as well as respondents living in the Koulikoro and Bamako regions (as compared to those in Kayes).

Table B1: Non-response: composition and rate

	Treatment group	Control group
# non-responses	100	110
# not reached (erroneous or never-answering phone number)	63	87
# reached, no consent	37	23
Non-response rate	11.8%	12.6%

Our replacement protocol possibly leads to a final sample that is slightly different from the initially selected sample. Respondents might, in fact, be living in a different household and in a different location. Data from the 2018 initial sample will thus not be available in the case of respondents differing from the ones initially selected.

C ACLED data

To avoid using (likely biased) self-reported measures of violence experience and to avoid increasing the salience of the sensitive items, we did not ask respondents about their past

Table B2: Determinants of non-response

Dep. var.:	Non-response (not reached or no consent)		
	(1)	(2)	(3)
Treated	-0.003	0.004	0.018
	(0.02)	(0.02)	(0.01)
Male		-0.167***	-0.172***
		(0.01)	(0.01)
Region: Koulikoro			-0.078**
			(0.04)
Region: Sikasso			0.048
			(0.04)
Region: Segou			-0.044
			(0.04)
Region: Mopti			0.070
			(0.05)
Region: Tombouctou			0.056
			(0.05)
Region: Gao			-0.041
			(0.05)
Region: Kidal			0.073
			(0.06)
Region: Bamako			-0.109***
			(0.04)
Interviewer FE	Yes	Yes	Yes
Observations	1719	1719	1714
R^2	0.074	0.137	0.165
Adjusted R^2	0.068	0.131	0.155

OLS estimations; Region of reference: Kayes;
Standard errors in parentheses; * p<0.10, ** p<0.05, *** p<0.01

exposure to the Malian conflict. Instead, we rely on the Armed Conflict and Location and Event Data (ACLED) to measure exposure to conflict.

ACLED makes a real-time geo-localised census of all political violence and protests reported in the media (Raleigh et al., 2010). It defines political violence as "the use of force by a group with a political purpose or motivation". It compiles data from newspaper articles, press agencies dispatch to non-governmental and international organisations report, and provides the precise date and GPS coordinates of each event. ACLED reports six types of events, namely (i) battles involving two armed groups, (ii) remote violence, (iii) violence against unarmed civilians, (iv) strategic development, (v) protests or demonstrations, and (vi) riots. We limit the analysis to conflict-related violent events, meaning (i) battles involving two armed groups, (ii) remote violence and (iii) violence against unarmed civilians. Overall the ACLED data contains 3,887 geo-located violent events in Mali from 1997 to July 2021, 97% of which are registered after the outburst of violence in 2012.

ACLED identifies the groups involved in each event. Despite the heterogeneity of active armed groups in Mali, they can be broadly divided into eight categories: national security forces (including the Malian national army); foreign and international armed forces (foreign national and international troops with a mandate to intervene in Mali); jihadist armed groups (with

explicit religious claims and terrorist-related activity); Northern rebel groups (with separatist claims); ethnic or communal militias (identified to be representative/protectors of a specified group); other armed groups, unidentified armed groups and civilians. To push further our heterogeneity analysis, we particularly focus on government and foreign armed forces involvement in conflict-related events. Their implication directly relates to sensitive items 4 and 5 (support for military regime to rule today's Mali and trust in foreign armed forces).[23] The response biases could increase if true preferences deviate further from the norm in areas exposed to such interventions. The response biases could decrease if the social norm is closer to true preferences following national and/or foreign armed forces itnervention.

D Exploring possible design effects

Table D1: Share of positive answers by question, number of items and treatment status.

	Q1	Q2	Q3	Q4	Q5
At least one item (T)	86.58	96.11	91.95	97.85	91.01
At least one item (C)	88.22	96.07	88.61	85.08	85.47
Diff. T-C	-1.64	0.03	3.33	12.77	5.54
At least two items (T)	57.99	71.95	63.49	71.54	60.00
At least two items (C)	51.57	67.80	50.79	46.07	46.34
Diff. T-C	6.42	4.15	12.70	25.47	13.66
At least three items (T)	16.78	26.58	21.61	26.58	20.94
At least three items (C)	11.13	21.34	12.70	8.25	10.21
Diff. T-C	5.65	5.24	8.91	18.33	10.73

Table D2: Average number of items and share of respondents agreeing with j items, by list (Q1-Q5) and treatment status.

	Q1	Q2	Q3	Q4	Q5
Zero items (T)	13.42	3.89	8.05	2.15	8.99
Zero items (C)	11.78	3.93	11.39	14.92	14.53
One item (T)	28.59	24.16	28.46	26.31	31.01
One item (C)	36.65	28.27	37.83	39.01	39.14
Two items (T)	41.21	45.37	41.88	44.97	39.06
Two items (C)	40.45	46.47	38.09	37.83	36.13
Three items (T)	13.02	22.68	17.58	20.54	15.57
Three items (C)	11.13	21.34	12.70	8.25	10.21
Four items (T)	3.76	3.89	4.03	6.04	5.37

E Additional results

[23] Government and foreign armed forces got engaged respectively in 25% and 16% of all violent events registered since January 2012. 25 and 85 of the 234 surveyed communes were exposed to violent events involving foreign armed forces and government armed forces (corresponding to 584 and 980 respondents).

Table E1: Robustness to various vectors of control variables

	Q1 Physical assault	**Q2** Firearm possession	**Q3** Ready to take up arms	**Q4** Support for military regime	**Q5** Trust in foreign A.F.
DQ	0.122***	0.112***	0.309***	0.740***	0.247***
	(0.012)	(0.011)	(0.017)	(0.017)	(0.016)
Panel 1: Control for unbalanced characteristics					
LE	0.165***	0.157***	0.299***	0.630***	0.364***
	(0.047)	(0.044)	(0.048)	(0.045)	(0.050)
$LE - DQ$	0.043	0.045	-0.010	-0.109	0.117
p-value	0.378	0.321	0.842	0.023	0.030
Panel 2: With extensive control var. vector					
LE	0.164***	0.155***	0.294***	0.630***	0.362***
	(0.047)	(0.044)	(0.048)	(0.045)	(0.050)
$LE - DQ$	0.042	0.043	-0.015	-0.110	0.115
p-value	0.385	0.346	0.763	0.022	0.032
N LE	1509	1509	1509	1509	1509
N DQ	761	760	739	665	693

Notes: OLS estimations; A.F.: Armed Forced; all LE equations include strata dummies (sex, and nine region fixed effects) and enumerator fixed effects as control variables; in Panel 1 we control for unbalanced characteristics between Treatment and Control groups (a dummy for speaking Bambara as mother tongue and a dummy for past exposure to violent conflict-related events since 2012); Panel 2 adds controls for respondent's age, Songhai as mother tongue, other mother tongue dummy, education (primary level achieved), marital dummy and a migrant dummy.
Robust standard errors in brackets; *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

Table E2: Prevalence rates by sex

	Q1 Physical assault	**Q2** Firearms possession	**Q3** Ready to take up arms	**Q4** support for mil. regime	**Q5** Trust in foreign A.F.
Female *LE* prevalence	0.210**	0.199**	0.298***	0.590***	0.355***
	(0.065)	(0.062)	(0.068)	(0.066)	(0.071)
N	689	689	689	689	689
Female *DQ* prevalence	0.110***	0.073***	0.268***	0.725***	0.302***
	(0.017)	(0.014)	(0.024)	(0.026)	(0.026)
N	356	355	340	298	315
Male *LE* prevalence	0.137*	0.143*	0.309***	0.668***	0.370***
	(0.066)	(0.061)	(0.065)	(0.061)	(0.067)
N	820	820	820	820	820
Male *DQ* prevalence	0.133***	0.146***	0.343***	0.752***	0.201***
	(0.017)	(0.018)	(0.024)	(0.023)	(0.021)
N	405	405	399	367	378
N *LE*	1509	1509	1509	1509	1509
N *DQ*	761	760	739	665	693

Notes: OLS estimations; A.F.: Armed Forced; LE equations include strata variables (sex and nine region fixed effects), and enumerator fixed effects as control variables.
Robust standard errors in brackets; *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

Figure E1: Response bias by sub-group.

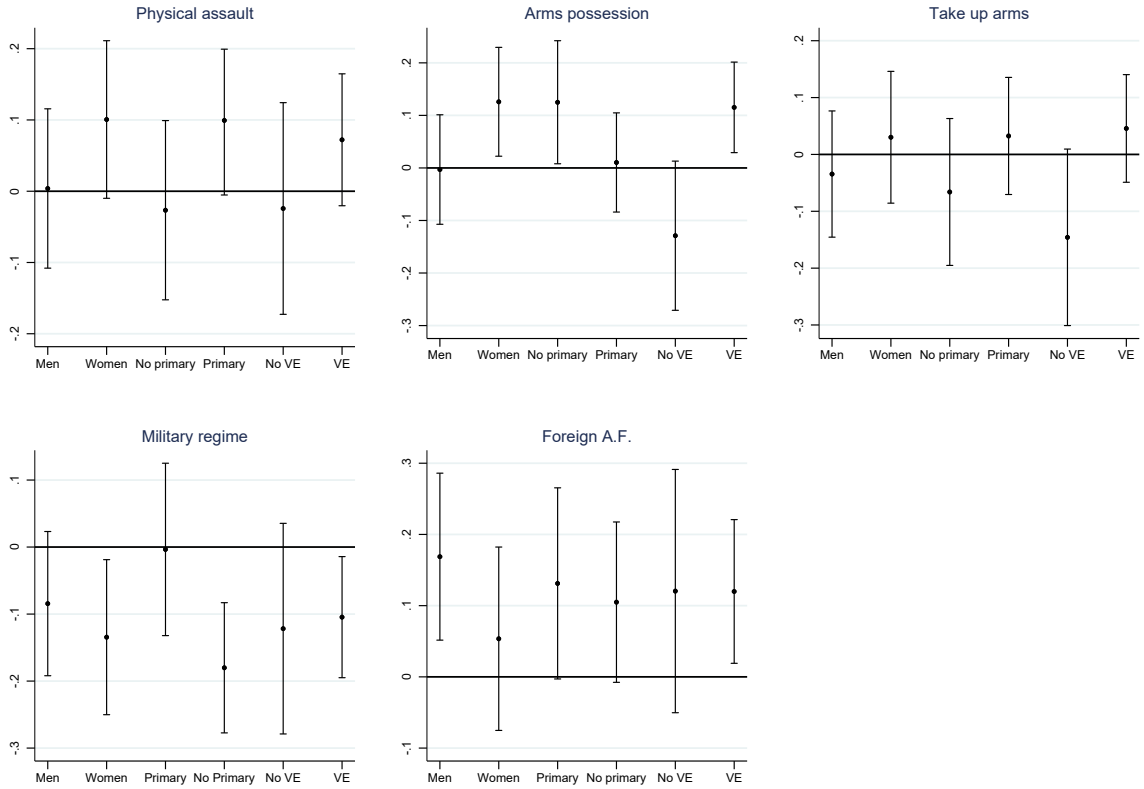


Table E3: Prevalence rates by level of education

	Q1 Physical assault	**Q2** Firearms possession	**Q3** Ready to take up arms	**Q4** support for mil. regime	**Q5** Trust in foreign A.F.
Primary not achieved *LE*	0.197***	0.122*	0.301***	0.582***	0.339***
	(0.059)	(0.056)	(0.059)	(0.055)	(0.064)
N	907	907	907	907	907
Primary not achieved *DQ*	0.098***	0.111***	0.268***	0.762***	0.234***
	(0.014)	(0.015)	(0.021)	(0.022)	(0.021)
N	451	450	436	387	398
Primary achieved *LE*	0.131	0.238***	0.300***	0.705***	0.396***
	(0.074)	(0.069)	(0.076)	(0.074)	(0.074)
N	602	602	602	602	602
Primary achieved *DQ*	0.158***	0.113***	0.366***	0.709***	0.264***
	(0.021)	(0.018)	(0.028)	(0.027)	(0.026)
N	310	310	303	278	295
N *LE*	1509	1509	1509	1509	1509
N *DQ*	761	760	739	665	693

Notes: OLS estimations; A.F.: Armed Forced; educational level is measured by the achievement or not of the primary level; LE equations include strata variables (sex and nine region fixed effects), and enumerator fixed effects as control variables.
Robust standard errors in brackets; *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

Table E4: Prevalence rates, according to violence exposure (VE)

	Q1 Physical assault	**Q2** Weapons possession	**Q3** Ready to take up arms	**Q4** support for mil. regime	**Q5** Trust in foreign A.F.
No violent events *LE*	0.073	0.051	0.151	0.615***	0.350***
	(0.088)	(0.081)	(0.087)	(0.088)	(0.094)
N	425	425	425	425	425
No violent events *DQ*	0.097***	0.179***	0.297***	0.737***	0.230***
	(0.021)	(0.028)	(0.033)	(0.034)	(0.031)
N	196	195	192	171	187
violent events *LE*	0.203***	0.204***	0.358***	0.636***	0.373***
	(0.055)	(0.051)	(0.056)	(0.052)	(0.057)
N	1084	1084	1084	1084	1084
violent events *DQ*	0.131***	0.088***	0.313***	0.741***	0.253***
	(0.014)	(0.012)	(0.020)	(0.020)	(0.019)
N	565	565	547	494	506
N LE	1509	1509	1509	1509	1509
N DQ	761	760	739	665	693

Notes: OLS estimations; A.F.: Armed Forced; "No violent events" means living in a commune without any violent events since 2012; "violent events" means living in a commune with at least one violent event that has occurred since 2012; LE equations include strata variables (sex and nine region fixed effects), and enumerator fixed effects as control variables.
Robust standard errors in brackets; *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

Table E5: Prevalence rates Q4-Q5, according to violence exposure (VE) and actors involved

	VE	VE Govt	VE foreign A.F.
	(1)	(2)	(3)
Q4: support for the military regime			
Violent events LE	0.636***	0.634***	0.678***
	(0.052)	(0.054)	(0.072)
N	1084	980	584
Violent events DQ	0.741***	0.741***	0.705***
	(0.020)	(0.021)	(0.028)
N	494	448	258
No violent events LE	0.615***	0.614***	0.608***
	(0.088)	(0.088)	(0.090)
N	425	425	425
No Violent events DQ	0.737***	0.737***	0.737***
	(0.034)	(0.034)	(0.034)
N	171	171	171
N LE	1509	1405	1009
N DQ	665	619	429
Q5: Trust in foreign A.F.			
Violent events LE	0.373***	0.383***	0.435***
	(0.057)	(0.060)	(0.079)
N	1084	980	584
Violent events DQ	0.253***	0.255***	0.232***
	(0.019)	(0.020)	(0.026)
N	506	459	263
No violent events LE	0.350***	0.351***	0.366***
	(0.094)	(0.094)	(0.095)
N	425	425	425
No Violent events DQ	0.230***	0.230***	0.230***
	(0.031)	(0.031)	(0.031)
N	187	187	187
N LE	1509	1405	1009
N DQ	693	646	450

Notes: OLS estimations; A.F.: Armed Forced; "Violent events" means living in a commune with at least one violent event that has occurred since 2012; Violent events are defined differently for each column. In VE column, violent events category is equal 1 if the respondent lives in a commune with at least one violent event, 0 otherwise (No violent events); VE Govt equals 1 if the respondent lives in a commune with at least one violent event that has occurred since 2012 and that involved the government; whereas VE foreign A.F. equals 1 if foreign armed forces were involved; the omitted category of each of these variables are communes without any violent event; "No violent events" means living in a commune without any violent events since 2012; LE equations include strata variables (sex and nine region fixed effects), and enumerator fixed effects as control variables. The coefficients associated to No violent events" of the LE estimations ($\hat{\gamma_k}$, Equation 4) change across columns 1 to 3 due to the fact that total sample sizes are different, respectively 1,509; 1,405 and 1,009 observations, which induces slight differences in the coefficients of the control variables.
Robust standard errors in brackets; *: $p < 0.10$, **: $p < 0.05$, ***: $p < 0.01$.

F Refusal to answer to direct questions

As discussed in Section 3, a few respondents of the control group refused to answer to one or more direct questions. The measured response bias is thus the result of hiding true preferences through both untruthful answers and refusal to answer.[24]

As noted in Section 5.1, the share of respondents who refused to answer is almost null for the first two questions (below 0.5% of respondents), and reaches 3.3% for question 3 (25 observations). Although still limited, refusal to answer is more frequent for questions 4 and 5, with 13% of respondents refusing to directly declare whether or not they support a military regime (99 observations), and 9.3% whether or not they trust the foreign armed forces (71 observations). This reflects the fact that these two questions are more sensitive and raise more embarrassment than the three others.

In this Appendix, we first document the individual determinants of the refusal to answer to each of the last three direct questions (Table F1)[25]. We then explore how refusal to answer affects the misreporting bias measured when comparing the DQ-based and the LE-based prevalence rates (Table F2).

Let us start by noticing that the three sub-groups of respondents who refuse to answer to questions 3, 4 and 5 overlap, as one could expect, but not entirely. In particular, among the 25 individuals who did not answer about their willingness to take up the arms, 72% (18 individuals) provided an answer to the question about the military regime (question 4), and 80% (20 individuals) to the question about the foreign armed forces (question 5). Symmetrically, among the 99 who refused to declare whether or not they support the military regime, 93% (92 individuals) did answer the question about the willingness to take up the arms (question 3), and 69% (68 individuals) answered the fifth question about the foreign armed forces. Last, among the 71 respondents who decided not to answer the final question about trust in the foreign armed forces, 93% (66 individuals) answered to the direct question about willingness to take up the arms (question 3), and more than 56% (40 individuals) declared whether or not they support the military regime (question 4).

Results in Table F1 shows that the only characteristics consistently and significantly related to the probability of refusal to answer is the gender of the respondent. Depending on the sensitive question, women are on average 2 to 10 pp more likely than men to refuse to answer. Education is negatively related to the probability of refusing to answer, although the coefficient is not statistically significant for question 3. It also appears that respondents living in localities exposed to violence are 6 pp more likely to refuse to answer the question about trust in the foreign armed forces – but violence exposure is unrelated to the probability of refusing to answer to the two other questions. Interestingly, the adjusted R^2 in the three columns is low, only 4 to 7% of the variance in the probability to refuse to answer is explained by respondents' observed characteristics and enumerator fixed effects.

[24] Question misunderstanding or incapacity to give a definite answer should be marginal given interviewers' training and efforts to make the questionnaire comprehensible.

[25] The refusal to answer to the first two questions is too rare to explore its determinants.

Table F1: Determinants of refusal to answer in DQ

Dep. var.: Refusal to answer	Ready to take up arms	Support for military regime	Trust in foreign A.F.
Male	-0.023*	-0.095***	-0.055**
	(0.01)	(0.03)	(0.02)
Educ	-0.016	-0.046*	-0.075***
	(0.01)	(0.03)	(0.02)
Exposed to violence	0.004	-0.012	0.061**
	(0.02)	(0.03)	(0.03)
Age	-0.001	0.002*	0.001
	(0.00)	(0.00)	(0.00)
Married	-0.038*	-0.012	-0.002
	(0.02)	(0.04)	(0.03)
Migrant	0.006	0.033	0.003
	(0.02)	(0.03)	(0.02)
Bambara	Yes	Yes	Yes
Region dummies	Yes	Yes	Yes
Surveyor dummies	Yes	Yes	Yes
Adjusted R^2	0.061	0.066	0.043
Observations (#refusals)	764 (25)	764 (99)	764 (71)

Standard errors in parentheses.
* p<0.10, ** p<0.05, *** p<0.01

We then run two tests to assess the impact of refusal to answer on our main results.

In the main analysis, we include responses to the lists of those who refuse to answer sensitive direct questions. One could thus be concerned by the fact that their answers are accounted for to compute the LE-based prevalence rates, but not the DQ-based prevalence rates. We test whether our main results are not driven by the fact that respondents who refuse to answer the sensitive items under DQ provide specific answers to the baseline items' lists. As a robustness check, we compute the LE-based prevalence rate of each sensitive item (and the associated response bias) by excluding from the control group those who refused to answer the corresponding direct question. The results of this exercise are displayed in Column 2 of Table F2, while Column 1 reproduces our baseline results from Table 2.

Reassuringly, our main results are robust to the exclusion of respondents who refused to answer a sensitive direct question. In particular, support for a military regime and trust in the foreign armed forces remain, respectively, over- and under-reported under DQ in similar magnitudes.

Second, instead of dropping respondents who refused to answer the corresponding direct questions, we successively impute them three different DQ answers to sensitive items. We assume that, had they answered, they would all have answered Yes, all have answered No, or have answered Yes in a proportion in line with the results of our list experiment. While relying on arbitrary hypotheses, these three exercises aim at illustrating the extent to which the response bias can be explained by refusals to answer in DQ, rather than by untruthful responses. The results of these three tests are displayed in Columns 3 to 6 of Table F2.

As already discussed, the first two questions (physical assault and firearm possession) are characterized by very few refusals, yielding very similar response biases regardless of the test implemented. While no significant bias is obtained over the whole sample for these first two items (column 1), evidence of misreporting in certain segments of the population is confirmed (see Section 5.2), which we can attribute to untruthful responses rather than refusal to answer in DQ.

We do not observe a response bias for the third question, the one about the willingness to take up the arms (column 1). Had the 25 DQ respondents who refused to directly answer that question declared that they do feel ready to take up the arms, the estimated response bias would have been larger in absolute value but still not significantly different from zero (column 3).

Results are of greater interest for items 4 and 5, that are characterized by a higher share of refusals to answer in DQ and a significant misreporting bias in the baseline estimates (column 1). Had all the respondents who refused to answer directly provided the socially desirable answer ('Yes' in the case of question 4, 'No' in the case of question 5), the estimated bias would have reached an upper bound approximately equal to 14 pp in both cases (column 3 for item 4 and column 5 for item 5). On the contrary, had all these respondents declared the plausibly undesirable answer, the estimated response bias would have shrunk (reaching 1.2 pp for question 4 and 4.6 pp for question 5), remaining of the same sign but turning non-significant in the statistical sense. This suggests that an important part of the observed misreporting is likely to be due to refusals to answer rather than untruthful responses in DQ in these cases.

Finally, when we attribute the LE-based prevalence rate to DQ respondents who refused to answer the direct questions, thus assuming they are a random selection of truthful respondents, we estimate a misreporting bias that is in between the previously mentioned upper and lower bounds, and that is very close to our baseline estimates.

Table F2: Response bias across various scenarios

	Baseline	Refusals to answer in DQ...			
		...Missing in LE	...Yes in DQ	...LE-based prev. rate in DQ	...No in DQ
Item:	(1)	(2)	(3)	(4)	(5)
Physical assault	0.049	0.051	0.045	0.048	0.049
p-value	(0.314)	(0.292)	(0.351)	(0.316)	(0.309)
Firearm possession	0.057	0.053	0.052	0.056	0.057
p-value	(0.21)	(0.238)	(0.249)	(0.212)	(0.205)
Ready to take up arms	-0.005	-0.012	-0.027	-0.005	0.005
p-value	(0.924)	(0.818)	(0.581)	(0.926)	(0.914)
Support for military regime	-0.108	-0.126	-0.141	-0.094	-0.012
p-value	(0.025)	(0.011)	(0.003)	(0.048)	(0.806)
Trust in foreign A.F.	0.116	0.129	0.046	0.106	0.139
p-value	(0.03)	(0.019)	(0.39)	(0.047)	(0.009)

Note: OLS estimations; LE equations include strata dummies (sex, and nine region fixed effects), and enumerator dummies as control variables.

Column (1) reproduces the baseline results from Table 2. The LE estimates are based on the complete sample of 1,509 observations (764 in the control group, 745 in the treated group). The DQ estimates are based on sub-samples that change across questions due to refusals to answer: among the 764 control individuals, respectively 3, 4, 25, 99 and 71 refused to answer to questions 1, 2, 3, 4 and 5.

In Column (2), the DQ estimates are identical to those in Column (1) (sample size = 761; 760; 739; 665; 693; respectively). We exclude from the LE estimates control respondents who refused to answer to the corresponding question in DQ (sample size = 1,506; 1,505; 1,484; 1,410; 1,438; respectively).

In Columns (3), (4) and (5), we replace the missing answers of control respondents who refused to answer a direct question by 'Yes' (Column (3)), by 'Yes' in the proportion indicated by the results of the LE (Column (4)), and by 'No' (Column (5)). The LE estimates are identical to those in Column (1). The DQ estimates are based on the complete 764 observations control group.