

# Selective Exposure, Heterogeneous Responses, and Sign Reversal in the Political Consequences of Violence\*

Carolina Torreblanca<sup>†</sup>

May 3, 2026

## Abstract

The burden of violence is unequally shared, and reactions to it are divergent. Yet our understanding of its consequences often comes from populations unlikely to experience it. I show that when exposure is selective, and reactions diverge, as with violence, population-wide and treated-group effects can differ in magnitude and direction. I develop a principal stratification framework that encodes selective exposure and heterogeneous effects across countervailing response types, and derive a closed-form sign-reversal condition for any binary treatment and outcome. I reassess findings linking criminal victimization to increased participation by applying a partial identification procedure derived from this framework to survey data from Latin America. I find that population-wide averages indeed favor mobilization. However, effects among the victimized are predominantly negative. Panel evidence from Mexico confirms that those likely to demobilize are more exposed. For selective treatments we would never universalize, the causal contrast of first-order importance is among the treated.

**Word count:** 9441 words

---

\*I thank Guilherme Duarte, Jane Esberg, Guy Grossman, Kun Heo, Nicolás Idrobo, Dorothy Kronick, Giacomo Lemoli, Beatriz Magaloni, Justin Melnick, Alexis Palmer, Pablo Querubín, Arturas Rozenas, Cyrus Samii, Tara Slough, David Stasavage, Elisa Wirsching, Yuhan Zheng, audiences at NEWEPS 24, APSA 2024, and at the University of Pennsylvania's Comparative Politics Workshop for excellent comments and suggestions.

<sup>†</sup>Postdoctoral Fellow, PDRI-DevLab, University of Pennsylvania, [catba@sas.upenn.edu](mailto:catba@sas.upenn.edu).

Empirical social science often takes treatment effects that average across all units in a population as a benchmark for causal knowledge. For policies that could, in principle, reach anyone or that we may want to universalize, this benchmark speaks to a question of central empirical, theoretical, and political importance. For instance, what would happen if scholarships, cash transfers, or new technologies were adopted? However, many treatments of central interest in political science are not policies we would want to universalize. Instead, they impose high costs and fall selectively on distinct subpopulations. For this class of treatments, the population-wide average informs about a counterfactual of universal exposure, which would never occur and that no one would wish for. I argue that the causal contrast of first-order importance in such instances is among those exposed.

Violence is a central example of selective, harmful treatments, and its participatory consequences have organized a substantial body of contemporary political research. Whatever its form, a central finding is that violence falls unevenly across subpopulations. Criminals target vulnerabilities (Cohen and Felson, 1979; Browning, Pinchak and Calder, 2021), police discriminate in stops (Knox, Lowe and Mummolo, 2020), or civil war combatants exercise violence selectively (Balcells and Stanton, 2021). Whether the subpopulations most exposed to violence engage politically or withdraw is a question of central importance, as it determines whether those who bear the burden of violence can shape the politics of security provision.

Existing theories expect both that violence will draw some victims into politics and that it will push others out. Specifically, research posits violence can mobilize victims through post-traumatic growth, grievance, or instrumental demand for state response (Blattman, 2009; Bateson, 2012), or demobilize them through fear and eroded trust (Ley, 2018; Trelles and Carreras, 2012). Importantly, traits that predict political participation, like race (Anoll, 2022), age (Plutzer, 2002), socioeconomic status (Brady, Verba and Schlozman, 1995), extraversion (Gerber et al., 2011), and risk tolerance (Kam, 2012), may also predict exposure to violence, including criminal victimization (Cawvey et al., 2018), police contact (Knox, Lowe and Mummolo, 2020), and civil war violence (Balcells and Stanton, 2021).

As a result, the effect of violence among those exposed can differ sharply from its effect in the population as a whole. If the question of interest is what violence does to those who suffer it, rather than what it could do to those it spares, the relevant causal contrast is not the average effect across the entire population but the average effect among those exposed. I show that the former corresponds to the average treatment effect (ATE), the latter to the average effect among the treated (ATT). I formalize when and how the ATE and ATT may differ in magnitude and direction when treatment exposure is selective and responses to it diverge. I adapt a principal stratification (Frangakis and Rubin, 2002), departing from its standard compliance application. I sort individuals into four strata by their potential outcomes. For political participation under exposure to violence, for instance, these are individuals who would participate regardless of violence exposure, only if exposed, only if not exposed, or would never participate. I further allow the probability of treatment to vary across these strata, capturing selective exposure.

I show the ATE depends only on the population shares of two countervailing response types. In the running example, those who would participate only if treated and those who would participate only if not. The ATT, by contrast, weights each population share by their probability of exposure. That is, subpopulations that are more likely to be exposed contribute more to the effect among the treated. I show that the two estimands, the ATE and the ATT, diverge in magnitude when these exposure probabilities differ across types, and that they diverge in sign when the less prevalent response type in the population is sufficiently overrepresented among the exposed.

From this framework, I derive a closed-form expression for the threshold at which sign reversal occurs, given the relative shares of countervailing response types and the ratio of their exposure probabilities. Table 2 reports this threshold across plausible ATE magnitudes and selection ratios. Researchers who have estimated a population-wide effect can use the table to determine what they would need to believe about exposure and stratum sizes for the effect among the exposed to differ in sign. The result is applicable to all settings with binary treatment and outcome, follows directly from the framework, and requires no additional data input.

In Latin America, the most violent region of the world (Vilalta, 2020), influential studies have found that crime spurs participation (Bateson, 2012; Dorff, 2017), a potential silver lining in a context of pervasive insecurity. I empirically investigate whether the conclusion that crime mobilizes survives when the contrast is defined over those victimized. To do so, I rely on 94 waves of the LAPOP survey, covering approximately 150,000 respondents across 20 Latin American countries (LAPOP, 2022), an extended version of the Latin American data used in Bateson (2012).

I develop a novel partial identification procedure (Manski, 1990) based on my principal stratification framework and apply it to the LAPOP data. I observe four moments in the data: the prevalence of victimization, the prevalence of participation, and participation rates among the victimized and non-victimized. What I cannot observe are the structural parameters of the framework: the stratum-specific population sizes or the treatment probabilities. Many parameter configurations are consistent with the observed moments in the data. Rather than pinning down a single configuration, the procedure recovers millions consistent with the data moments across the entire parameter space. Each specifies a share for each response type, an exposure rate for each, and the (ATE, ATT) pair produced by those parameters. The collection of all such pairs is the joint identified set.

Across the set, I assess how often the ATE is larger than the ATT, whether they share sign, and which selection patterns sit behind sign disagreement. I find that the majority of parameter vectors in the identified set yield a negative ATT of crime on non-electoral participation, ranging from 53% to 91% of plausible effects depending on the outcome. The ATE, by contrast, is positive in 63%–85% of those vectors, consistent with the mobilization finding in the literature. Willingness to vote in the next election, conversely, shows mostly negative ATEs and positive ATTs.

In most plausible configurations in the identified set, demobilizers face a higher risk of victimization than mobilizers, one empirical condition required for a negative ATT. But which configuration is empirically realized is something the LAPOP cross-section cannot tell us. I turn to a quarterly rotating panel survey of approximately 174,000 Mexican city residents observed over 27 quarters (INEGI, 2024). The panel allows me to separately assess who is more likely to be exposed

to victimization and how individuals respond behaviorally when they are victimized. Between-individual comparisons show that soon-to-be victims are more mobile, witness more criminal and gang activity, and report more violent conflicts than their neighbors before victimization, indicating higher baseline exposure to crime. Within-individual comparisons show that after victimization, individuals become less mobile, less trusting of law enforcement, and more fearful than before, consistent with victims coming from the demobilization stratum.

Together, these findings suggest that those most exposed to crime are also those whose participation declines after victimization. Such selection is a necessary but not sufficient condition for a negative ATT. Even when demobilizers face higher exposure, a large enough number of mobilizers in the population can keep the ATT positive. I filter the identified, retaining only the parameter configurations where demobilizers face higher exposure than mobilizers. Within the filtered set, I ask how few mobilizers per demobilizer are required for the ATT to be sign-identified as negative. Sign-identification is a stringent bar, requiring every parameter vector in the filtered set to yield a negative ATT. At a bound of no more than 4 percent more mobilizers than demobilizers, the ATT is sign-identified as negative for protest and political meetings, and over 97 percent of compatible parameter vectors yield a negative ATT for community meetings and community problem-solving. This bound clears estimated ATEs in Bateson (2012) for LATAM.

The participatory consequences of violence have also been studied for civil war and policing. In both of these settings, exposure is plausibly selective, and responses to it, heterogeneous. I apply the partial identification procedure to data from Blattman (2009) on forced abduction by the LRA in northern Uganda and from Walker (2020) on direct police contact in the United States, recovering the identified sets of ATEs and ATTs for electoral and non-electoral participation. In both cases, the direction of disagreement between the ATE and the ATT mirrors the crime application.

This paper contributes to research on causal estimands under selective treatment exposure. Heckman and Vytlacil (2005) use a structural model to show estimands differ in how they weight underlying treatment effects when treatment is endogenous to potential outcomes. I formalize an

analogous relationship for discrete response types using a simple stratification framework. I show that the ATE averages outcomes using their population shares, while the ATT reweights those shares by the probability of exposure to treatment. Consequently, the two oft-targeted estimands can differ in magnitude and have opposite signs. I derive a closed-form condition for such sign reversals that depends only on how exposure varies across response types and the prevalence of those types in the population. I also show that the joint identified set of both estimands can be recovered from four observable moments in cross-sectional survey data, contributing to recent work on partial identification (Knox, Lowe and Mummolo, 2020; Duarte et al., 2024).

The paper also contributes to scholarship that foregrounds estimand choice in causal empirical research. Prior work emphasizes that estimands, or the population of study, should align with the theoretical question and be policy-relevant (Lundberg, Johnson and Stewart, 2021; Samii, Paler and Daly, 2017; Egami and Hartman, 2023). I show that for selective harmful treatments, the policy-relevant treatment effect is defined over those exposed, not over the population as a whole. When exposure is correlated with response type, causal effects that average over an entire population, even those unlikely to experience the treatment, will differ systematically from effects defined among those exposed and may even have opposite signs. In so doing, I make explicit how theoretical knowledge about how violence is produced and by whom it is experienced should guide the choice of estimand in research on its consequences.

Last, it contributes to the literature on the participatory consequences of crime victimization. Influential studies find that victimization increases political participation (Bateson, 2012; Dorff, 2017). Others document withdrawal and reduced engagement among those exposed (Ley, 2018; Trelles and Carreras, 2012). I show that these findings are not in conflict. The ATE, informed by the counterfactual responses of people who will seldom experience crime, is positive. However, among those who actually bear its costs, the effect of crime on non-electoral participation is negative. Evidence from civil war and policing (Blattman, 2009; Walker, 2020) suggests similar reversals could be common across forms of violence.

# 1 A Principal Strata Framework

In this section, I develop a principal stratification framework that accounts for both heterogeneous responses and selective treatment exposure. The framework makes precise the conditions under which the population-average effect and the effect among the exposed diverge, and yields a partial identification procedure for recovering both from observable data.

Consider a population of units indexed by  $i$  and a binary treatment  $Z \in 0, 1$ , for example, experiencing violence or not. Denote the potential outcome for unit  $i$  under treatment assignment  $z$  as  $Y_i(z)$ . Following Frangakis and Rubin (2002), I classify units into principal strata ( $\theta$ ) defined by their joint potential outcomes  $(Y_i(1), Y_i(0))$ , where the outcome itself serves as the post-treatment variable with respect to which strata are defined. Thus, by definition, the individual treatment effect for each  $i$ ,  $ITE_i = Y_i(1) - Y_i(0)$ , will be identical for all units in the same stratum.

Denote the proportion of units in each stratum  $\pi_\theta$ , with  $\sum_\theta \pi_\theta = 1$ . I depart from the canonical principal-strata setup by allowing units to have stratum-specific probabilities of receiving treatment,  $\rho_\theta$ . Table 1 synthesizes the setup for the case of a binary outcome  $Y_i(z) \in 0, 1$ , such as whether an individual participates in a given political activity or not.

<b>Stratum (<math>\theta</math>)</b>	<b>Share</b>	<b>Pr(Z=1)</b>	<b>Y(1)</b>	<b>Y(0)</b>	<b>ITE</b>
Always Participate	$\pi_A$	$\rho_A$	1	1	0
Participate if Treated	$\pi_T$	$\rho_T$	1	0	1
Participate if Untreated	$\pi_U$	$\rho_U$	0	1	-1
Never Participate	$\pi_N$	$\rho_N$	0	0	0

Table 1: Stratum-specific individual treatment effects and treatment probabilities. Each row defines a principal stratum by its potential outcomes under treatment and control. The treatment probability  $\rho_\theta$  is allowed to vary across strata.

For a binary treatment such as exposure to violence, we can classify individuals into four strata: those who always participate (A), those who never participate (N), those who participate if and only if treated (T), and those who participate if and only if untreated (U).<sup>1</sup>

**Proposition 1.** *The average treatment effect is  $ATE = \pi_T - \pi_U$ .*

Proposition 1 applies the principal stratification framework of Frangakis and Rubin (2002) to binary outcomes. It states that the ATE depends exclusively on the relative shares of strata  $T$  and  $U$ . This result is straightforward, given that all units in strata  $A$  and  $N$  contribute an ITE of zero. Consequently, the magnitude and direction of the ATE will only depend on the frequency with which  $T$  appears in the population, relative to  $U$ .<sup>2</sup>

**Proposition 2.** *The average treatment effect on the treated is*

$$ATT = \frac{\pi_T \rho_T - \pi_U \rho_U}{Pr(Z = 1)},$$

where  $Pr(Z = 1) = \sum_{\theta} \pi_{\theta} \rho_{\theta}$ .

Conversely, the ATT, or the average expected change in outcomes for those who are treated, will be a function of population shares and also of the stratum-specific probabilities of receiving treatment. These stratum-specific probabilities scale the relative contributions of the  $T$  and  $U$  strata within the treated population. Intuitively, strata with higher treatment probabilities constitute a larger share of the treated group and therefore should contribute more to the ATT. Further rearranging the numerator of the expression for the ATT yields an identity for the population-level impact of treatment:

$$ATT \times Pr(Z = 1) = \pi_T \rho_T - \pi_U \rho_U. \tag{1}$$

I denote this quantity  $\Delta Y$ , the average change in  $Y$  attributable to the realized treatment assignment.

---

<sup>1</sup>I make the stable unit treatment value assumption (SUTVA) throughout.

<sup>2</sup>Formal proofs for all expressions in this section are provided in Appendix Appendix A.

**Corollary 1.** *The ATE and ATT are equal when  $\rho_T = \rho_U = Pr(Z = 1)$ .*

Intuitively, when  $\rho_T = \rho_U = Pr(Z = 1)$ , all strata face the same probability of treatment, so conditioning on being treated does not alter the relative representation of any stratum. That is, the treated group is a random draw from the population, so the ATE and the ATT average over identical weights in expectation.<sup>3</sup>

**Corollary 2.** *When  $\rho_T = \rho_U$ , the ATE and ATT have the same sign. Specifically,  $ATT = \frac{\rho_T}{Pr(Z=1)} \cdot ATE$ .*

Corollary 2 relaxes Corollary 1 by allowing the treatment-insensitive strata ( $A$  and  $N$ ) to have treatment probabilities that differ from those of the treatment-sensitive strata. Because the  $A$  and  $N$  strata have a treatment effect of zero, their over- or under-representation in the treated group cannot change the sign of the ATT relative to the ATE. Only the relative weighting of  $T$  and  $U$  determines the sign, so  $\rho_T = \rho_U$  suffices for sign agreement. The treatment-insensitive strata do, however, affect the *magnitude*: the scaling factor  $\rho_T/Pr(Z = 1)$  can amplify or attenuate the ATT but cannot reverse its direction. Corollary 1 is the special case where  $\rho_T = Pr(Z = 1)$ , so the scaling factor equals one.

**When do the ATE and ATT disagree in sign?** The ATE and ATT need not agree in sign. Signs agree when composition and selection reinforce each other: for example, when the  $T$ -stratum is both larger than the  $U$ -stratum and at least as likely to be treated, or when the  $U$ -stratum is both larger and at least as likely to be treated. Signs can diverge when one stratum is larger but the other is more likely to be treated, so that the stratum that dominates the population average is under-represented among the treated. I now characterize exactly when this occurs.

---

<sup>3</sup>Corollary 1 also parallels the condition identified by Heckman and Vytlacil (2005) under which all treatment parameters coincide. They derived this result within a structural framework that explicitly models the selection process, requires an instrument, and recovers the full marginal treatment effect curve. Here, the same condition emerges from the structure of principal stratification alone, without parametric assumptions on the selection process or an exclusion restriction.

To do so, I define two summary parameters. Let  $\lambda = \rho_U/\rho_T$  capture differential selection between the responsive strata. When  $\lambda > 1$ , the  $U$ -stratum is more likely to be treated than the  $T$ -stratum; when  $\lambda < 1$ , the reverse holds. Last, when  $\lambda = 1$ , we have equal selection and necessarily sign agreement (Corollary 2). Let  $\eta = \pi_T/\pi_U$  capture the relative size of the two responsive strata. Hence  $\eta > 1$  when the  $T$ -stratum is larger ( $ATE > 0$ ) and  $\eta < 1$  when the  $U$ -stratum is larger ( $ATE < 0$ ).

**Proposition 3.** *Let  $\pi_T > 0$  and  $\pi_U > 0$ . Then  $\text{sign}(ATT) = \text{sign}(\pi_T\rho_T - \pi_U\rho_U)$ , and the ATE and ATT have opposite signs if and only if  $\lambda > \eta$  when  $ATE > 0$ , or  $\lambda < \eta$  when  $ATE < 0$ .*

The sign of the ATT is determined by whether the  $T$ -stratum's exposure-weighted size ( $\pi_T\rho_T$ ) exceeds that of the  $U$ -stratum ( $\pi_U\rho_U$ ). When  $\lambda > 1$ , the  $U$ -stratum is over-represented among the treated, pulling the ATT toward negative values. A sign reversal occurs when this differential selection more than compensates for the size advantage of the  $T$ -stratum, that is, when  $\lambda > \eta$ . The condition is symmetric: when  $ATE < 0$  ( $\eta < 1$ ), the ATT is positive only if  $\lambda < \eta$ . In that case, the  $T$ -stratum must be sufficiently more exposed to treatment than the  $U$ -stratum to overcome its smaller population share.

**Corollary 3.** *When  $ATE > 0$  and  $\lambda > 1$ :*

$$ATT < 0 \iff \pi_U > \frac{ATE}{\lambda - 1}. \quad (2)$$

*By symmetry, when  $ATE < 0$  and  $\lambda < 1$ , the ATT is positive if and only if  $\pi_T > |ATE| \cdot \lambda / (1 - \lambda)$ .<sup>4</sup>*

For the sign of the ATT to oppose the sign of the ATE, the countervailing stratum must constitute a sufficiently large share of the total population. How large depends on two quantities: (i) the magnitude of the ATE, and (ii) how much more exposed the countervailing stratum is to treatment. When the ATE is small or  $\lambda$  is far from one, the threshold is low, so sign reversal requires only

---

<sup>4</sup>Define  $\lambda' = \rho_T/\rho_U = 1/\lambda$  and  $\eta' = \pi_U/\pi_T = 1/\eta$ . The sign-flip condition  $\lambda < \eta$  is equivalent to  $\lambda' > \eta'$ , and the same algebra yields  $\pi_T > |ATE|/(\lambda' - 1) = |ATE| \cdot \lambda / (1 - \lambda)$ .

a modest population share of the countervailing stratum. Table 2 illustrates this relationship for a range of ATE magnitudes and selection ratios  $\lambda$ . The table applies to any binary treatment and binary outcome; it follows entirely from the framework and requires no empirical input.

Consider the bottom-left region of the table, where the ATE is positive and the  $T$ -stratum is also more likely to be treated ( $\lambda < 1$ ). These cells show check marks when composition and selection reinforce each other, so the ATE and ATT always agree in sign regardless of stratum shares. Now consider the bottom-right region, where the ATE is still positive but the  $U$ -stratum is more likely to be treated ( $\lambda > 1$ ). Here, the threshold cells show the minimum population share of the  $U$ -stratum ( $\pi_U$ , as a fraction of the total population) above which the ATT turns negative. For example, at  $ATE = 0.04$  and  $\lambda = 2$  (meaning that for every member of the  $T$ -stratum who is treated, two members of the  $U$ -stratum are treated, per capita), the ATT is negative whenever the  $U$ -stratum constitutes more than 4% of the population. Conversely, the top-left region applies when the ATE is negative and the  $T$ -stratum is more likely to be treated ( $\lambda < 1$ ). At  $ATE = -0.10$  and  $\lambda = 1/3$  (meaning for every member of the  $U$ -stratum who is treated, three members of the  $T$ -stratum are treated, per capita), the ATT is positive whenever the  $T$ -stratum exceeds 5% of the population.

		Selection ratio $\lambda = \rho_U/\rho_T$						
		<i>T</i> more likely treated				<i>U</i> more likely treated		
		$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	1	$\frac{3}{2}$	2	3
<i>ATE</i>	-0.10	$\pi_T > 5\%$	$\pi_T > 10\%$	$\pi_T > 20\%$	✓	✓	✓	✓
	-0.04	$\pi_T > 2\%$	$\pi_T > 4\%$	$\pi_T > 8\%$	✓	✓	✓	✓
	-0.02	$\pi_T > 1\%$	$\pi_T > 2\%$	$\pi_T > 4\%$	✓	✓	✓	✓
	0.02	✓	✓	✓	✓	$\pi_U > 4\%$	$\pi_U > 2\%$	$\pi_U > 1\%$
	0.04	✓	✓	✓	✓	$\pi_U > 8\%$	$\pi_U > 4\%$	$\pi_U > 2\%$
	0.10	✓	✓	✓	✓	$\pi_U > 20\%$	$\pi_U > 10\%$	$\pi_U > 5\%$

Table 2: Minimum share of the countervailing stratum required for the ATT to have opposite sign from the ATE, as a function of the ATE and the selection ratio  $\lambda = \rho_U/\rho_T$ . When  $\lambda > 1$ , the *U*-stratum (demobilizers) is more likely to be treated than the *T*-stratum (mobilizers); when  $\lambda < 1$ , the reverse. Cells show the threshold from Corollary 3. Off-diagonal check marks indicate regions where differential selection reinforces population share differences, so the ATE and ATT always agree in sign. The  $\lambda = 1$  column indicates no differential selection, so signs agree by Corollary 2. The table is symmetric around  $\lambda = 1$  and  $ATE = 0$ .

The framework and Table 2 are entirely data-agnostic: they apply to any setting with a binary treatment and a binary outcome in which treatment plausibly generates opposing responses across distinct subpopulations, and in which exposure to treatment covaries with response type. Any empirical researcher can use the table to assess, for a given ATE magnitude and belief about differential selection, whether the ATE and ATT might diverge in sign. In the next section, I apply this framework to criminal victimization and political participation.

## 2 Application to Criminal Victimization

Extant theoretical accounts of the participatory consequences of criminal victimization, drawing heavily from work on civil war and policing violence, posit that victimization may affect participation decisions through two competing mechanisms. The first encompasses processes that increase

participation. For instance, post-traumatic growth can transform victims into civic actors, anger at victimization can generate demand for political action, shared exposure can build solidarity that lowers collective action costs, or victims may acquire instrumental incentives to demand security or punish incumbents (Bateson, 2012; Blattman, 2009; Koos and Traunmüller, 2024; Gilligan, Pasquale and Samii, 2014; Ley, 2022). The second encompasses processes that suppress participation. For example, by engendering fear, crime can raise the perceived costs of public engagement. Alternatively, victimization can erode trust in the capacity of the state to protect (Ley, 2018; Hale, 1996). These two broad channels, and the many mechanisms they encompass, map onto the *Participate if Treated* and *Participate if Untreated* strata, respectively.

A central feature of crime is that exposure is not randomly distributed (Cohen and Felson, 1979; Gottfredson, 1986). Victimization risk varies systematically with mobility, risk preferences, socioeconomic status, and access to protection (Miethe, Stafford and Long, 1987). Wealthier and more risk-averse individuals tend to be better able or more willing to avoid victimization (Skogan, 1995; Hale, 1996), whereas more mobile and risk-prone individuals embedded in their communities face greater exposure and often weaker protection (Rader, 2004; Browning, Pinchak and Calder, 2021; Boggs, 1965; Guedes, Domingos and Cardoso, 2018). These same characteristics are also correlates of political participation (Gerber et al., 2011; Kam, 2012); in Latin America, the same profiles predict both (Cawvey et al., 2018). Crime is therefore a setting in which who is exposed and how they respond are linked.

Influential empirical research concludes that victimization, on average, increases non-electoral civic and political engagement (Bateson, 2012; Dorff, 2017; Ley, 2022). This evidence is largely generated by research designs that target population-average or conditional average treatment effects. Bateson (2012), for example, conditions on observed covariates, estimating how participation would change if victimization were randomly assigned within covariate strata. This targets a coherent counterfactual question, the conditional average treatment effect, but it does not describe how participation changes among those who are victimized. If subpopulations facing higher vic-

timization risk are also more likely to be demobilized by victimization, population-wide average effects need not describe effects among victims.

The conceptual distinction is unrelated to confounding concerns, or bias when estimating any one given causal estimand.<sup>5</sup> It is about the antecedent decision of what causal quantity to target, whether one defined over the entire population or defined over the treated and how that shapes the answers we get. For example, Sønderskov et al. (2022), using Danish administrative panel data, find that estimated associations between criminal victimization and turnout yield the opposite sign when the comparison is made between victims and non-victims, targeting a population-wide effect, and when it is made within-victims only, targeting the ATT.

## 2.1 Data

I focus on Latin America, where victimization is widespread (Vilalta, 2020; LAPOP, 2022) and where much of the research on the political consequences of crime has been conducted (Visconti, 2020; Ley, 2018; Blanco, 2013; Marshall, 2022; Kronick, 2014). As the primary data source, I employ an extended version of the LAPOP survey, which Bateson (2012) uses for their Latin American results. I pool all surveys conducted in 20 Latin American countries between 2010 and 2019, yielding approximately 150,000 respondents from 94 survey waves. Each survey round was designed to represent the voting-age population of the country in that year.<sup>6</sup> LAPOP measures self-reported victimization in the past 12 months and several forms of civic and political engagement. I focus on five outcomes: protest participation, political meeting attendance, community meeting attendance, participation in community problem-solving activities, and willingness to vote in upcoming national elections if they were held next week.

---

<sup>5</sup>Others have questioned the mobilization finding on confounding grounds; see Boulding, Mullenax and Schauer (2022).

<sup>6</sup>See Appendix A2.1.1 for sampling details.

## 2.2 Empirical Strategy

I implement the partial identification procedure using the LAPOP data. From the pooled data, I can retrieve the proportion of respondents who report being victims,  $Pr[Z = 1]$  and its converse probability, the proportion of respondents who report engaging in each of the participatory activities,  $Pr[Y = 1]$ , and its converse, and the proportion of victimized/unvictimized respondents who report participating  $Pr[Y = 1|Z = 1]$ ,  $Pr[Y = 1|Z = 0]$ , respectively. Table 3 reports the observed probabilities in the data for each of the outcomes.<sup>7</sup>

Leveraging the structure of the principal-strata framework, these observed quantities restrict the possible values of the population shares of each stratum ( $\pi_\theta$ ) and their corresponding treatment probabilities ( $\rho_\theta$ ). The expressions derived in Section 1 imply that any feasible parameter vector must reproduce the observed participation moments through the identities linking  $(\pi_\theta, \rho_\theta)$  to  $(Pr[Y = 1], Pr[Y = 1 | Z = 1], Pr[Y = 1 | Z = 0], Pr[Z = 1])$ . For each feasible vector, I then compute the ATE and ATT directly using the closed-form expressions established earlier:

$$ATE = \pi_T - \pi_U, \quad ATT = \frac{\pi_T \rho_T - \pi_U \rho_U}{Pr(Z = 1)}.$$

To characterize the full set of parameter values consistent with the data, I explore the feasible parameter space for  $\{\pi_\theta, \rho_\theta\}$  for all  $\theta \in \{A, T, U, N\}$ . The approach follows the logic of partial identification: instead of selecting assumptions that point-identify either estimand, I recover the entire identified set of parameter combinations implied jointly by the observed moments and the model's structure (Kline and Tamer, 2023). This proceeds as follows.

1. I construct a fine grid over all possible population shares  $(\pi_A, \pi_T, \pi_U, \pi_N)$  so  $\sum_\theta \pi_\theta = 1$ .
2. For each candidate vector of population shares, I use the observed moments in Table 3 to

---

<sup>7</sup>These proportions are estimated from survey data and therefore contain uncertainty. However, the large sample size yields very precise estimates, rendering the incorporation of this uncertainty negligible and immaterial to the results. For simplicity and clarity, I omit this uncertainty from the explanation.

Participation Type	N	$Pr(Z = 1)$	$Pr(Y = 1)$	$Pr(Y = 1   Z = 1)$	$Pr(Y = 1   Z = 0)$
Attended Protest	149,728	0.21 (0.001)	0.08 (0.0007)	0.14 (0.002)	0.07 (0.0007)
Community Problem-Solving	94,022	0.19 (0.001)	0.36 (0.001)	0.43 (0.004)	0.34 (0.002)
Attended Community Meeting	149,405	0.21 (0.001)	0.29 (0.001)	0.32 (0.003)	0.28 (0.001)
Attended Political Meeting	137,308	0.20 (0.001)	0.16 (0.0009)	0.18 (0.002)	0.15 (0.001)
Would vote next week	127,256	0.22 (0.001)	0.81 (0.001)	0.85 (0.002)	0.80 (0.001)

Table 3: Descriptive statistics from the LAPOP survey for each measure of political participation.  $N$ : respondents with non-missing participation and victimization responses.  $Pr(Z = 1)$ : victimization rate.  $Pr(Y = 1)$ : unconditional participation rate.  $Pr(Y = 1 | Z = 1)$  and  $Pr(Y = 1 | Z = 0)$ : participation rates among victims and non-victims, respectively. Standard errors in parentheses.

solve for the implied treatment probabilities  $\rho_T$  and  $\rho_U$  that rationalize the observed conditional participation rates.

3. I then solve for  $\rho_A$  and  $\rho_N$  using the marginal probability of victimization,  $Pr(Z = 1)$ .
4. Any parameter vector that violates logical or probabilistic constraints (e.g., treatment probabilities outside  $[0, 1]$  or participation rates inconsistent with potential outcomes) is discarded.
5. For every remaining feasible parameter vector, I compute the corresponding ATE and ATT using the expressions above, as well as the implied differential selection ratio  $\lambda = \rho_U / \rho_T$  and stratum-size ratio  $\eta = \pi_T / \pi_U$ , which jointly govern whether the ATT agrees in sign with the ATE (Section 5).

The procedure yields the identified set, that is, the set of parameter vectors consistent with the LAPOP moments. Each vector implies an ATE, ATT pair along with stratum-specific treatment probabilities, and stratum-specific population shares. Together, these vectors characterize the causal effects compatible with the data and the framework. Whenever individuals who would reduce participation after victimization are more likely to be victimized (i.e.,  $\rho_U > \rho_T$ ), the ATE and ATT diverge, often in magnitude and sometimes in sign. All algebraic derivations, feasibility constraints, and computational details are reported in Appendix C.

### 3 Results: Crime Victimization

After repeating the algorithm detailed in the previous section for each of the participatory outcomes, I obtained 12.1 million, 16 million, 20.6 million, 18.34 million, and 3.63 million feasible parameter vectors  $\{\pi_\theta, \rho_\theta\} \forall \theta \in \{A, T, U, N\}$ , each implying a specific  $\lambda$  and  $\eta$ , for the protest attendance, political meeting attendance, community meeting attendance, community problem-solving outcomes, and willingness to vote respectively. These are an internally consistent set of values that bound the actual value of each parameter. The difference in the sizes of the parameter sets depends on how informative the data is; for infrequent outcomes, fewer parameter sets are consistent with the data.

For each set, I can then calculate the corresponding ATE and ATT for victimization on each measure of political participation. Figure 1 plots the results. Each point in the graph corresponds to a plausible combination of  $\pi_\theta, \rho_\theta$  for a given participatory outcome. The red diagonal line that passes through the origin indicates the points where the ATE and the ATT are equal. For all the points below, the ATE is larger than the ATT. Conversely, the ATT is larger than the ATE for all points above that line. The color of the point indicates the difference between the probability of becoming a victim for individuals who become mobilized due to victimization and the probability of becoming a victim for individuals who respond by decreasing political participation in the set.

The implied ATE is almost always larger than the corresponding ATT across all measures of non-electoral political participation. When the outcome is participating in a protest, the ATE is larger than the ATT in 99% of the parameter sets. When attending political meetings, is the outcome the same for 96% of the parameter sets. When it is attending community meetings, the ATE is larger in 83% of parameter sets. When community problem-solving activities is the outcome, the percentage is 64%. Further, we can see that the points above the red diagonal line correspond to the largest differences between  $\rho_T - \rho_U$ . For the ATT to be larger than the ATE, selection into victimization has to be strongly correlated with the treatment, making individuals

more participative.

While the majority of the ATEs compatible with the data are positive for all four measures of non-electoral participation (85% for protest attendance, 79% for political meeting attendance, 68% for community meeting attendance, and 63% for community problem solving), the converse is true for the ATTs across measures. Specifically, only 9% of the ATTs are positive for protest attendance, only 13% for political meeting attendance, 29% for community meeting attendance, and 47% for community problem-solving. Even for the most frequently reported outcome, community meeting attendance, the ATTs compatible with the data are negative in most cases. As reported in Table 3, the larger share of positive ATEs and negative ATTs reflects the conditional probabilities of participating in these activities given victimization implied in the respondents' answers.

The results from the willingness-to-vote question, if elections were held the following week, show that electoral participation shows a different pattern than non-electoral forms of engagement. While the ATE exceeds the ATT across most non-electoral participation measures, this relationship reverses for voting intention. Here, the ATE exceeds the ATT in only 1% of parameter sets. More strikingly, in 89% of parameter sets compatible with the data, the ATE of victimization on willingness to vote is *negative*, while in 95% of the sets, the ATT is positive.

Three considerations help explain this reversal. First, voting may be relatively less costly than non-electoral participation. For the marginal individual, it requires no collective organization, no sustained community presence, and no resources beyond showing up. The mechanisms that suppress costly participation after victimization may therefore not suppress voting. Second, strata may be outcome-specific. An individual can be in the *Participate if Untreated* stratum for protest but in the *Participate if Treated* stratum for voting. With roughly 8% of respondents reporting willingness to vote and 20% victimization, the population is dominated by *Always* voters for whom voting behavior is unaffected by victimization, consistent with the empirical stickiness of electoral participation (Coppock and Green, 2016). Third, the reversal is consistent with extant evidence. Sønderskov et al. (2022), find that victimization increases turnout by 2 to 3 percentage points

within individuals but that the between-individual association is large and negative.

The divergence between the ATE and ATT could be driven by the relative size of the mobilized and demobilized strata, their differential exposure to victimization, or a combination of both. Figure 2 plots the cumulative distribution of these two quantities across all parameter vectors compatible with the LAPOP data. The left panel shows  $\eta = \pi_T/\pi_U$ , the ratio of mobilizers to demobilizers; the right panel shows  $\lambda = \rho_U/\rho_T$ , the ratio of demobilizer to mobilizer victimization risk. Both panels use a log scale, with the dashed line indicating the reference value of 1.

The left panel shows that across all four non-electoral outcomes, the vast majority of the identified set has  $\eta > 1$ , with mobilizers outnumbering demobilizers in the population at large. Protest has the highest  $\eta$  values, with a median around 3.5; the community outcomes cluster lower, with medians between 1.7 and 2.6. The cumulative proportion at which each curve crosses the dashed line equals the share of vectors with  $ATE > 0$ , since  $ATE = \pi_T - \pi_U$ . For voting, the pattern reverses and  $\eta < 1$  for most of the identified set, indicating that demobilizers outnumber mobilizers and that the ATE is predominantly negative.

The right panel tells the complementary story. For all four non-electoral outcomes,  $\lambda > 1$  across the vast majority of the identified set, meaning demobilizers face a higher victimization risk than mobilizers. Protest and political meetings show the strongest pattern, with  $\lambda$  rising above 1 early in the cumulative distribution. Voting again reverses:  $\lambda < 1$  for most vectors, meaning mobilizers are more likely to be victimized than demobilizers.

The results suggest that if the probability of victimization were uniform across the entire population, it would likely increase non-electoral participation, consistent with Bateson (2012) and the field's prevailing understanding.<sup>8</sup> However, because individuals who would participate less after victimization are disproportionately likely to become victims, the identified sets suggest that crime reduces all four types of non-electoral participation in the region. The pattern reverses for will-

---

<sup>8</sup>Results in Bateson (2012) are based on one wave of the LAPOP data. Section A2.3 in the Appendix replicates the results from Bateson (2012) with repeated cross-sections of the LAPOP data, as well as within-country analyses.

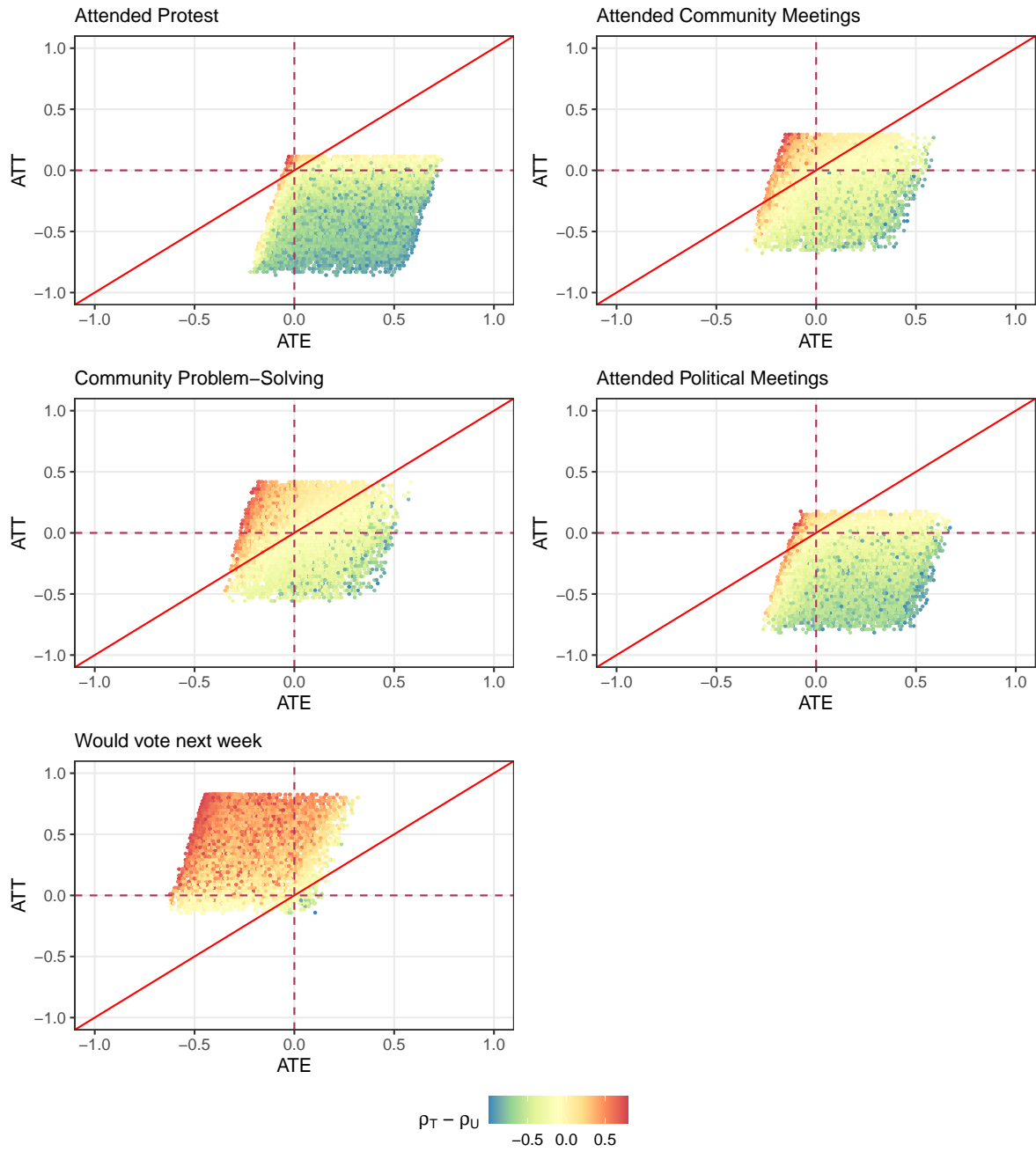


Figure 1: Figure shows the combination of potential ATTs and ATEs of victimization on four measures of political participation, conditioning on observed participation, and criminal victimization (LAPOP, 2022).

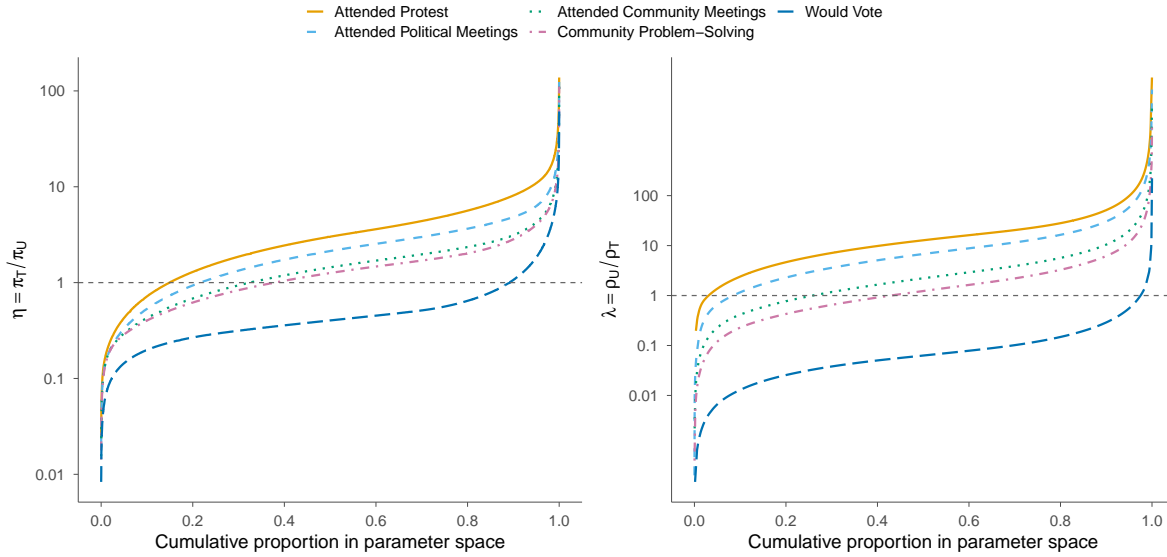


Figure 2: Cumulative distribution of the stratum-size ratio  $\eta = \pi_T/\pi_U$  (left) and the differential selection ratio  $\lambda = \rho_U/\rho_T$  (right) across parameter vectors compatible with the LAPOP data. Log scale. The dashed line marks 1: above it, mobilizers outnumber demobilizers (left), or demobilizers face a higher risk of victimization (right). The cumulative proportion at which a curve crosses 1 equals the share of the identified set with  $ATE > 0$  (left) or  $\rho_U > \rho_T$  (right).

ingness to vote: crime, were it random, would likely reduce willingness to vote, but given actual patterns of victimization, it likely increases it.

Participation Type	$Pr(Z = 1)$	ATE			ATT			$\Delta Y$	
		Min	Max	$Pr(ATE < 0)$	Min	Max	$Pr(ATT < 0)$	Min	Max
Attended Protest	0.19	-0.23	0.76	0.15	-0.84	0.12	0.91	-0.18	0.02
Community Problem-Solving	0.19	-0.38	0.61	0.37	-0.57	0.42	0.53	-0.11	0.08
Attended Political Meetings	0.20	-0.28	0.70	0.21	-0.82	0.16	0.87	-0.17	0.03
Attended Community Meetings	0.21	-0.36	0.63	0.32	-0.68	0.30	0.71	-0.14	0.06
Would vote next week	0.22	-0.65	0.33	0.89	-0.13	0.84	0.05	-0.03	0.18

Table 4: Identified sets of ATE, ATT, and  $\Delta Y = ATT \times Pr(Z = 1)$  from all feasible parameter vectors compatible with the LAPOP data, without constraining  $\rho_U$  or  $\rho_T$ . Min and Max report the bounds of each identified set.  $Pr(ATE < 0)$  and  $Pr(ATT < 0)$  report the share of feasible parameter vectors for which each estimand is negative.

What are the magnitudes of the changes in participation? Recall from expression (1) that the

realized change in population outcomes equals the ATT scaled by the probability of treatment. Table 4 synthesizes all the results. Participation in protests could decrease by as much as 18 pp, engagement in community problem-solving could decrease by 11 pp, political meeting attendance could decrease by 17 pp, while community meeting attendance could decrease by as much as 14 pp. Willingness to vote next week, however, at worst could decrease by 3 pp. Conversely, conditional on the data, protest attendance could *at best* increase only by 2 pp; community problem solving by 8 pp; political meeting attendance by 3 pp; and community meeting attendance by 6 pp. Hypothetical willingness to vote could increase by at most 18 pp. The *upper bound effects* within the identified set of non-electoral participation estimands are comparable in magnitude to, and often smaller than, the ATEs identified in Bateson (2012).

One concern is that these results reflect peculiarities of the LAPOP data. For example, we might be worried that LAPOP over-represents urban areas, under-represents dangerous areas, or yields survey responses that do not generalize. To address such concerns, I analyze data from Dorff (2017), who conducted a nationally representative survey of 1,000 Mexican respondents that included questions about victimization and both electoral and non-electoral participation. I apply the same partial identification procedure from the main paper to this alternative dataset. I present the results in Table A6 in the Appendix. I find strikingly similar results.

The partial identification procedure that underpins all results in this section takes the observed moments as given. A reasonable concern is that self-reported victimization is subject to differential *reporting* that covaries with political participation, with politically engaged individuals more likely to classify an experience as criminal (Skogan, 1982; Boulding, Mullenax and Schauer, 2022). Appendix A5.1 formalizes how such bias would propagate through the framework. If it takes the form assumed in the literature, where the politically active are more likely to report, it would inflate both the ATE and the ATT, making a negative ATT harder to detect. Appendix E also presents evidence that self-reported victimization tracks municipal homicide rates in Mexico, Brazil, and Colombia, and that Mexican respondents who report being asked for a bribe are also more likely to

report being extorted, suggesting that crime experience maps onto reporting, assuaging concerns that the measure is not construct-valid.

## **4 Panel Evidence from Mexico**

The previous section established that demobilization among the victimized is consistent with the majority of the parameter space arising from pooled data from 20 Latin American countries. The natural next question is where in that space we actually are. Answering it requires evidence on who is selected into victimization and how they respond, which cross-sectional data cannot provide. I turn to a panel survey from Mexico, one of the countries in the LAPOP sample and the second largest country in Latin America, that follows the same individuals over time.

I use a quarterly rotating panel survey of Mexican city residents that asks questions related to crime and insecurity, the National Survey of Urban Public Safety, ENSU, for its name in Spanish (INEGI, 2024). This survey has been conducted four times per year since 2013<sup>9</sup> and, starting in 2017, has included direct questions regarding personal victimization. This rolling survey panel allows me to observe individuals at five different points over 15 months. In each quarter, respondents are asked about experiences with the state, crime, and conflict, and in at least two waves per year, they are asked about direct victimization. I keep all survey waves from 2017, when crime victimization was first incorporated, through the last quarter of 2023, resulting in 560k responses from 174K unique respondents collected over 27 quarters.

Unlike the LAPOP data, the panel structure of ENSU allows me to provide direct evidence on both conditions that the framework requires. First, by comparing future victims with non-victims before any victimization occurs, I assess whether exposure is selective: if it is, future victims should differ systematically from their neighbors even before becoming victims. Second, by exploiting within-individual variation, I estimate how victimization changes attitudes and behavior among those who experience it, providing evidence on the direction of response among the

---

<sup>9</sup>Except the second quarter of 2020 due to the COVID emergency.

treated.<sup>10</sup>

To assess whether exposure to victimization is selective, I identify ENSU questions that map onto the characteristics the framework calls for to vary across strata: mobility patterns, exposure to criminal and violent behavior, feelings of unsafety, and interpersonal conflicts. Higher mobility increases exposure to crime (Cohen and Felson, 1979; Miethe, Stafford and Long, 1987; Rader, 2004; Browning, Pinchak and Calder, 2021); witnessing criminal behavior directly indexes victimization risk (Gottfredson, 1986); and fear of crime correlates with perceived vulnerability (Hale, 1996; Guedes, Domingos and Cardoso, 2018). Importantly, these same characteristics are also correlates of political participation, as extroverted, mobile, community-oriented, and risk-prone individuals participate more (Gerber et al., 2011; Kam, 2012; Ley, 2022). In Latin America, these are also the individuals who are victimized more often (Cawvey et al., 2018).

To assess reactions to victimization, I estimate how victimization changes these same attitudes and behaviors among those who experience it. I exploit within-individual variation using two estimators: a two-way fixed effects estimator and the fixed-effect counterfactual estimator proposed by Liu, Wang and Xu (2022). I focus on measures of crime prevention behavior, fear of crime, trust in law enforcement, and mobility. If victimization demobilizes, victims should become less mobile, more fearful, and less trusting; the behavioral profile of the *Participate if Untreated* stratum.

#### **4.1 Exposure to Crime**

I subset the data and retain only respondents who never reported criminal victimization and respondents who reported criminal victimization in some, but not all, survey waves. For the latter, I retain responses only before their first reported instance of victimization. Consequently, I compare the responses of this group of future victims *before they reported victimization* with those of their never-victimized neighbors—either from the same sampling unit or the same locality—in the same

---

<sup>10</sup>Like any survey measure, self-reported victimization may be subject to heterogeneous reporting: individuals may interpret events as criminal or non-criminal differently, and such sensitivity could correlate with political attitudes and behavior (Skogan, 1982; Boulding, Mullenax and Schauer, 2022). Two descriptive analyses in Appendix Appendix E show that self-reported victimization tracks municipal-level criminal victimization rates as expected.

quarter.

I estimate

$$PM_{i[q,g]} = \beta_1 * FutureVictim_{i[q,g]} + \theta_{q \times g} + \epsilon_{i[q,g]} \quad (3)$$

where  $PM_{i[q,g]}$  is respondent  $i$ 's self-reported behavior, experiences, or attitude in year-quarter  $q$ , and neighborhood/census tract or locality  $g$ , or a binary measure that takes the value of 1 if respondent  $i$  reports that type of behavior or experience and 0 otherwise.  $FutureVictim_{i[q,g]}$  is a binary variable that takes the value of 1 when respondent  $i$  will report having been a victim of crime in a future survey and 0 otherwise.  $\theta_{q \times g}$  represent either *Neighborhood*  $\times$  *Quarter* fixed effects, or *Locality*  $\times$  *Quarter* fixed effects.  $\epsilon_{i[q,g]}$  are robust errors clustered at the primary sampling unit.

Future victims differ systematically from non-victims before victimization occurs (Figure 3). When we compare future victims with never-victims living in the same locality in that same quarter (dark blue) or the same small neighborhood (light blue), future victims report having more conflicts, including conflicts that resulted in physical violence, conflicts with authorities, conflicts with the police, and conflicts with strangers on the street, *before* being victimized. Future victims also witnessed gangs, robberies, and vandalism, and heard gunshots more frequently. They also report feeling unsafe in the city, on the streets, on public transport, and even in their house. Last, they report leaving their home more frequently.

Overall, the results link future victimization with increased exposure to crime and show that victims and non-victims were differentially exposed to crime even before becoming victims. Future victims leave home more frequently and report more interpersonal conflicts—including conflicts with authorities, police, and strangers—than their neighbors in the same census tract and quarter. By comparing neighbors to each other, we can ensure that the results are not mechanically driven by living in dangerous areas. Instead they measure differences in mobility, social exposure, and embeddedness in conflictual interactions that the participation literature associates with higher political engagement (Gerber et al., 2011; Kam, 2012) and higher victimization risk

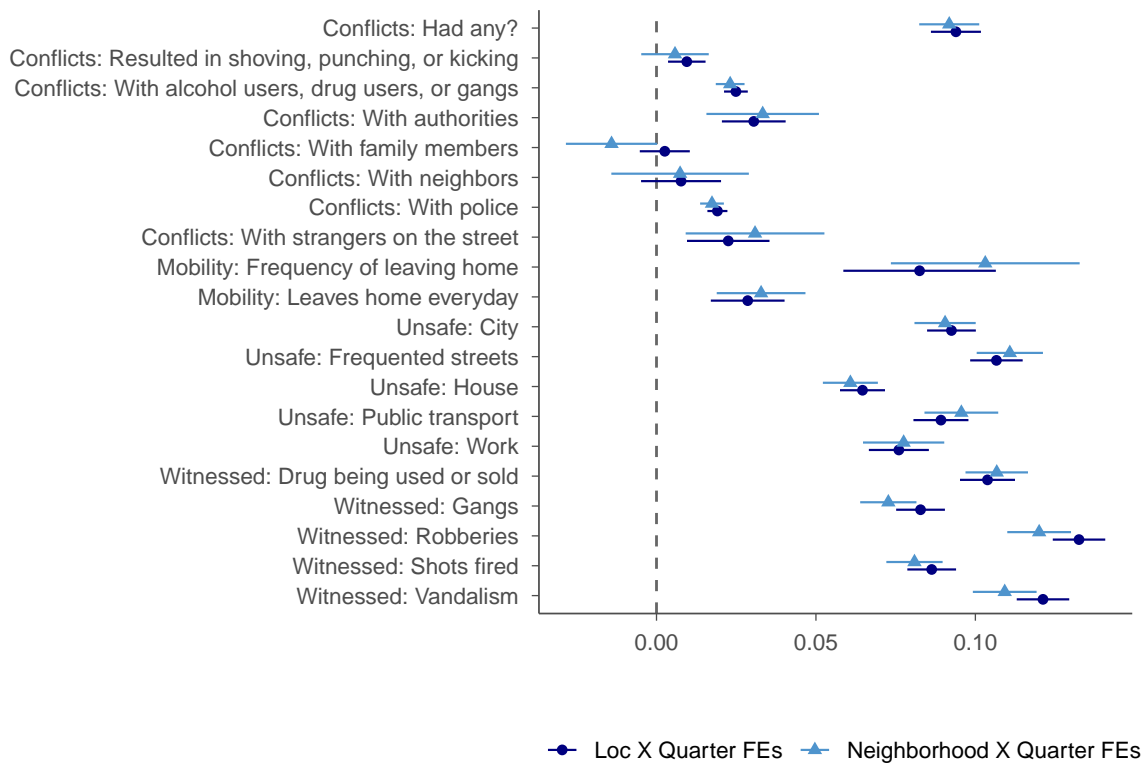


Figure 3: Cross-sectional differences between future victims and non-victims living in the same locality (dark blue) and the same neighborhood/census tract (light blue) on self-reported outcomes relating to exposure to crime. Robust standard errors clustered at the primary sampling unit. Full regression results in Table A13.

(Cawvey et al., 2018). Some of the remaining differences, particularly in witnessing crime, may partly reflect micro-geographic sorting within census tracts rather than individual behavioral profiles. But the mobility and conflict results are difficult to explain solely through geography.

## 4.2 Reactions to Crime

I next estimate how victimization changes attitudes and behavior among those who experience it.

$$PM_{i[q]} = \beta_1 * Victim_{i[q]} + \theta_i + \gamma_q + \epsilon_{i[q]} \quad (4)$$

where  $PM_{i[q]}$  is either a continuous measure of respondent  $i$ 's self-reported behavior, experiences, or attitude in year-quarter  $q$ , and neighborhood or locality  $g$ , or a binary measure that takes the value of 1 if respondent  $i$  reports that type of behavior, experience or attitude and 0 otherwise.  $Victim_{i[q]}$  is an indicator variable that takes the value of 1 when respondent  $i$  reports having been victimized in the past six months in quarter  $q$  and 0 otherwise. Because victimization is asked every six months, while response variables are asked every three months, each respondent's victimization report is extended to cover the prior period.  $\theta_i$  are individual fixed effects while  $\gamma_q$  are period fixed effects.  $\epsilon_{i[q]}$  are robust errors clustered at the primary sampling unit. I use a TWFE estimator of the ATT as well as the counterfactual estimator with equal weights proposed by Liu, Wang and Xu (2022).

Victimization changes attitudes and behaviors in a way consistent with demobilization, regardless of the preferred estimator (Figure 4). Compared with their behavior before becoming crime victims, respondents report leaving home less frequently, visiting friends and family less often, and taking more steps to prevent crime. They become less mobile and less likely to say they leave home every day. They trust law enforcement less, fear crime more, and adopt more constrained behaviors.

These behavioral and attitudinal changes are mechanisms through which demobilization is theorized to operate. Reduced mobility and social withdrawal limit the interpersonal interactions

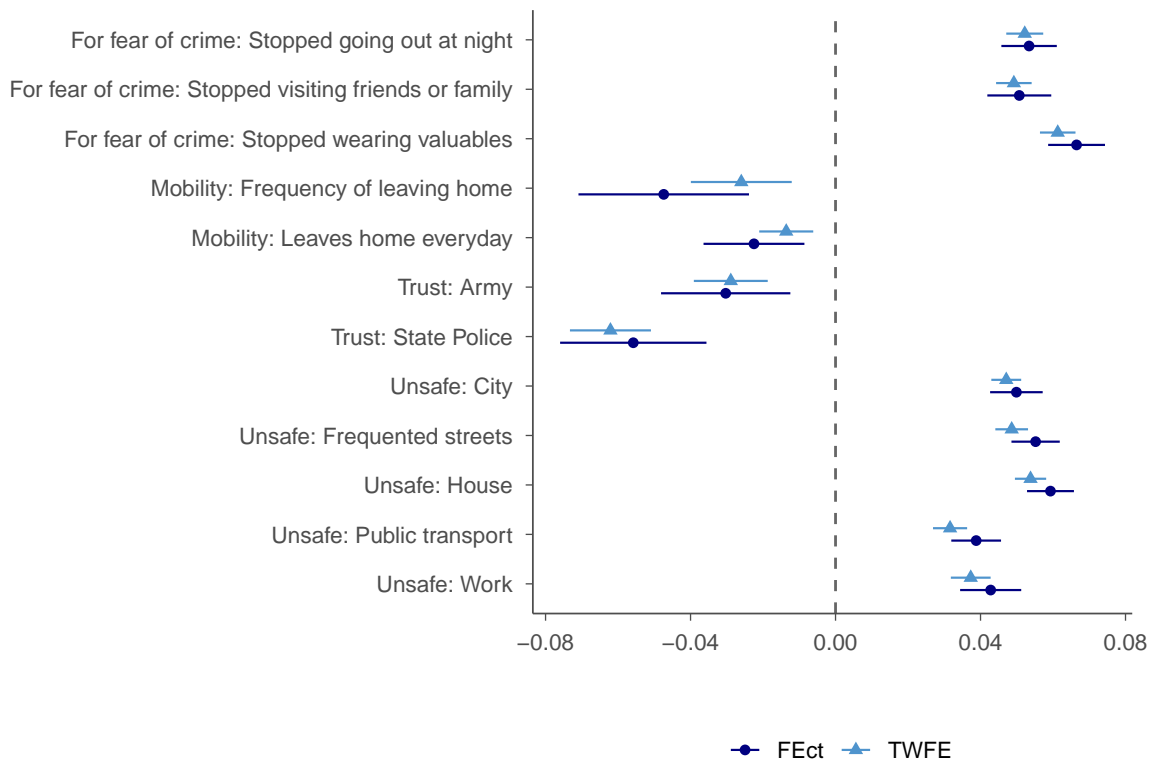


Figure 4: Within-individual effects of victimization on self-reported behavioral and attitudinal measures, estimated with the fixed-effect counterfactual estimator from Liu, Wang and Xu (2022) (dark blue) and a two-way fixed effects estimator (light blue). Robust standard errors clustered at the primary sampling unit. Full regression results in Table A14.

that facilitate political engagement (Gerber et al., 2011; Kam, 2012). Increased fear raises the perceived costs of public engagement (Ley, 2018; Skogan, 1986). ENSU does not measure political participation directly, but the within-individual changes it documents consistently move in the direction that the participation literature predicts should reduce engagement.

## 5 Disciplining the Identified Set with Panel Evidence

The LAPOP cross-section mapped the range of plausible effects in a pooled sample from Latin American countries. The ENSU panel showed that, in Mexico, individuals with the behavioral profile of demobilizers are more exposed to crime, and that victimization moves behavior in the demobilizing direction. If this pattern generalizes across the region, the remaining question is whether mobilizers are numerous enough relative to demobilizers to keep the ATT positive despite their lower exposure. To address this question, recall from Section 1 that whether the ATT agrees in sign with the ATE depends on two quantities:

1. The *differential selection ratio*,  $\lambda = \rho_U / \rho_T$ : how much more likely demobilizers are to be victimized compared to mobilizers. When  $\lambda = 1$ , both strata face identical victimization probabilities and sign agreement is guaranteed (Corollary 2). When  $\lambda > 1$ , demobilizers are disproportionately exposed to treatment.
2. The *stratum-size ratio*,  $\eta = \pi_T / \pi_U$ : how numerous mobilizers are relative to demobilizers. When  $\eta > 1$ , mobilizers outnumber demobilizers in the population.

When  $ATE > 0$ , Proposition 3 establishes that the ATT is negative if and only if  $\lambda > \eta$ . That is, the ATT reverses sign relative to the ATE when the overexposure of demobilizers exceeds their relative scarcity. The ENSU panel showed that in Mexico, future victims are systematically more exposed than non-victims before victimization, and that victimization moves behavior in the demobilizing direction within individuals. Together, these findings support the conclusion that  $\lambda > 1$  for Mexico: demobilizers likely face a higher risk of victimization than mobilizers.

I assume this directional condition holds across Latin America, without imposing any additional assumptions on the magnitude of  $\lambda$ . Given  $\lambda > 1$ , the sign of the ATT depends entirely on  $\eta$ , the ratio of mobilizers to demobilizers in the population. Since  $ATE = \pi_T - \pi_U$ , when the ATE is small, the two groups are of comparable size and  $\eta$  is close to 1. Published ATE estimates for non-electoral participation are on the order of 1 to 4 percentage points (Bateson, 2012). Small ATEs imply that mobilizers do not vastly outnumber demobilizers. The question is then whether this modest imbalance in the population is enough to sustain a positive ATT when demobilizers are overexposed.

I ask how restrictive the bound on  $\eta$  must be for *all* of the millions of plausible ATTs compatible with the data to be negative. When 100% of compatible vectors agree on direction, the ATT is *sign-identified*. I fix  $\lambda > 1$  (the directional condition from ENSU), restrict attention to LAPOP-compatible parameter vectors with  $ATE > 0$ , and gradually tighten the upper bound on  $\eta$ . For each bound, I compute the proportion of vectors where  $ATT < 0$ . Table 5 shows the results.

	Bound on $\eta = \pi_T/\pi_U$ (mobilizers per demobilizer)				
	Unrestricted	$\leq 2$	$\leq 1.5$	$\leq 1.1$	$\leq 1.04$
<i>Share of feasible vectors with <math>ATT &lt; 0</math>, given <math>ATE &gt; 0</math> and <math>\lambda &gt; 1</math></i>					
Protest	0.90	0.98	0.99	>0.99	<b>Signed –</b>
Community Problem-Solving	0.71	0.80	0.87	0.97	>0.99
Community Meetings	0.79	0.89	0.93	0.97	0.98
Political Meetings	0.89	0.97	0.98	>0.99	<b>Signed –</b>

Table 5: Sensitivity of ATT sign identification to restrictions on the stratum-size ratio. Cells report the share of LAPOP-compatible parameter vectors with  $ATT < 0$ , conditional on  $ATE > 0$  and  $\lambda > 1$ . Columns impose progressively tighter upper bounds on  $\eta = \pi_T/\pi_U$ .

Even without restricting  $\eta$ , imposing only  $\lambda > 1$ , between 71% and 90% of parameter vectors yield a negative ATT across the four non-electoral outcomes. At  $\eta \leq 1.5$ , all four outcomes exceed

87%. At  $\eta \leq 1.1$ , protest and political meetings exceed 99%. At  $\eta \leq 1.04$ , protest and political meetings are signed negatively. Since published ATE estimates are small, on the order of 1 to 4 percentage points, mobilizers cannot vastly outnumber demobilizers, placing  $\eta$  well within this range. Overall, results show that the mobilization finding for non-electoral participation does not hold up when shifting from the population average to those actually victimized. A corresponding analysis for voting, where the sign-flip runs in the opposite direction, is reported in Appendix A4.1.

## **6 Applications to other Forms of Violence**

Criminal victimization is not the only form of violence where selective exposure and heterogeneous responses plausibly coincide. I apply the partial identification procedure to civil war abduction in Uganda and police contact in the United States. Unlike the crime application, I cannot constrain the identified sets with panel evidence in these settings. The extensions, therefore, cannot establish sign reversal with the same force. What they can establish is structural consistency. In both settings, the region of the parameter space where the ATE and ATT disagree in sign is large, and the direction of disagreement mirrors the crime application.

### **6.1 Civil War Violence**

Studies of civil war violence have generally found that exposure increases political participation, often attributing the increase to post-traumatic growth and increased prosociality (Bellows and Miguel, 2009; Blattman, 2009; Voors et al., 2012). Complementary work emphasizes that civil war violence can also reshape political identities, generating durable forms of political engagement rooted in conflict experience (Balcells, 2012). One influential paper on the participatory consequences of civil war violence is Blattman (2009), which examines how forced recruitment into the LRA insurgency in Northern Uganda shaped downstream political participation. Leveraging quasi-random variation in abduction, the author targets the ATE and finds that abduction significantly increased voting and doubled the likelihood of becoming a community leader.

I apply the partial identification procedure to data from Blattman (2009) to derive identified

sets of ATEs and ATTs. Detailed results are reported in Section A6.1. For electoral participation, 77% of the ATEs in the identified set are positive, and 75% of the ATTs are as well — unlike the LAPOP results, the two estimands largely agree. For non-electoral participation, however, the pattern mirrors the crime application. Only 35% of the identified ATEs imply that abduction increased community leadership, despite the reported point estimate being positive and significant. And between 93% and 96% of the identified ATTs are negative. If abduction were random, it would likely increase voting and some forms of civic engagement. If it was selectively experienced, the identified set suggests it may have depressed non-electoral participation among the abducted.

Under random abduction, the ATE equals the ATT (Corollary 1), and the positive estimate corresponds to both. The partial identification exercise relaxes this assumption and assesses what would happen if abduction were even moderately selective, with those most exposed to abduction also most likely to demobilize. In that case, the effect on non-electoral participation among the treated is plausibly negative, even if the population-average effect is positive. The partial identification exercise does not adjudicate whether abduction was random, nor does it imply that the original research design was flawed. Instead, it clarifies the stakes of that assumption for the substantive conclusion.

## **6.2 Contact with the Police**

Research on the political consequences of police contact presents competing findings. Some work documents that contact teaches individuals they are politically marginal, suppressing engagement (Weaver and Lerman, 2010; Soss and Weaver, 2017). Others find that police contact, whether direct or indirect, stimulates participation (Walker, 2014, 2020). Police engage in statistical discrimination, and police contact is patterned (Pierson et al., 2020; Ba et al., 2021; Mummolo, 2017), making exposure in this setting selective.

Walker (2020) finds that direct and indirect police contact is associated with higher rates of both electoral and non-electoral participation, targeting the ATE as the primary estimand. I apply

the partial identification procedure to one of the data sources used, the National Crime and Politics Survey, a nationally representative U.S. survey conducted in 2013. I focus on direct police contact and use all available measures of electoral and non-electoral participation. Results are reported in Table A12.

The pattern mirrors the crime application. While most parameter sets show positive ATEs for non-electoral participation, the ATTs are predominantly negative. For voting, the pattern reverses again: 89% of parameter sets yield negative ATEs, while only 7% yield negative ATTs. If police contact were random, it would increase non-electoral participation. If it is selectively experienced in ways that coincide with heterogeneity in reactions, as the abundant literature suggests it is, the identified set suggests it depresses non-electoral engagement among those most frequently stopped while increasing their electoral participation.

## 7 Guidance for Researchers

In this paper, I showed that the substantive conclusion of an empirical analysis can reverse depending on the population over which the causal effect is defined. Specifically, the ATE and the ATT can carry opposite signs when exposure is selective, and responses are heterogeneous. This section distills the insights from the framework and the results into guidance for empirical research, applicable to the study of violence and beyond.

**1. Choose your estimand.** The estimand links theory to empirical analysis by specifying the population for which a causal effect is defined and the counterfactual comparison supported (Lundberg, Johnson and Stewart, 2021). The research design determines which estimand the analysis recovers. Sometimes a design allows researchers to recover multiple estimands. At a minimum, researchers should know which estimand the design delivers and why that estimand is relevant to the theoretical claim.

**2. When responses are heterogeneous and exposure is selective, target the ATT.** The ATE and the ATT are both weighted averages of stratum-specific treatment effects, differing only in their

weights; the ATE weights each stratum by its share in the population, while the ATT weights each stratum by its share among those who are treated. The two coincide when treatment is assigned independently of stratum membership and diverge under selective exposure; their signs disagree when the asymmetry in stratum-specific exposure exceeds the asymmetry in stratum-specific population shares (formally, when  $\lambda > \eta$ ). For harmful treatments where this divergence is plausible, the ATT is the substantively meaningful contrast. It tells us, on average, what becomes of the people the treatment in fact reaches. The ATE, by contrast, answers a counterfactual in which everyone is treated, an outcome no one would seek to bring about for treatments of this class.

**3. When the ATT cannot be point-identified, or when the link between the ATE and the ATT is the question, use partial identification on any binary treatment-outcome pair.** Four moments from cross-sectional data,  $Pr(Z = 1)$ ,  $Pr(Y = 1)$ ,  $Pr(Y = 1|Z = 1)$ , and  $Pr(Y = 1|Z = 0)$ , are sufficient to characterize the full set of (ATE, ATT) pairs consistent with the data. While this set always contains pairs with both possible signs (Appendix Appendix A), researchers can use the set to characterize the structural relationship between the ATE and the ATT across all configurations of selective exposure, and to ask what values of  $\eta$  (the relative size of mobilizing and demobilizing response types) and  $\lambda$  (their differential exposure to treatment) would be required for the two estimands to share a sign. While the procedure is defined only for binary treatment and outcomes, continuous variables can be dichotomized. Computational details and replication code appear in Appendix C.

**4. When building on the literature, compare effects defined for similar populations.** When engaging with the empirical literature, researchers should consider whether the studies under comparison define effects for the same population. With selective exposure, apparent disagreement across studies may reflect differences in the estimands targeted rather than in mechanisms or contexts. Stating the estimand explicitly clarifies whether findings are commensurable, which body of literature the study contributes to, and what the prior expectation for effects should be.

## 8 Conclusion

Most of what we know about the political consequences of violence from empirical causal research comes from counterfactuals in which everyone is exposed to a treatment that falls on a select few. In this paper, I showed that the population-average effect of violence on political participation and the effect among the exposed can yield opposite answers. To make the stakes of this distinction concrete, I reassessed the literature on criminal victimization in Latin America. That literature has read positive population-average effects on non-electoral participation as evidence that violence produces a more engaged citizenry.

Yet I find that those who become victims of crime were already more exposed than their neighbors before victimization, and that they react to it by becoming less mobile, more fearful, and less trusting of law enforcement. These are the behavioral changes that the participation literature associates with withdrawal from political life. Violence could well mobilize on average if everyone were equally exposed to it. But the victimized are often the vulnerable, and among them, violence is more likely than not to demobilize.

That targeting is selective and that the vulnerable are disproportionately exposed is well established in the theoretical literature on violence. What has not been established is the consequence for empirical causal research. When these features coincide, the ATE and ATT can yield estimates of opposite sign. Empirical research on the consequences of violence should therefore be designed around estimands that account for heterogeneous treatment probabilities and the possibility of divergent responses to treatment. A better understanding of who is exposed to violence, and how they differ from those who are not, is central to any account of its political consequences. Future research should pursue what structural features produce differential exposure, who is most exposed, and whether exposure to violence reinforces other forms of structural inequality.

Whether violence erodes or strengthens civic engagement among those who experience it determines whether affected communities can hold their governments accountable. If the victim-

ized withdraw from political life while the population average suggests mobilization, the political system registers the preferences of those least affected by violence, not those most in need of a governmental response. What is true of violence is true of any treatment that falls selectively and that we would not seek to universalize. For this class of treatments, the effect among the exposed is the quantity of first-order interest.

## References

- Anoll, Allison P. 2022. *The Obligation Mosaic: Race and Social Norms in U.S. Political Participation*. Chicago: University of Chicago Press.
- Ba, Bocar A., Dean Knox, Jonathan Mummolo and Roman Rivera. 2021. “The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago.” *Science* 371(6530):696–702.
- Balcells, Laia. 2012. “The Consequences of Victimization on Political Identities: Evidence from Spain.” *Politics & Society* 40(3):311–347.
- Balcells, Laia and Jessica A. Stanton. 2021. “Violence Against Civilians During Armed Conflict: Moving Beyond the Macro- and Micro-Level Divide.” *Annual Review of Political Science* 24:45–69.
- Bateson, Regina. 2012. “Crime Victimization and Political Participation.” *American Political Science Review* 106(3):570–587.
- Bellows, John and Edward Miguel. 2009. “War and Local Collective Action in Sierra Leone.” *Journal of Public Economics* 93(11):1144–1157.
- Blanco, Luisa R. 2013. “The Impact of Crime on Trust in Institutions in Mexico.” *European Journal of Political Economy* 32:38–55.
- Blattman, Christopher. 2009. “From Violence to Voting: War and Political Participation in Uganda.” *American Political Science Review* 103(2):231–247.
- Boggs, Sarah L. 1965. “Urban Crime Patterns.” *American Sociological Review* 30(6):899–908.
- Boulding, Carew, Shawna Mullenax and Kathryn Schauer. 2022. “Crime, Violence, and Political Participation.” *International Journal of Public Opinion Research* 34(1):edab032.

- Brady, Henry E., Sidney Verba and Kay Lehman Schlozman. 1995. "Beyond Ses: A Resource Model of Political Participation." *The American Political Science Review* 89(2):271–294.
- Browning, Christopher R., Nicolo P. Pinchak and Catherine A. Calder. 2021. "Human Mobility and Crime: Theoretical Approaches and Novel Data Collection Strategies." *Annual Review of Criminology* 4(Volume 4, 2021):99–123.
- Cawvey, Matthew, Matthew Hayes, Damarys Canache and Jeffery J. Mondak. 2018. "Personality and Victimization in the Americas." *International Review of Victimology* 24(1):123–139.
- Cohen, Lawrence E. and Marcus Felson. 1979. "Social Change and Crime Rate Trends: A Routine Activity Approach." *American Sociological Review* 44(4):588–608.
- Coppock, Alexander and Donald P. Green. 2016. "Is Voting Habit Forming? New Evidence from Experiments and Regression Discontinuities." *American Journal of Political Science* 60(4):1044–1062.
- URL:** <https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12210>
- Dorff, Cassy. 2017. "Violence, Kinship Networks, and Political Resilience: Evidence from Mexico." *Journal of Peace Research* 54(4):558–573.
- Duarte, Guilherme, Noam Finkelstein, Dean Knox, Jonathan Mummolo and Ilya Shpitser. 2024. "An Automated Approach to Causal Inference in Discrete Settings." *Journal of the American Statistical Association* 119(547):1778–1793.
- Egami, Naoki and Erin Hartman. 2023. "Elements of External Validity: Framework, Design, and Analysis." *American Political Science Review* 117(3):1070–1088.
- Frangakis, Constantine E. and Donald B. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58(1):21–29.

- Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling, Connor Raso and Shang E. Ha. 2011. "Personality Traits and Participation in Political Processes." *The Journal of Politics* 73(3):692–706.
- Gilligan, Michael J., Benjamin J. Pasquale and Cyrus Samii. 2014. "Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal." *American Journal of Political Science* 58(3):604–619.
- Gottfredson, Michael R. 1986. Substantive Contributions of Victimization Surveys. In *Crime and Justice: An Annual Review of Research*, ed. Michael Tonry and Norval Morris. Vol. 7 Chicago: University of Chicago Press pp. 251–287.
- Guedes, Inês Maria Ermida Sousa, Sofia Patrícia Almeida Domingos and Carla Sofia Cardoso. 2018. "Fear of Crime, Personality and Trait Emotions: An Empirical Study." *European Journal of Criminology* 15(6):658–679.
- Hale, C. 1996. "Fear of Crime: A Review of the Literature." *International Review of Victimology* 4(2):79–150.
- Heckman, James J. and Edward Vytlacil. 2005. "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica* 73(3):669–738.
- INEGI, Instituto Nacional de Estadística y Geografía. 2024. "Encuesta Nacional de Seguridad Pública Urbana (ENSU).".
- Kam, Cindy D. 2012. "Risk Attitudes and Political Participation." *American Journal of Political Science* 56(4):817–836.
- Kline, Brendan and Elie Tamer. 2023. "Recent Developments in Partial Identification." *Annual Review of Economics* 15:125–150.

**URL:** <https://www.annualreviews.org/content/journals/10.1146/annurev-economics-051520-021124>

Knox, Dean, Will Lowe and Jonathan Mummolo. 2020. “Administrative Records Mask Racially Biased Policing.” *American Political Science Review* 114(3):619–637.

Koos, Carlo and Richard Traunmüller. 2024. “The Gendered Costs of Stigma: How Experiences of Conflict-Related Sexual Violence Affect Civic Engagement for Women and Men.” *American Journal of Political Science* 69(2):763–778.

Kronick, Dorothy. 2014. Electoral Consequences of Violent Crime: Evidence from Venezuela. Working Paper 687 CAF Development Bank of Latin America.

**URL:** <https://ideas.repec.org/p/dbl/dblwop/687.html>

LAPOP, Lab. 2022. “The AmericasBarometer.” [www.vanderbilt.edu/lapop](http://www.vanderbilt.edu/lapop).

Ley, Sandra. 2018. “To Vote or Not to Vote: How Criminal Violence Shapes Electoral Participation.” *Journal of Conflict Resolution* 62(9):1963–1990.

Ley, Sandra. 2022. “High-Risk Participation: Demanding Peace and Justice amid Criminal Violence.” *Journal of Peace Research* 59(6):794–809.

Liu, Licheng, Ye Wang and Yiqing Xu. 2022. “A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data.” *American Journal of Political Science* 68(1):160–176.

Lundberg, Ian, Rebecca Johnson and Brandon M. Stewart. 2021. “What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory.” *American Sociological Review* 86(3):532–565. Publisher Copyright: © American Sociological Association 2021.

Manski, Charles F. 1990. “Nonparametric Bounds on Treatment Effects.” *American Economic Review* 80(2):319–323.

Marshall, John. 2022. "Tuning In, Voting Out: News Consumption Cycles, Homicides, and Electoral Accountability in Mexico."

**URL:** [https://john-l-marshall.github.io/files/tuning\\_in\\_voting\\_out.pdf](https://john-l-marshall.github.io/files/tuning_in_voting_out.pdf)

Miethe, Terance D., Mark C. Stafford and J. Scott Long. 1987. "Social Differentiation in Criminal Victimization: A Test of Routine Activities/Lifestyle Theories." *American Sociological Review* 52(2):184–194.

Mummolo, Jonathan. 2017. "Modern Police Tactics, Police-Citizen Interactions, and the Prospect for Reform." *Journal of Politics* 80(1):1–15.

Pierson, Emma, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff and Sharad Goel. 2020. "A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States." *Nature Human Behaviour* 4(7):736–745.

Plutzer, Eric. 2002. "Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood." *American Political Science Review* 96(1):41–56.

Rader, Nicole. 2004. "The Threat of Victimization: A Theoretical Reconceptualization of Fear of Crime." *Sociological Spectrum* 24(6):689–704.

Samii, Cyrus, Laura Paler and Sarah Zukerman Daly. 2017. "Retrospective Causal Inference with Machine Learning Ensembles: An Application to Anti-recidivism Policies in Colombia." *Political Analysis* 24(4):434–456.

Skogan, Wesley. 1986. "Fear of Crime and Neighborhood Change." *Crime and Justice* 8:203–229.

Skogan, Wesley G. 1982. Methodological Issues in the Measurement of Crime. In *The Victim in International Perspective: Papers and Essays Given at the "Third International Symposium on Victimology" 1979 in Münster/Westfalia*, ed. Hans Joachim Schneider. de Gruyter.

- Skogan, Wesley G. 1995. "Crime and the Racial Fears of White Americans." *The Annals of the American Academy of Political and Social Science* 539:59–71.
- Sønderskov, Kim Mannemar, Peter Thisted Dinesen, Steven E. Finkel and Kasper M. Hansen. 2022. "Crime Victimization Increases Turnout: Evidence from Individual-Level Administrative Panel Data." *British Journal of Political Science* 52(1):399–407.
- Soss, Joe and Vesla Weaver. 2017. "Police Are Our Government: Politics, Political Science, and the Policing of Race–Class Subjugated Communities." *Annual Review of Political Science* 20(1):565–591.
- Trelles, Alejandro and Miguel Carreras. 2012. "Bullets and Votes: Violence and Electoral Participation in Mexico." *Journal of Politics in Latin America* 4(2):89–123.
- Vilalta, Carlos. 2020. "Violence in Latin America: An Overview of Research and Issues." *Annual Review of Sociology* 46(1):693–706.
- Visconti, Giancarlo. 2020. "Policy Preferences after Crime Victimization: Panel and Survey Evidence from Latin America." *British Journal of Political Science* 50(4):1481–1495.
- Voors, Maarten J., Eleonora E. M. Nillesen, Philip Verwimp, Erwin H. Bulte, Robert Lensink and Daan P. Van Soest. 2012. "Violent Conflict and Behavior: A Field Experiment in Burundi." *American Economic Review* 102(2):941–964.
- Walker, Hannah L. 2014. "Extending the Effects of the Carceral State: Proximal Contact, Political Participation, and Race." *Political Research Quarterly* 67(4):809–822.
- Walker, Hannah L. 2020. "Targeted: The Mobilizing Effect of Perceptions of Unfair Policing Practices." *The Journal of Politics* 82(1):119–134.
- Weaver, Vesla M. and Amy E. Lerman. 2010. "Political Consequences of the Carceral State." *American Political Science Review* 104(4):817–833.