Moral Violence: Unbundling Social Preferences At the Heart of a Major Armed Group in Congo*

Louis Dunia† **Raúl Sánchez de la Sierra**‡ **Hilary Yu**§

This version: February 2025
Latest version: Here

Abstract: Economists generally view prosocial preferences as a foundation for cooperation. Yet, they can also fuel lethal violence. We unbundle social preferences into preferences over *outcomes* (others' welfare) and preferences over *acts* (violence against others). We embedded a real-time notification system within the HR division of a major militia in the DRC, allowing us to survey future joiners before enlistment alongside non-joiners. Three patterns reveal a link from prosociality to lethal violence. First, prior to enlistment, joiners, compared to non-joiners, show greater outcome-prosociality for their collective and lower violence aversion for enemy members—via *self-deception*. Psychophysiological data and militia supervisor evaluations of joiners' behavior after enlisting validate this pattern. Second, moral dilemma experiments reveal a prevailing '*morality of war*' that defines enlistment as outcome-prosocial for collective defense and the act of lethal violence against enemy members—even genocidal acts—as a moral duty of collective retaliation and liberation. This leads those most outcome-prosocial toward the collective to suppress their violence aversion toward enemy members. Third, following enlistment, such aversion declines further through *desensitization*. These findings suggest that the canonical conflict-as-bargaining framework and, more broadly, welfare analysis could be expanded to incorporate preferences over *acts*, such as violence, shaped by the moral frameworks of participants—rather than the observer's. **JEL Codes:** D91, O10, D74.

*IRB approval was obtained from the University of Chicago. This project was supported by the Wess Fund, JPAL, IPA, the IGC, the Pearson Institute for the study of conflicts, the Becker Friedman Institute, the Kiphart Center for Public Health at the University of Chicago, and the Harry Frank Guggenheim Foundation. We are indebted for excellent research assistance to the 23 researchers who collected the data, and to Kashif Ahmed, Sebastien Annan-Phan, Rimin Cherian, Soeren Henn, Erica Hogan, Christelle Inema, Benjamin Krause, Harrison Lee, Hersheena Rajaram, Ghifari Firman, Miguel Ortiz, Adidth Pillai, Yeduo Pu, Franziska Schwingeler, Lucina Rodriguez Guillen, Sol Rivas Lopez, Nancy Smith, Vincent Tanutama, Mwangi Thuita, Tyler Thompson, Zeray Wu, Jiemin Xu, Blair (Xinyi) Yao, and Yinghui Zhou for excellent research assistance, and especially to Akanksha Kala, Margot Barranco, Clara Brandt, Jan Fasnacht, Sebastian Hernandez Leiva and Stan Xie for their relentless yearlong support and commitment, without whom it would not have been possible. We thank Sam Altman, Moustafa El-Kashlan, Jan Fasnacht, Simon Franklin, Lucie Gadenne, Sara Lowes, Gauthier Marchais, Eduardo Montero, Nathan Nunn, Sandra Polanyía-Reyes, Jonathan de Quidt, James Robinson, Noam Yuchtman, and participants of BFI's Institutions Organizations and Growth, ERINN's conference, Brown development seminar, Boston University's development seminar, London School of Economics/Bocconi crime seminar, Stockholm IIES development seminar, Harvard Alberto Alesina seminar, and Queen Mary University of London Applied Micro seminar as well as Jean-Francois Bonnefon and Jean Decety for suggestions.

†Independent Researcher, Democratic Republic of the Congo.

‡University of Chicago and NBER. Corresponding author. Email: raul@uchicago.edu

§University of California, Berkeley

1. Introduction

Economists typically view participation in violence as driven by material incentives—akin to rent-seeking (Becker, 1968, Besley and Persson, 2011). Yet across disciplines—from anthropology to evolutionary biology—violence is often seen as a product of prosocial behavior: the drive to protect and serve one's group (Bowles and Gintis, 2011). This tension cuts to a fundamental puzzle at the intersection of economics and the social sciences: humans are both uniquely cooperative and uniquely violent—two capacities thought to have co-evolved (Pinker, 2011, Gómez, Verdú, González Megías, and Méndez, 2016, Choi and Bowles, 2007, Wrangham, 2019). Could the same preferences that sustain cooperation also harbor the potential for lethal violence?

This paradox is not merely theoretical; it is echoed in the accounts of those who enlist. Consider Lewis, born in 1984 in Eastern Congo (Butinda, 2021). At 11, while foreign armed groups violently dispossessed his district and the state was absent to defend them, he picked up an AK-47 for the first time, thinking: "If we do not do anything, we are going to be exterminated." His enlistment—which continued for over two decades—was not driven by gain, but by what he experienced as a duty toward his collective. His is not an isolated case. Across militias, national armies, and self-defense groups, individuals who enlist willingly—knowing they will engage in lethal violence—often describe their choice not as self-interest but as "*service*" and "*duty*."

Yet, despite qualitative evidence across disciplines, we lack evidence on the role of *social preferences* in participation in violence. Two challenges stand in the way. First, existing concepts of social preferences in economics focus on *outcomes*—to improve or worsen others' welfare. But while enlisting is often experienced as a prosocial *service* to a collective, the *act of killing* bundled within that service imposes unique psychological costs tied to empathy. Their anticipation might influence the decision to enlist and may systematically vary with motivation for service.

Second, violent organizations operate in secrecy. This forces research to rely on ex-combatants, risking survivor bias and conflating selection with the effects of participation, or on historical data from national armies, with little measurement of social preferences. As a result, we know little about how social preferences shape who enlists—the stage where such preferences are revealed.

This paper addresses these challenges with two innovations. First, we develop a conceptual framework that unbundles social preferences into preferences over outcomes (others' welfare) and preferences over acts (violence against others), distinguishing both by their targets, and formalizing how these preferences jointly shape enlistment decisions. Second, we embed a real-

time notification system within the HR division of a major Congolese militia, allowing us to track enlistment decisions free from survivor bias. Together, they reveal that social preferences, when structured by moral values, can selectively suppress empathy and drive lethal violence.

We address the conceptual challenge by unbundling social preferences. First, we distinguish: (a) preferences over *outcomes* (over others' welfare), and (b) preferences over the *act* of violence against others, capturing an aversion or inclination toward perpetrating interpersonal violence, drawing on ideas from moral psychology (Decety and Jackson, 2004a, Decety, 2011). Second, we distinguish each of these by the identity of their target. Enlisting generates a payoff with two additive components: (1) outcome preferences toward those perceived to benefit from violence (henceforth, *the collective*), a form of mission motivation (Besley and Ghatak, 2005a); and (2) violence aversion toward enemy members (henceforth, *the enemy*). Enlistment follows a Roy-like selection process, driven by the correlation between these two components, which determines whether outcome-prosocial types are violence averse toward enemies, and by the ratio of their variances, which governs the scope for sorting—henceforth, the *key relationship*.

We addressed the logistical challenge by embedding a real-time notification system within the HR division of a major Congolese militia (hecenfroth, *the group*). From July 2019 to April 2020, we trained all 55 HR commanders (henceforth, *S1s*) and their 79 assistants, establishing a system that generated real-time phone notifications with the identity and location of individuals who contacted an S1 *intending* to enlist, along with information on their households. This allowed us to survey all joiners *before enlistment* between January and September 2022 ($n = 213$), alongside young adults randomly sampled from (1) their households ($n = 213$) and (2) their villages ($n = 213$, across 133 villages). We cross-referenced these notifications with the group's administrative records, confirming that this system captured the universe of joiners during that period.

The group is well-suited to provide a proof of concept for the role of social preferences in violence. First, membership is voluntary, and all volunteers are accepted, making enlisting a revealed preference. Second, the group, which had 2,789 members as of May 2023, enrolls an average of 36 new members monthly, allowing us to study selection within a feasible time frame. Third, the group frames its mission as defending its district—a multi-ethnic area larger than Switzerland (henceforth, *the District*)—from a Rwandan armed group, whose 20 years of violent dispossession are well-documented. It is thus framed as a contribution to the District's welfare.

Leveraging this system, we collect two sets of *measures* among joiners and non-joiners: (a) outcome preferences—both pro- and anti-social—using globally validated experimental games

and survey tools (Falk, Becker, Dohmen, Huffman, and Sunde, 2022); and (b) *violence preferences*, which capture both violence *aversion* and violence *inclination*. Violence aversion is measured through the affective-motivational components of empathy (henceforth, *A.M. empathy*), comprising *empathic concern*—an emotional response to others' well-being—and *affective empathy*—sharing another's emotional state—both elicited in a *suffering empathy task* (Ferguson and Inzlicht, 2022). We measure *violence inclination* by adapting the *suffering empathy task* to elicit *enjoyment of others' pain* and by developing a video game to capture revealed violence inclination. *Manifest* violence preferences rest on *cognitive empathy*—beliefs about others' emotional states—which, in the *suffering empathy task*, captures beliefs about others' pain and suffering. All measures are collected toward the collective (operationalized as "the District') and the enemy (operationalized as "Rwandan Hutus")—henceforth, *the District* and *the enemy*.

Using these unique data, our analysis presents three descriptive findings that provide a proof of concept for the role of social preferences, shaped by a '*morality of war*,' in driving violence.

Our first finding is that social preferences influence enlistment decisions. After documenting civilians' widespread low aversion to violence against the enemy, we establish this in three steps.

First, outcome prosocial preferences toward the District correlate positively with both outcome prosocial preferences toward the enemy and violence aversion toward the District but weakly negatively with violence aversion toward the enemy. This pattern, driven by cognitive empathy toward the enemy, suggests that those most outcome-prosocial toward the District downplay enemy suffering, thereby reducing their violence aversion—henceforth, *self-deception*.

Second, building on the correlation analysis, incorporating variance, the *key relationship* satisfies the condition that if social preferences structure enlistment payoffs, those who self-select into enlistment should exhibit stronger outcome prosocial preferences toward the District and disproportionately lower violence aversion toward the enemy—especially relative to young adults in their households. The data confirm this selection pattern. Furthermore, joiners' lower violence aversion toward the enemy is entirely driven by the joiners' differential *self-deceptio*n.

Third, this pattern of selection reflects the role of unbundled social preferences.

(i) These two components of social preferences are part of a broader selection structure. Joiners are younger, have fewer siblings, and are economically vulnerable. They are also more likely to belong to the dominant ethnic group, be Protestant, and exhibit signs of low mental health—yet they have comparable cognitive and noncognitive skills, marginalization, and social ties to the group. Still, social preferences independently predict enlistment, accounting for these covariates.

(ii) Joiners exhibit the same capacity for violence aversion as non-joiners. While they display lower A.M. empathy toward the enemy, their A.M. empathy toward District members is comparable. Underscoring this interpretation, they are no more likely to test positive for *psychopathy*. Their lower A.M. empathy toward enemy members persists after accounting for empathy toward the District. This suggests joiners' violence preferences are more "biased," disfavoring the enemy, but does not necessarily indicate that joiners' violence preferences are more parochial.

(iii) Joiners exhibit greater outcome prosociality toward the District than toward the enemy, but their prosociality toward the enemy is comparable to that of non-joiners. This pattern suggests that joiners' outcome preferences are more "biased," but does clearly correspond to differential outcome preference parochiality. Rather, they exhibit greater outcome prosociality on average.

(iv) The joiners' lower violence aversion to the enemy is not due to associating enemy identity to armament, as revealed by violence aversion to the District's armed group members.

Our second finding is that a prevailing *morality of war*, governing both enlistment and the act of violence, shapes the expression of *outcome prosociality* and *violence aversion* in enlistment.

The act of enlisting is best understood through a prevailing morality that defines enlistment as right based on its outcomes for the District—a moral service. Moral dilemmas reveal that 65% of civilians view enlistment as morally right. For many, this belief is grounded in the perceived consequences of enlistment, aligning with moral consequentialism. Joiners' first-, second-, and third-order beliefs indicate that enlistment is perceived as serving the collective defense of the District's territorial integrity— which is also their predominant motivation.

The act of violence toward the enemy, anticipated at enlistment and bundled within that service, is best understood through a prevailing morality that defines killing enemy members as a *moral duty*. Moral dilemmas reveal that two-thirds of civilians view killing an enemy member as morally right, even if the target is a harmless civilian, pregnant woman, or infant—groups often targeted in genocidal violence—or if it results in 100 enemy deaths—potentially aligning with *genocidal morality*. For 15% of civilians, this view remains unchanged even if it leads to large numbers of District civilian deaths—aligning with *collective narcissism*. In open-ended answers, 55% of civilians justify violence against enemy civilians—56% as collective retaliation and 10% as a means for liberation. Its roots lie in negative reciprocity, past family victimization, and moral disengagement (Bandura, 2011), including dehumanization. As hostile moral convictions can selectively suppress empathy (e.g., Workman, Yoder, and Decety (2020)), this is consistent with a morality of killing as the mechanism enabling the suppression of violence aversion.

This analysis of moral convictions helps interpret both enlistment and the *key relationship*. First, the selection of outcome-prosocial individuals toward the District and violence-inclined individuals toward the enemy is unaffected by inclusion of individual-level controls for these moral convictions. Thus, if this morality shapes selection, it is through defining social preferences. Second, the negative correlation between outcome prosociality toward the District and violence aversion toward the enemy is consistent with those most committed to the District upholding this morality more strongly, suppressing cognitive empathy toward the enemy.

Our third finding is that, after enlisting, violence aversion against the enemy declines further through *desensitization*. Tracking joiners eight months after enlistment and administering the same measures, we find that their violence aversion (as measured in A.M. empathy) toward the enemy declines and their violence inclination toward the enemy accentuates, while they fully retain *cognitive* empathy toward the enemy—which we label *desensitization*. Violence aversion toward District members increases, emphasizing that desensitization is also *selective*.

Having presented these findings, we empirically validate the measures of violence preferences. First, *violence aversion* is positively associated with *emotional arousal* upon observing the pain and suffering of others in the *suffering empathy task* measured in psychophysiological data (Avenanti, Sirigu, and Aglioti, 2010); furthermore, psychophysiological data replicates our findings on *violence aversion*. Second, violence preferences are strongly associated with contemporaneously measured violent behavior, based on supervisor and peer evaluations of joiners' conduct eight months after enlistment. Outcome preferences, in contrast, fail both validations.

Our findings make several contributions to economics. First, we provide the first survey-based measure of selection into violence free from survivor bias, overcoming logistical barriers that previously made this impossible across the social sciences and psychology (e.g., Humphreys and Weinstein (2008), Pape, Decety, Ruby, Albanez Rivas, Jessen, and Wegner (2017), Cerina, Barrie, Ketchley, and Zelin (2023), Benmelech and Klor (2020), Henn and Huff (2024)).

This access allows us to shift the focus beyond self-interest as the primary driver of violence in economics (Fearon, 1995, Besley and Persson, 2011), contributing to a growing literature on the role of social emotions in violence (Grosjean, 2014, Nunn, Cao, Enke, Falk, and Giuliano, 2021, Mugaruka, Marchais, Sánchez de la Sierra, and Wu, 2024). Combined with these results, our findings suggest that while threats to *private* property can foster cultures of *individual* honor, *existential collective* threats can foster a morality of *collective* retaliation and liberation, moralize acts of genocidal violence and potentially induce *collective narcissism*—mirroring individual honor.

Second, we provide the first empirical evidence for the hypothesis in evolutionary biology that human altruism is linked to our capacity for violence (Choi and Bowles, 2007, Bowles and Gintis, 2011, Wrangham, 2019). Conversely, although not designed to measure its effect, our finding of a "morality of war" reveals a new mechanism through which war fosters cooperation: the notion that war can restructure moral frameworks, including the sanctification of lethal violence (Bauer, Blattman, Chytilová, Henrich, Miguel, and Mitts, 2016, Hartman and Morse, 2020).

Third, our findings suggest that the perceived moral righteousness of coercive organizations' goals—and thus of the state, one of the most extensively studied coercive organizations (Tilly, 1985)—(a) facilitates "empathy" suppression among its members, providing a *moral* foundation for state *capacity* to exert violence (Besley and Persson, 2009, Athias, 2025); and (b) shapes recruitment, contributing to the literature on participation in national armies (Campante and Yanagizawa-Drott, 2015, Cagé et al., 2023). Our study complements seminal work by Fouka (2020) and Qian and Tabellini (2024), which shows how alignment with the social contract influences enlistment, by providing the first direct measures of individual preferences in this context.

Fourth, while economists increasingly differentiate social preferences by *identity* (e.g., Jha (2013), Jha and Shayo (2019), Tabellini (2020), Sánchez de la Sierra (2021), Cikara, Fouka, and Tabellini (2022), Fouka and Tabellini (2022), Enke, Rodriguez-Padilla, and Zimmermann (2023), Bernini, Facchini, Tabellini, and Testa (Forthcoming)), joining a broader literature in the social sciences (e.g., Habyarimana, Humphreys, Posner, and Weinstein (2007)), we introduce a new type of social preference, drawing on moral psychology (Zaki and Ochsner, 2012b). Incorporating preferences over the *act* of violence reveals a stronger *communitarian morality* in war. Yet, those outside the in-group are not merely disfavored but perceived as existential threats—enemies to be "eliminated," with "empathy" selectively suppressed—revealing a distinct conception of the out-group. Simultaneously, the production of collective goods such as defense can expand the in-group beyond traditionally communitarian groups (e.g., kin, family, or ethnicity), fostering a form of *universalist categorization*. Building on the moral universalism-parochialism spectrum articulated by Enke, Rodriguez-Padilla, and Zimmermann (2023), our findings reveal that war can unleash both simultaneously—intensifying exclusion while expanding the scope of inclusion.

Fifth, these findings contribute to the study of genocide (e.g., Arendt and Elon (2006), Zimbardo (2008), Browning (2017)) by providing evidence from interviews with perpetrators of lethal violence and direct measures of genocidal morality. These reveal a blurring of the cognitive boundary between enemy combatants and civilians, linking genocidal violence to prosociality.

2. How Social Preferences, Unbundled, Drive Selection into Violent Social Action

This section develops a parsimonious framework to illustrate how social preferences shape the decision to enlist in violent action when that decision may be perceived as advancing the welfare of a group of individuals and/or contributing to a cause. By unbundling social preferences into two components relevant to enlisting, the framework emphasizes that selection into violence depends not only on the *levels* of these preferences but also on their *covariance structure*. Even in its simplest formulation, the framework yields nontrivial, Roy-like selection implications for the social preferences of those who self-select into violence, informing the empirical analysis.

There is one period. There is a continuum of individuals. Each individual chooses whether to participate in violent action against an enemy (henceforth, *enlisting*). Enlisting benefits a set of individuals (henceforth, *the collective*).[1] The payoff when not enlisting is $U_0 = w_0$, where $w_0 \in \mathbb{R}^+$ is the exogenous wage if not enlisting. The payoff from enlisting is:

$$U_1 = w_1 + \underbrace{\bar{\theta} + \tilde{s}_C - \tilde{\lambda}_E}_{\text{Social Preferences}},$$

where $w_1 \in \mathbb{R}^+$ is the common wage upon enlisting, $\bar{\theta}$ is a common intrinsic payoff from enlisting, and $\tilde{s}_C$ and $\tilde{\lambda}_E$ are individual idiosyncratic deviations, both mean zero across individuals. Each individual's decision to enlist does not affect the draws $\tilde{s}_C$ and $\tilde{\lambda}_E$ of others.

The first, $\tilde{s}_C$, is the idiosyncratic deviation in preferences over *outcomes* for the Collective. It reflects the individual's draw of this deviation, capturing the intrinsic psychological reward from the perceived contribution of enlisting to the collective's welfare—akin to mission-driven intrinsic motivation (Besley and Ghatak, 2005b). The second, $\tilde{\lambda}_E$, is the idiosyncratic deviation in preferences over *the act of violence* toward the Enemy. It reflects the individual's draw of this deviation, representing the intrinsic utility or disutility associated with anticipating *perpetrating* violence—akin to a cost of effort, where effort after enlisting can be emotionally taxing.

These two components can be situated within a broader 2x2 partition of the preference space, defined by preferences over *outcomes* vs. *act* and the target of those preferences (*collective* vs. *enemy*). Because enlisting involves contributing to the collective's welfare and inflicting violence on the enemy, this exhibition draws only on the two components introduced above for simplicity. Panel A of Table 1 illustrates this partition. Panel B partitions individuals into sets along these components. Figure 1, upper-left quadrant, represents individuals in the $(\tilde{\lambda}_E, \tilde{s}_C)$ space.

[1] Unlike a *club good*, where access depends on selective membership, defense is best understood as a *collective good*. Although security benefits are technically excludable, mobility costs make this form of excludability less relevant.

Table 1: Unbundling Social Preferences and Classification of Individuals

Panel A: Unbundling Social Preferences into Outcomes/Acts and Target Identities

		Social Preference's Target	
		Toward the Collective	*Toward the Enemy*
Social Preference Domain	*Outcome*	✓	–
	Violence	–	✓

Panel B: Classification of Individuals Induced by Unbundling Social Preferences

		Violence Aversion Toward the Enemy, $\tilde{\lambda}_E$	
		High	*Low*
Outcome Prosociality Toward the Collective, $\tilde{s}_C$	*High*	Soft Prosocials*	"Zealots"*
	Low	Soft Selfish	Mercenaries*

Notes: Panel A represents the two components of social preferences relevant to enlisting (✓), situated within the 2x2 social preference space, defined by preferences over *outcomes* vs. *act* and the target of those preferences (*collective* vs. *enemy*). Panel B partitions individuals into sets along these components. Only three of those may self-select into enlisting (*). This section provides conditions for which type enlists.

An individual enlists if and only if enlisting maximizes their payoff:

$$U_1 > U_0 \iff \tilde{s}_C > w_0 - w_1 - \bar{\theta} + \tilde{\lambda}_E.$$

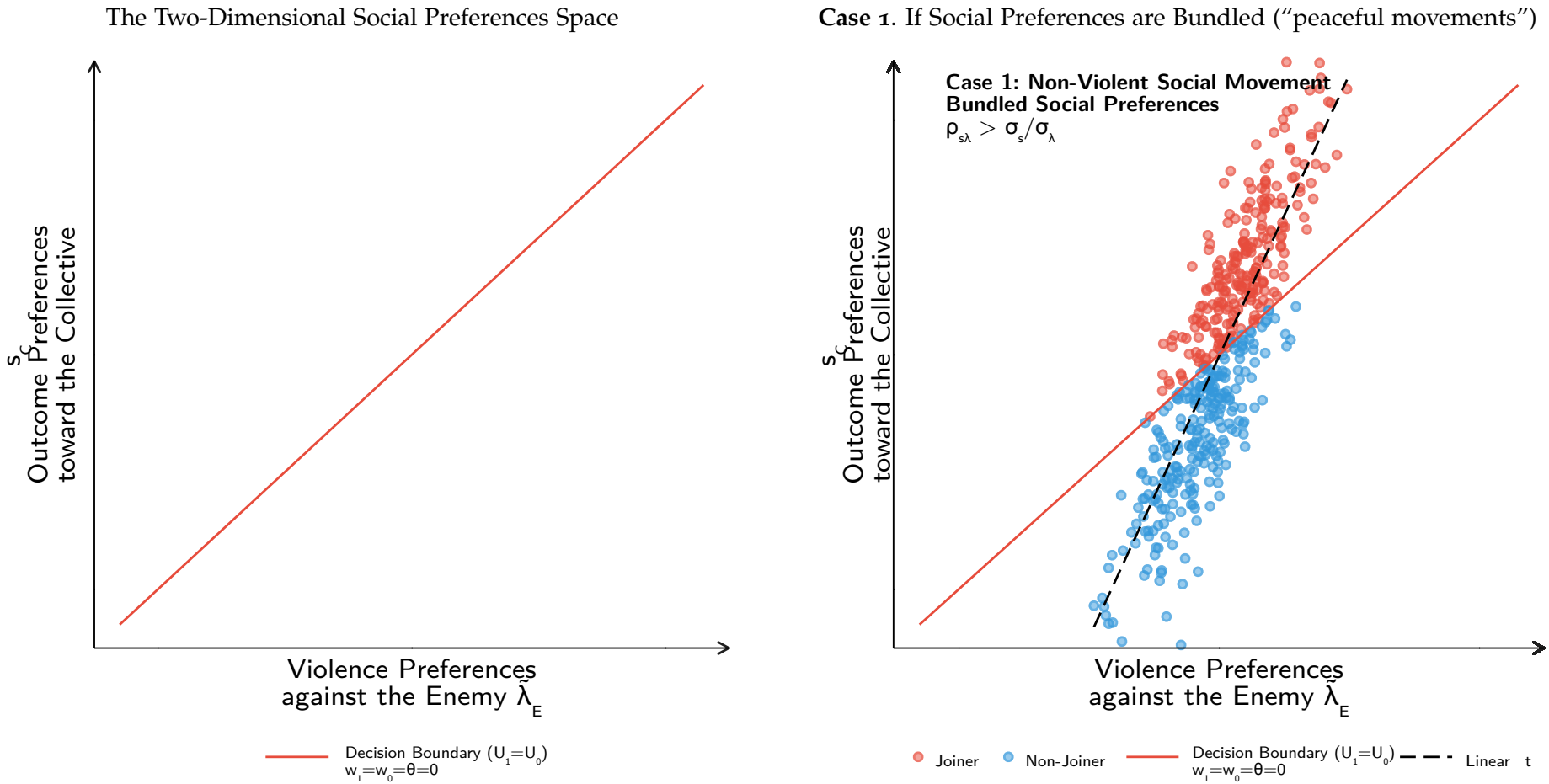
The enlistment threshold for $\tilde{s}_C$ increases with the wage opportunity cost $w_0 - w_1$ and decreases with both $\bar{\theta}$ and $\tilde{\lambda}_E$. The intuition is that the greater the individual idiosyncratic disutility from perpetrating violence against the enemy, the stronger the individual idiosyncratic outcome prosociality toward the collective's welfare must be to compensate for that disutility.

Who enlists depends not only on the *levels* of social preferences but also on their *covariance structure*. The correlation between outcome prosociality toward the collective and violence aversion toward the enemy is defined as $\rho_{s\lambda} = \frac{\sigma_{s\lambda}}{\sigma_\lambda \sigma_s}$. When $\rho_{s\lambda}$ is close to one, the two components largely reflect a common underlying trait (e.g., generic altruism), making unbundling less informative. When $\rho_{s\lambda}$ is lower, or negative, they represent distinct or even opposing dimensions of social preferences—unbundling is then essential for understanding selection.

Assumption A (Joint normal distribution)

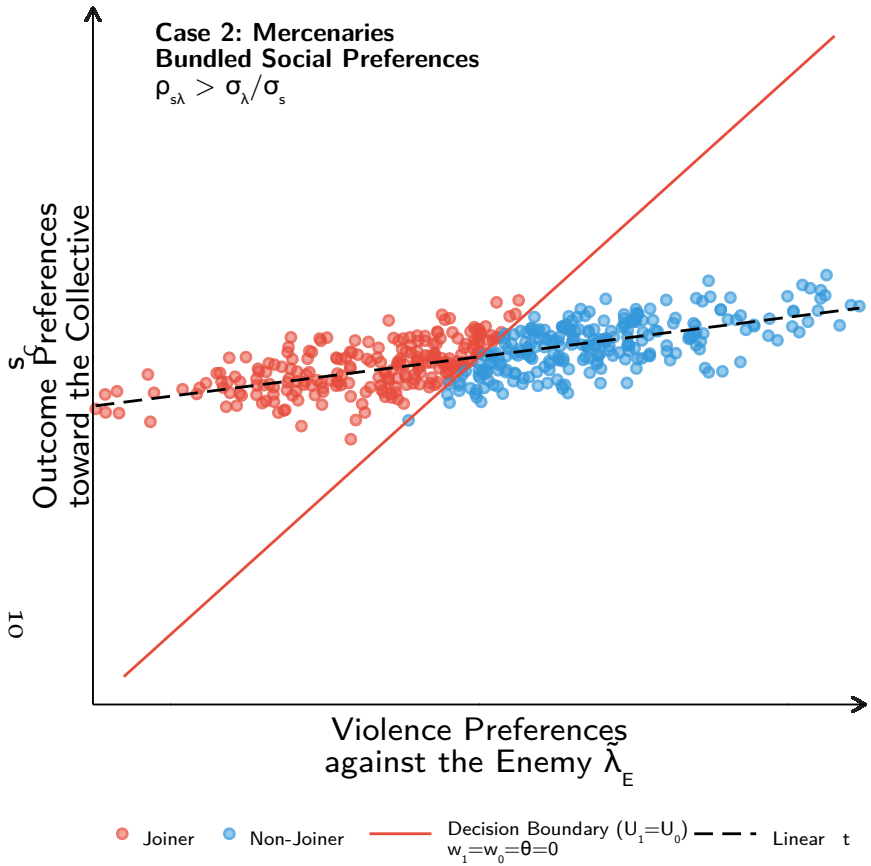
$$\begin{pmatrix} \tilde{s}_C \\ \tilde{\lambda}_E \end{pmatrix} \sim N\left(\boldsymbol{o}, \begin{pmatrix} \sigma_s^2 & \sigma_{s\lambda} \\ \sigma_{s\lambda} & \sigma_\lambda^2 \end{pmatrix} \right)$$

Figure 1: A Framework for Violent Social Action

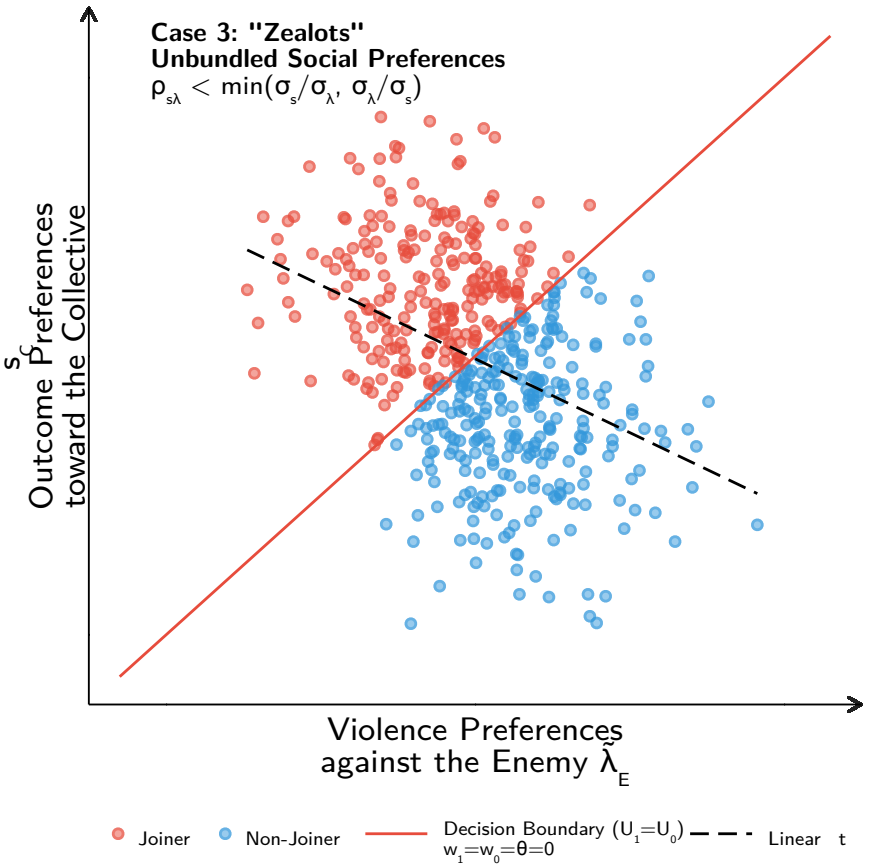


Notes: This figure presents the three cases implied by unbundling social preferences, derived from simulating the parsimonious framework presented in this section. When $\rho_{s\lambda}$ is *large and positive*, social preferences are *bundled*: those who are more prosocial toward the collective tend to be more violence-averse toward the enemy. In this case, the scope of selection depends on the relative variances of these components. There are two cases, with the first depicted on this page and the second in the next page. *Case 1: "Non-violent Social Movements."* When the variance in outcome prosociality toward the collective is large relative to violence aversion toward the enemy ($\rho_{s\lambda} > \frac{\sigma_s}{\sigma_\lambda}$), the most outcome prosocial toward the collective are *disproportionately* violence-averse toward the enemy, but this outcome prosociality compensates for that aversion. Individuals who are both highly outcome-prosocial toward the collective *and* highly violence-averse toward the enemy self-select into enlistment, corresponding to "non-violent" or "soft" social movements driven by prosocial motivations towards the collective's welfare and nonviolent toward the enemy.

Case 2. If Social Preferences are Bundled ("mercenaries")



Case 3. If Social Preferences are Unbundled ("zealots")



Notes: When $\rho_{s\lambda}$ is *large and positive*, social preferences are *bundled*: those who are more prosocial toward the collective tend to be more violence-averse toward the enemy. In this case, the scope of selection depends on the relative variances of these components. There are two cases, with the first depicted on the previous page and the second being *Case 2*: *Case 2: "Mercenaries."* When the variance in outcome prosociality toward the collective is small relative to violence aversion toward the enemy ($\rho_{s\lambda} > \frac{\sigma_\lambda}{\sigma_s}$), those who are more outcome-prosocial toward the collective are also more violence-averse toward the enemy, but their prosocial orientation is not strong enough to compensate. Individuals low on *both* dimensions self-select into enlistment—yielding an army of individuals distinguished by being the least concerned with the collective's well-being and the least averse to violence, a toward it.When $\rho_{s\lambda}$ is *low or negative*, social preferences are *unbundled*: outcome prosociality toward the collective and violence aversion toward the enemy are weakly correlated or even negatively correlated—i.e., being outcome prosocial toward the collective is no guarantee against low violence aversion toward the enemy, and may even be associated with greater inclination toward violence against the enemy. The intuition for this unbundling is captured by the insight that "the line dividing good and evil cuts through the heart of every human being" (Solzhenitsyn, 2007). In such scenario, there is only one case: *Case 3: "Zealots."* Individuals who are both highly outcome-prosocial toward the collective and minimally violence-averse toward the enemy self-select into enlistment. We refer to these as "zealots," reflecting their orientation toward the welfare of the collective alongside reduced violence aversion toward the enemy—without implying irrationality or moral judgment.

Assumption A implies a Roy-like taxonomy of *three* types of selection into violence, depending on the covariance structure of social preferences. These cases are illustrated in Figure 1.

When $\rho_{s\lambda}$ is *large and positive*, social preferences are *bundled*: those who are more prosocial toward the collective tend to be more violence-averse toward the enemy. In this case, the scope of selection depends on the relative variances of these components:

Case 1: "Non-violent Social Movements." When the variance in outcome prosociality toward the collective is large relative to violence aversion toward the enemy ($\rho_{s\lambda} > \frac{\sigma_s}{\sigma_\lambda}$), the most outcome prosocial toward the collective are *disproportionately* violence-averse toward the enemy, but this outcome prosociality compensates for that aversion.=. Individuals who are both highly outcome-prosocial toward the collective *and* highly violence-averse toward the enemy self-select into enlistment, corresponding to "non-violent" or "soft" social movements driven by prosocial motivations towards the collective's welfare and nonviolent toward the enemy.

Case 2: "Mercenary armies." When the variance in outcome prosociality toward the collective is small relative to violence aversion toward the enemy ($\rho_{s\lambda} > \frac{\sigma_\lambda}{\sigma_s}$), those who are more outcome-prosocial toward the collective are also more violence-averse toward the enemy, but their prosocial orientation is not strong enough to compensate. Individuals low on *both* dimensions self-select into enlistment—yielding an army of individuals distinguished by being the least concerned with the collective's well-being and the least averse to violence toward the enemy.

When $\rho_{s\lambda}$ is *low or negative*, social preferences are *unbundled*: outcome prosociality toward the collective and violence aversion toward the enemy are weakly correlated or even negatively correlated—i.e., being outcome prosocial toward the collective is no guarantee against low violence aversion toward the enemy, and may even be associated with greater inclination toward violence against the enemy. The intuition for this unbundling is captured by the insight that "the line dividing good and evil cuts through the heart of every human being" (Solzhenitsyn, 2007).

Case 3: "Zealot armies." Individuals who are both highly outcome-prosocial toward the collective and minimally violence-averse toward the enemy self-select into enlistment. We refer to these as "zealots," reflecting their orientation toward the welfare of the collective alongside reduced violence aversion toward the enemy—without implying irrationality or moral judgment.

In sum, selection into violence depends not only on the *levels* of social preferences but also on their *covariance structure*: the correlation between outcome prosociality toward the collective and violence aversion toward the enemy determines which types enlist. This insight generates the testable predictions that motivate the empirical comparisons we undertake in the next sections.

3. Empirical Strategy

A. *Empirical Setting: the Group*

Measuring selection into violent organizations is challenging due to secrecy. We overcome this by embedding a real-time notification system within a militia (henceforth, the group), allowing us to survey both joiners and non-joiners and study selection based on social preferences. The group is one of the most powerful in the region (Figure A1), with 2,789 combatants controlling 110 villages as of May 2023.[2]

Each month, 30 to 40 individuals enlist—providing a rare opportunity to measure the characteristics of selection into violence through surveys in real-time in a feasible time frame (Figure A2, Panel B). To enlist, individuals typically travel by foot to the nearest group post, often more than a day's journey, where they are referred to a Human Resource commander (S1) for enlistment. The group's HR structure includes 55 S1s, overseen by a regional commander (T1), as shown in Figure A3. ll applicants are accepted, allowing us to treat enlistment as revealed inclination.[3]

The group emerged in 2014 as a breakaway from its predecessor, which had collaborated with the Fronts De Libération du Rwanda (henceforth, *FDLR*), its own enemy, in attacks against its own civilians, violating its original defense mandate. The split partly restored the group's legitimacy among civilians and reinforced its self-proclaimed mission: defending the District—a multi-ethnic area larger than Switzerland—from the FDLR, responsible for two decades of displacement and massacres. Whether genuine or not, this framing provides the opportunity to examine the role of prosocial motivations toward the District's well-being in enlistment.

The group finances its operations through a tax system generating $1.5 million per month, alongside distribution monopolies (source: administrative records and authors' calculations). These revenues support combatant salaries, social insurance, and logistics. While enlisting exposes joiners to mortality risk (with an estimated eight-year survival rate of 55-70%, compared to over 95% for civilians), both joiners and civilians believe enlistment *reduces* lifetime income prospects, even accounting for expected promotion trajectories (Figure A4).

This combination of voluntary enlistment, a collective defense-oriented mission providing non-pecuniary prosocial rewards, and the perpetration of violence against an enemy creates an ideal setting to test the framework's predictions.

[2] As of 2023, 5% of adult males and 2% of the population were combatants (Figure A2, Panel A).

[3] Source: Vogel, Salvaggio, Boisselet, and Stearns (2021). This group is distinct from the one studied in Mugaruka, Marchais, Sánchez de la Sierra, and Wu (2024) and operates in a different province, unrelated to the M23.

B. *Data Collection Strategy*

To track joiners in real time, we established a notification system covering all 55 S1s and their 79 assistants, representing the complete set of "ports of entry" for enlisting.

The system relies on *real-time* phone call notifications to the study coordinator from the S1s. Each call is triggered when an individual reaches an S1 and declares their intention to enlist (henceforth, *joiners*). In each call, the coordinator records information about the joiner and details about their household and village of origin. Since the typical delay between first contact and enlistment is a few days, this system allows us to interview joiners immediately after they express their intent, before they enlist.[4] Figure 2, Panel A, shows that the notifications received through our system closely match the number of joiners recorded in the group's administrative data.[5]

We leverage this system to design a data collection process that allows us to compare the social preferences of joiners with a representative sample of non-joiners. Figure 2, Panel B, illustrates the data collection design for one joining event. For each notification, one team visited the joiner (henceforth, *the joiner team*), while another visited the joiner's village (henceforth, *the village team*). The joiner team met with the joiner to gather measures of their social preferences. Simultaneously, the village team visited two households: (a) the joiner's household and (b) a randomly selected household within the same village. In each of these households, the village team met with a randomly selected *youth* (aged 18–45) matching the joiner's gender (all were male), gathering measures of their social preferences—henceforth, referred to as the *household member* and the *villager*. Additionally, in both the joiner's household and the village household, the village team collected information about the household from the household head.[6]

From January 22 to August 25, 2022, we stationed ten survey teams, each consisting of two surveyors, across the District. During this period, we received 213 notifications through the system, responding to each by deploying both a joiner team and a village team. The final sample consists of 213 joiner households and 213 randomly selected households, covering a total of 426 households and 639 individuals. Figure A6 presents a map of the sample.[7]

[4]This system was developed between January and July 2019 and implemented from July 2019 to April 2020. The installation process included a survey of all S1 commanders conducted between July and August 2019.

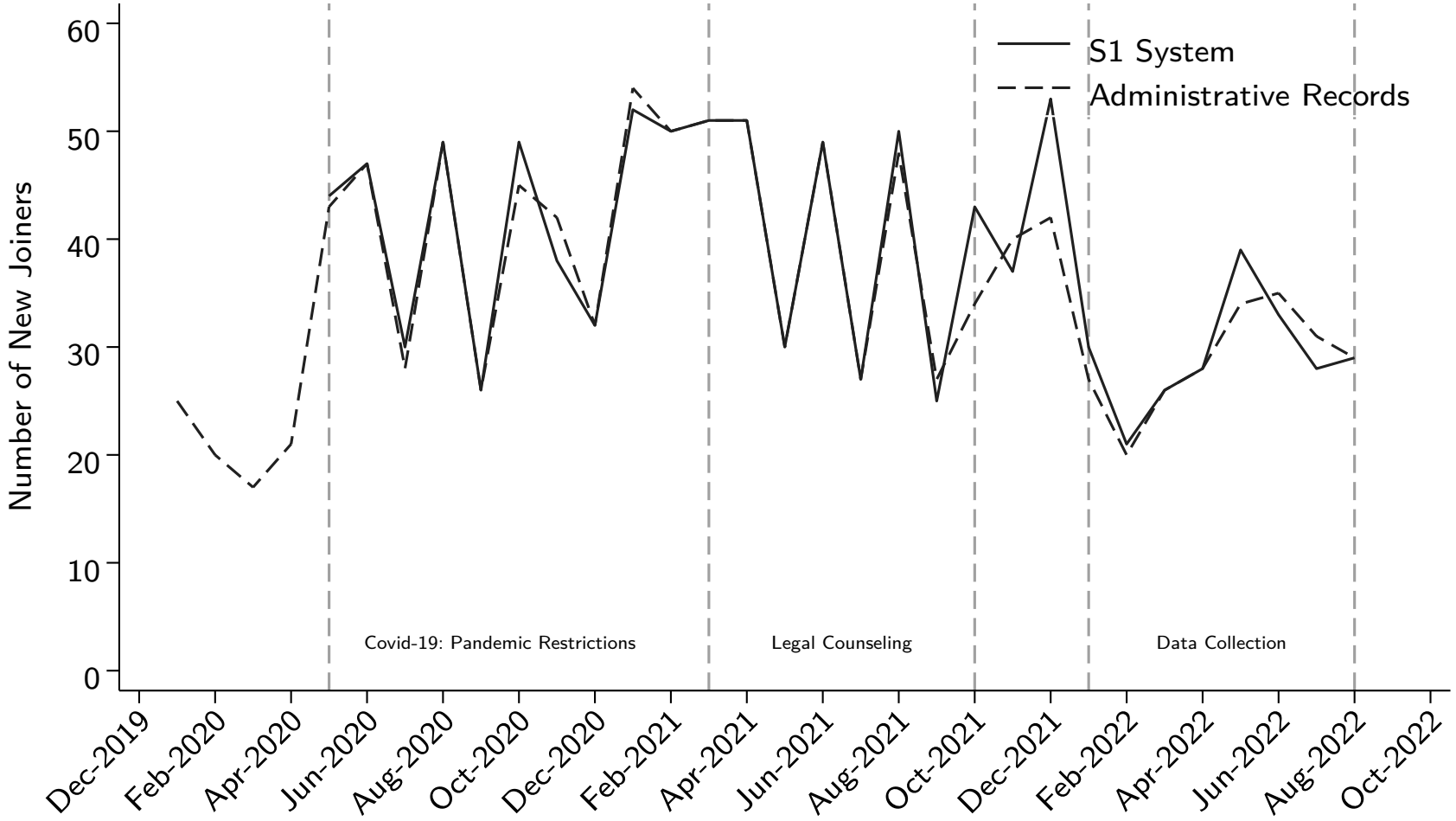
[5]These data are consolidated monthly and retrieved bi-monthly, making them unsuitable for interviewing joiners before enlistment. They are used solely as a benchmark for comparison with the enlistment data from our system.

[6]Village household lists were constructed from village censuses when available. In each selected household, enumerators created a roster of young males (aged 18–45) and randomly selected one. Interviews were conducted privately by surveyors from a respected local NGO trusted by the population.

[7]This design is the result of seven years of preparation and required addressing complex ethical and legal considerations. Section Appendix D provides detailed background on security, legality, and ethics.

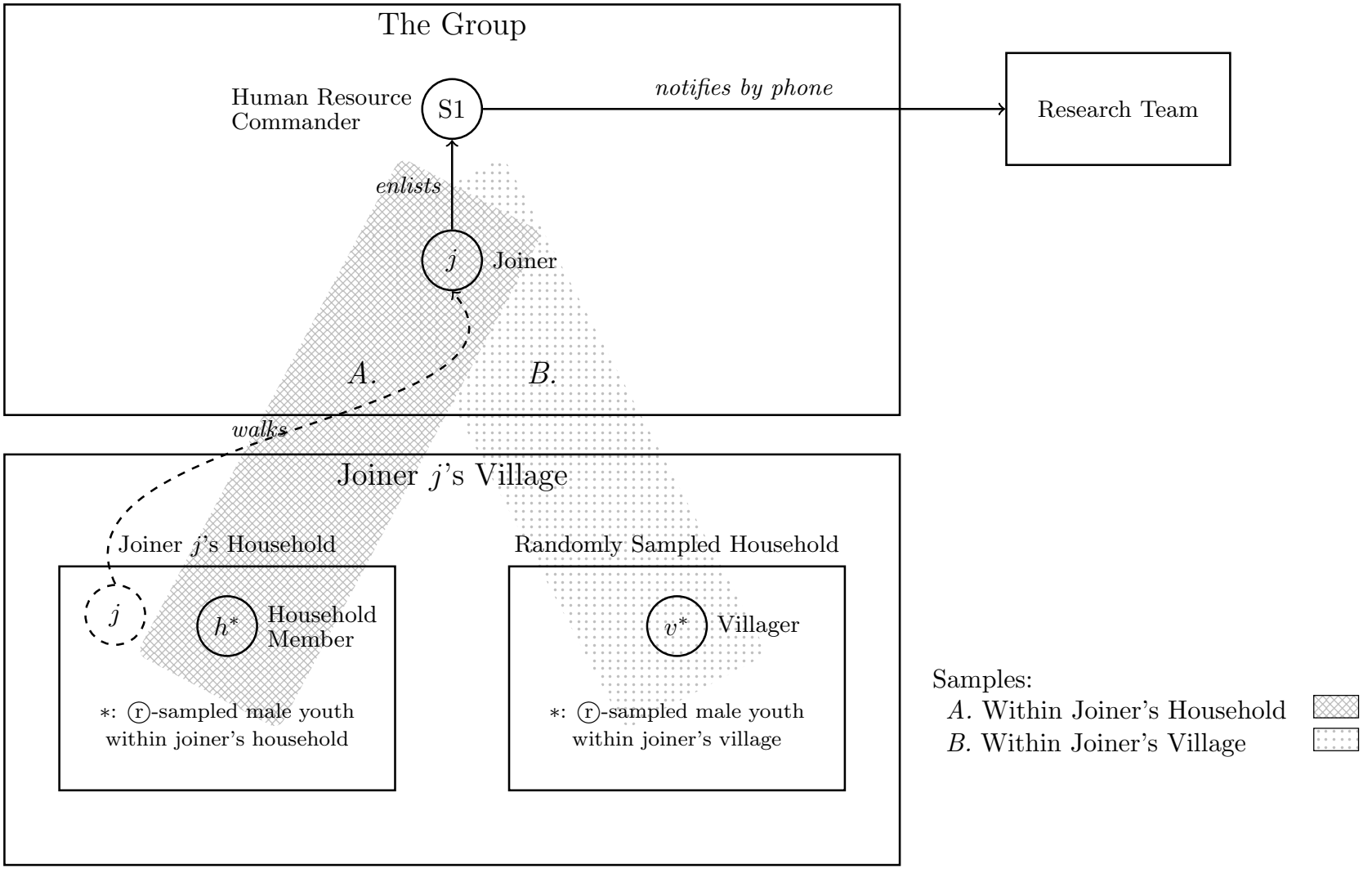
Figure 2: Data Collection Strategy

Panel A. The Real-Time Notification System About Enlistment Intent



Notes: Panel A presents the times series for the number of calls received from the notification system against the number of new joiners found in the group's administrative records. The notification system covered the group's all 55 human resource commanders (S1s) and their 79 assistants, representing the complete set of "ports of entry" for enlisting. The system relies on *real-time* phone call notifications to the study coordinator from the S1s. Each call is triggered when an individual reaches an S1 and declares their intention to enlist (henceforth, *joiners*). In each call, the coordinator records information about the joiner and details about their household and village of origin. Since the typical delay between first contact and enlistment is a few days, this system allows us to interview joiners immediately after they express their intent, before they enlist.

Panel B. Data Collection Design Triggered by One Enlistment Intent Event—Stylized Representation



Notes. This panel provides a graphical representation of the data collection design for one joining event. S1 indicates human resource commander, of which there are 55, all included in the embedded enlisting monitoring system. For each notification, one team visited the joiner, while another visited the joiner's village. The joiner team met with the joiner to gather measures of their social preferences. Simultaneously, the village team visited two households: (a) the joiner's household and (b) a randomly selected household within the same village. In each of these households, the village team met with a randomly selected *youth* (aged 18–45) matching the joiner's gender (all were male), gathering measures of their social preferences—henceforth, referred to as the *household member* and the *villager*. Additionally, in both the joiner's household and the village household, the village team collected information about the household from the household head. From January 22 to August 25, 2022, we stationed ten survey teams, each consisting of two surveyors, across the District.

C. *Measurement Strategy*

We leverage this system to collect measures of outcome preferences and violence preferences among both joiners and non-joiners. All measures are rescaled to a [0-10] scale and are graphically depicted in Figure 3. This section describes these measures; other measures, such as covariates, are introduced in subsequent sections when their description becomes relevant.

Outcome Preferences

Outcome *prosociality* is measured using two types of instruments: (1) established survey-based games and modules validated across diverse contexts (Falk, Becker, Dohmen, Enke, Huffman, and Sunde, 2018, Falk, Becker, Dohmen, Huffman, and Sunde, 2022)—specifically, a dictator game, a hypothetical donation task, and declared prosociality; and (2) a time-dictator game that we created and adapted to local norms around prosocial gestures to neutralize income effects that may arise in interpersonal comparisons of outcome preferences.[8]

Antisociality is measured using a *joy-of-destruction* game, in which respondents can reduce a target's endowment in a dictator game-like scenario (Abbink and Sadrieh, 2009).[9]

Violence Preferences

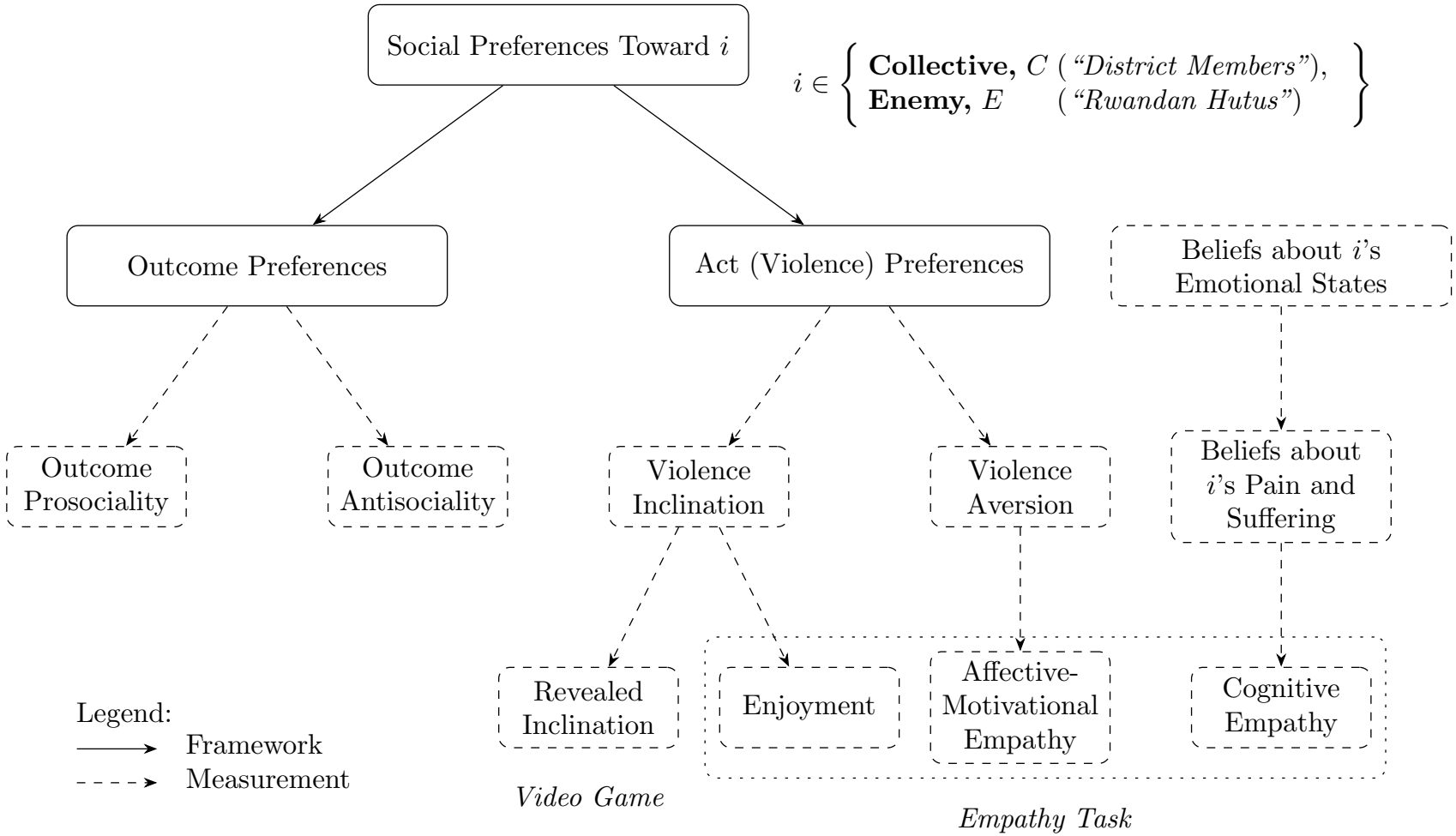
To measure *violence aversion*, we use a *suffering empathy task* to measure emotional responses to pain (Decety and Jackson, 2004b, Ferguson and Inzlicht, 2022). Respondents view an image depicting someone in pain (piloted to reflect local realities, e.g., an ear being cut or a fall from a tree) and rate: (a) "How sorry do you feel for this person?" (measuring *empathic concern*) and (b) "How are you feeling right now?" measuring *affective empathy*. Both scores are averaged to produce a composite measure of *affective-motivational empathy* (henceforth, *A.M. Empathy*).[10]

[8]In the dictator game, respondents allocate 10 tokens (each worth 10 US cents) to a target or destroy the target's endowment at a personal cost. In the donation task, respondents decide how much to donate from a $1,000 endowment to the District, Rwandan Hutus, or a local militia. Declared prosociality is measured using a self-reported scale based on the question: "How likely are you to take costly actions that benefit the community of [X]?" In the time-dictator game, respondents allocate up to 10 hypothetical hours to benefit a stranger, with varying imposed counterfactual use of time: 1. discretionary; 2. income-earning, 3. socializing/family.

[9]Following (Abbink and Sadrieh, 2009), destruction is costless and yields no payoff benefit. We implement the dictator and *joy-of-destruction* games sequentially and compactly: respondents first decide whether to transfer part of their 10-token endowment to the target or destroy part of the target's 5-token endowment, then choose the amount. This provides a complete set of dictator and destruction choices.

[10]These two measures are often referred to as empathic concern and personal distress (Batson, Fultz, and Schoenrade (1987)). Both yield qualitatively identical analysis in our data. The individual's face is not visible in the images. We are grateful to Professor Decety for sharing with us the task validated for use in the United States.

Figure 3: Unbundling Social Preferences



Notes: Outcome *prosociality* is measured using two types of instruments: (1) established survey-based games and modules validated across diverse contexts (Falk, Becker, Dohmen, Enke, Huffman, and Sunde, 2018, Falk, Becker, Dohmen, Huffman, and Sunde, 2022)—specifically, a dictator game, a hypothetical donation task, and declared prosociality; and (2) a time-dictator game that we created and adapted to local norms around prosocial gestures to neutralize income effects that may arise in interpersonal comparisons of outcome preferences. In the dictator game, respondents allocate 10 tokens (each worth 10 US cents) to a target or destroy the target's endowment at a personal cost. In the donation task, respondents decide how much to donate from a $1,000 endowment to the District, Rwandan Hutus, or a local militia. Declared prosociality is measured using a self-reported scale based on the question: "How likely are you to take costly actions that benefit the community of [X]?" In the time-dictator game, respondents allocate up to 10 hypothetical hours to benefit a stranger, with varying imposed counterfactual use of time: 1. discretionary; 2. income-earning, 3. socializing/family.

To measure *violence inclination*, we develop two tasks. First, adapting the *suffering empathy task*, respondents rate how much they *enjoyed* viewing the suffering depicted in the image. Second, we use a custom-designed video game where respondents encounter a person on a bridge and choose how to interact: non-violently (e.g., giving candy, doing nothing) or with lethal violence (e.g., shooting, stabbing, or pushing the person off the bridge, all resulting in visible death).

Beliefs About Others' Pain and Suffering [Cognitive Empathy]

We measure beliefs about others' suffering as a component of *cognitive empathy*, which refers to an individual's "recognition and understanding of another's emotional state" (Decety and Jackson, 2004b). Such beliefs may constrain affective-motivational empathy by limiting one's capacity to recognize others' suffering. To measure cognitive empathy, respondents in the suffering empathy task rate how much they believe the person in the image is suffering, on a scale from 0 to 10, immediately after observing pain being inflicted. This rating captures their assessment of the intensity of the other's pain, operationalizing cognitive empathy as the capacity to evaluate another person's suffering (Zaki and Ochsner, 2012a, Lamm, Batson, and Decety, 2007). This approach focuses on pain assessment, a task-specific application of cognitive empathy that is conceptually distinct from the affective-motivational dimensions described above.

Cognitive empathy may be *motivated*—distorted to serve self-interest or reduce discomfort (Bénabou and Tirole, 2016)—which we henceforth refer to as *self-deception*. Self-deception aligns closely with the "distorting consequences" mechanism and, to some extent, dehumanization—both elements of *moral disengagement* (Bandura, 2011), through which individuals adjust their beliefs to reduce discomfort with committing violence. Regardless of whether civilians systematically overestimate or underestimate others' pain, the rating in the cognitive empathy measure allows us to examine whether joiners exhibit *greater suppression* of cognitive empathy.[11]

Each of the measures is administered twice—once for a member of the collective and once for an enemy member. In the survey, we operationalize the collective as "members of the District," with defense framed as a collective good encompassing all individuals within the District. The enemy is represented by the "Rwandan Hutu," in accordance with local terminology.[12]

[11]Survey protocols ensured privacy and anonymity through encrypted digital interfaces. Enumerator training and piloting were conducted in three phases between 2019 and 2022, with approvals from local authorities.

[12]The choice of "Rwandan Hutu" reflects local social categorization. It also acknowledges the blurring of boundaries between enemy civilians and combatants, as discussed in later sections. Additional variations later distinguish preferences toward Rwandan Hutus as armed adversaries from those shaped by their social identity.

4. The Role of Social Preferences in Selection into Enlistment

The *key relationship* provides a testable implication for the role of social preferences in enlistment decisions. In this section, after documenting the patterns of unbundled social preferences in the general population of non-joiners, we estimate the *key relationship* and then test this implication.

A. *Patterns of Ununbundled Social Preferences and Beliefs in the General Population*

Before turning to the *key relationship*, we first present three patterns in unbundled social preferences in the general population. To ensure representativeness for the population of interest, this section relies on the sample of randomly selected *villagers* (Figure 4). Together, these patterns reveal an extreme form of violent parochial altruism directed against the enemy, driven by an inclination toward violence perpetration grounded in *desensitization* and *self-deception*.

First, social preferences are *parochial* across all dimensions: villagers exhibit stronger outcome prosociality and violence aversion toward District members than toward the enemy (Panels A and B), and even exhibit violence inclination toward enemy members (Panel B, Figure A9).[13] Violence inclination is quite telling: 77% enjoy seeing the enemy in pain, and 59% choose to kill them privately in the video game—compared to virtually none toward District members.

Second, villagers' violence inclination toward the enemy hinges mostly on *selective desensitization* and, to a lesser extent, on *self-deception*: despite their low violence aversion (as measured in A.M. empathy) toward the enemy, most maintain high average *cognitive empathy* toward them (driven by a cluster, the majority, with full *cognitive empathy* toward the enemy), and a large mass has suppressed cognitive empathy, amounting to *self-deception*.

Third, outcome prosociality and violence aversion are positively correlated, but their variances and patterns differ markedly, implying that each also captures unique, non-overlapping dimensions of social preferences.[14] This suggests that their unbundling could be informative.

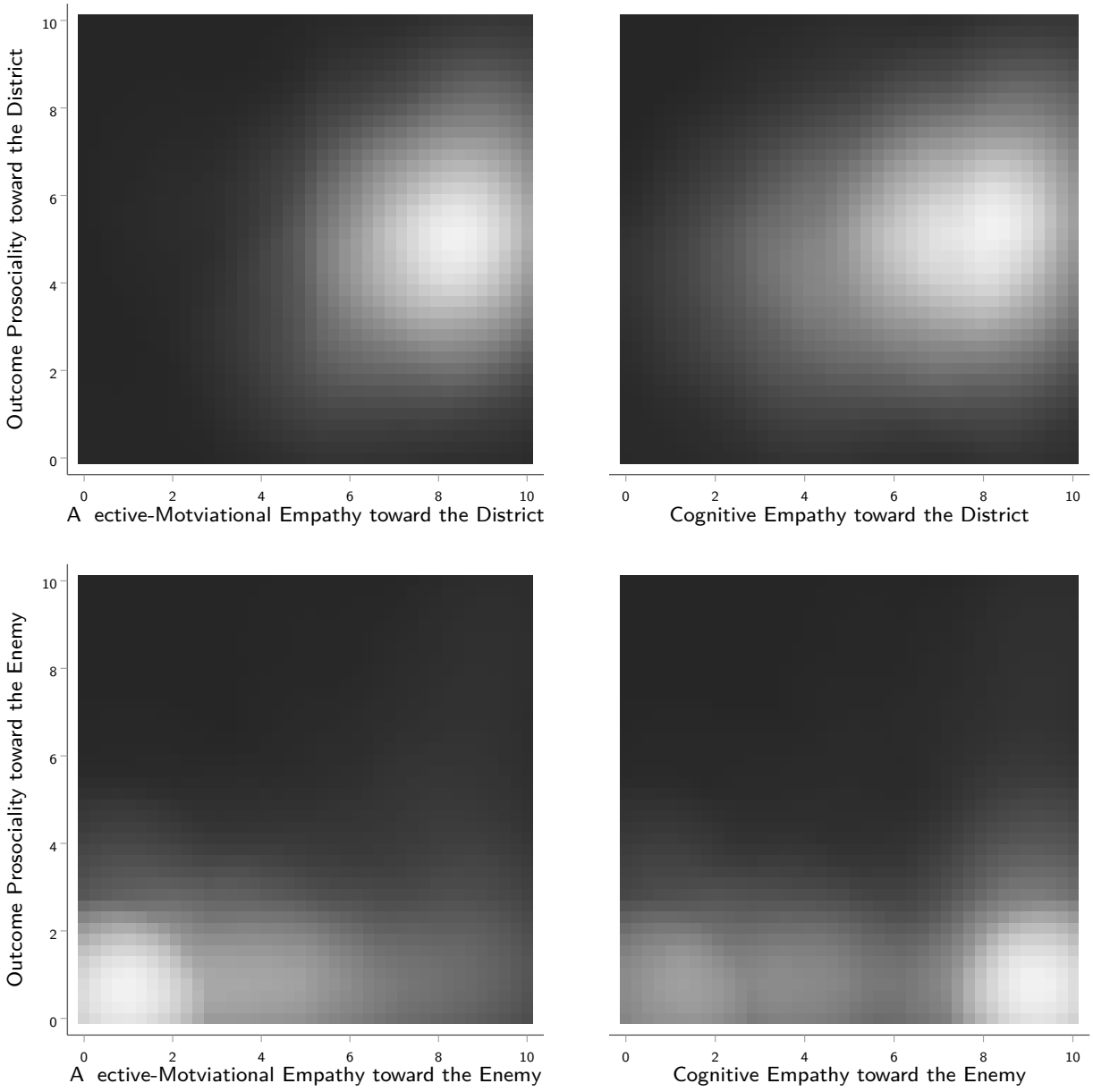
These patterns underscore the need to jointly consider outcome and violence preferences—alongside beliefs about others' suffering—when examining social preferences and enlistment. We now turn to the *key relationship* between outcome prosocial preferences toward the District and violence aversion toward enemy members.

[13] Figure A11 replicates Panel B, replacing the enemy with a member of the District's armed group. The similarity in preferences toward District civilians and armed members suggests parochialism is driven by social identity, not mere association with weapons. It also suggests there is very little animosity toward them.

[14] They are positively correlated for the same target (Figure A10; Table A1), with target identity accounting for 60% of the correlation (Table A2).

Figure 4: Patterns of Unundled Social Preferences and Beliefs in the General Population

Panel A. Joint Distributions—Outcome Preferences vs. Violence Aversion and Beliefs

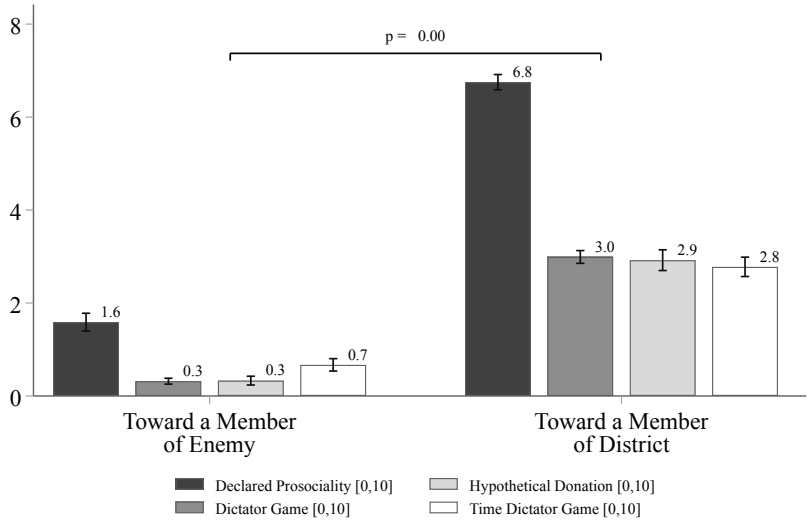


Notes. This figure presents the joint distribution (in heatmap form) of outcome prosociality toward a person of identity X (y-axis) and, on the x-axis: (i) violence aversion toward a person of identity X (left panels), and (ii) beliefs about others' pain and suffering, measured as cognitive empathy (right panels). In the upper panels, X represents a member of the District; in the lower panels, X represents a member of the enemy. Outcome prosociality is the average of: declared altruism [0-10], hypothetical donation [0-1000] (rescaled to [0-10]), and dictator game donation [0-10], all directed toward a person of identity X. Violence aversion is measured using a suffering *suffering empathy task* adapted to local contexts (Decety and Jackson, 2004b, Ferguson and Inzlicht, 2022). Respondents view images depicting individuals in pain (e.g., an ear being cut or a fall from a tree) and rate: (a) "How sorry do you feel for this person?" (empathic concern), and (b) "How are you feeling right now?" (affective empathy). The average of these scores yields a composite measure of affective-motivational empathy (A.M. Empathy). The individual's identity is obscured in the images to prevent identification. We pool these empathy measures for clarity, as they operate similarly across all analyses. Lighter colors in the heatmap indicate a higher concentration of observations; darker colors indicate a lower concentration. Sample: randomly selected young males in randomly selected households of each of the joiners' villages of origin (referred to as the *villagers*, totalling 213 individuals in 133 villages).

Panel B. Mean Comparisons—Toward the District vs. Toward the Enemy

Panel B1. Outcome Preferences

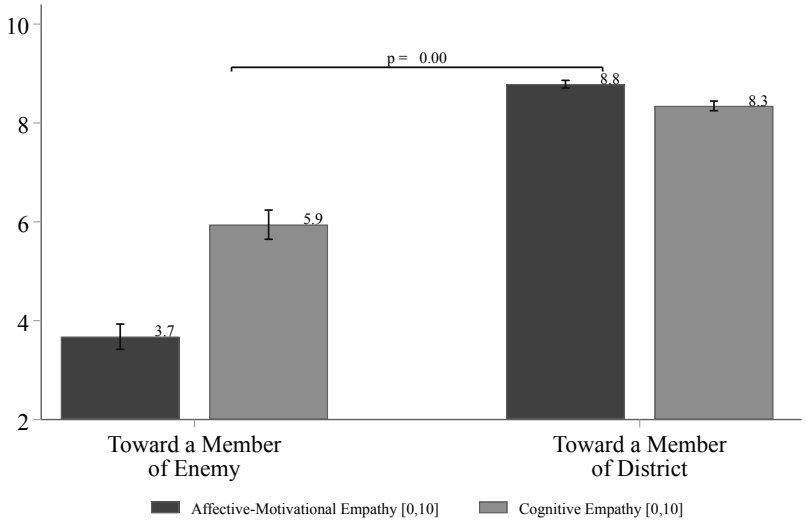
Prosocial



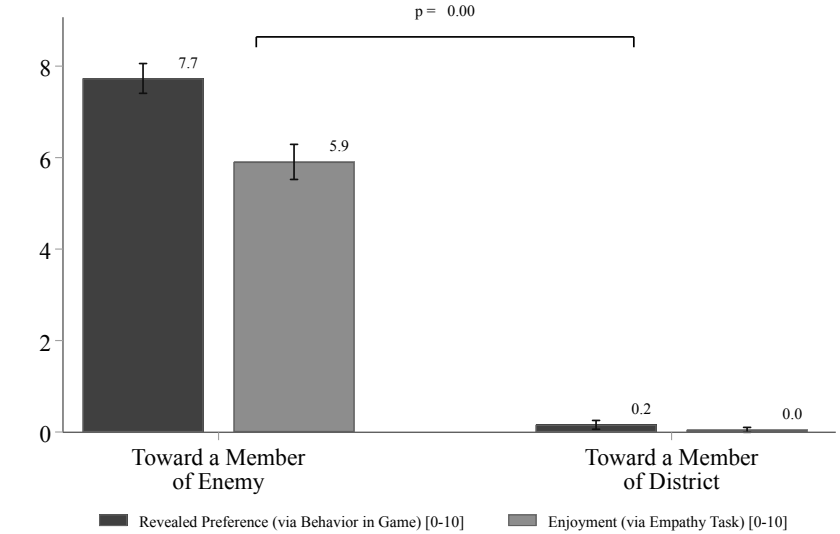
Anti-social

Panel B2. Violence Preferences

Aversion



Inclination



Notes. This figure presents the raw means of outcome preferences (Panel B1) and violence preferences (Panel B2). Outcome preferences are further decomposed into outcome prosociality and outcome antisociality, while violence preferences are decomposed into violence aversion and violence inclination. Each measure is shown separately for two targets: the District and the enemy. All measures range from 0 to 10. The p-value in each panel corresponds to a test of equality of means between the average of the measures in each panel toward the enemy vs. the average of the same measures toward the District. Sample: randomly selected young males in randomly selected households of each of the joiners' villages of origin (referred to as the *villagers*, totalling 213 individuals in 133 villages).

B. *The Key Relationship*

The *key relationship* hinges on the correlation between outcome prosociality toward the District and violence aversion toward the enemy, $\rho_{s\lambda}$, which reveals whether social preferences are bundled and, together with their relative variances, determines the scope of selection into enlistment.

We begin by assessing this correlation. Figure A12 shows that individuals who are more outcome prosocial toward the District exhibit weakly lower violence aversion toward the enemy (as measured by A.M. empathy), highlighting that social preferences are *unbundled*. Furthermore, Figure 4 shows that these same individuals exhibit *higher* violence aversion toward District members, implying that their reduced aversion to violence against the enemy reflects not a general lack of violence aversion, but a selective suppression of A.M. empathy toward the enemy.

What drives the suppression of violence aversion toward the enemy, particularly among those most outcome-prosocial toward the District? Analysis of *cognitive empathy* reveals a key mechanism: *self-deception*. Individuals most prosocial toward the District appear to suppress their *cognitive empathy* toward the enemy, thereby reducing their violence aversion. Two patterns support this interpretation. First, this correlation is more negative when *cognitive empathy* is used as the outcome measure, rather than violence aversion. Second, this negative correlation disappears when *cognitive empathy* toward the enemy is included as a control (Table A3).

We next examine the *key relationship* across sub-samples within households and villages. First, the *key relationship* satisfies the condition for Case 3—$\hat{\rho} < \min\left(\frac{\hat{\sigma_s}}{\hat{\sigma_\lambda}}, \frac{\hat{\sigma_\lambda}}{\hat{\sigma_s}}\right)$—both when comparing joiners to *household members* and *villagers*. Second, the evidence for Case 3 is stronger within households than across villagers: $\hat{\rho} - \min\left(\frac{\hat{\sigma_s}}{\hat{\sigma_\lambda}}, \frac{\hat{\sigma_\lambda}}{\hat{\sigma_s}}\right)$ is larger *within joiners' households* and than *within joiners' villages*.[15] Third, the evidence for Case 3 is also stronger when *cognitive empathy* toward the enemy is used as the outcome, instead of violence aversion alone.

Taken together, these findings highlight that those most committed to the District's welfare simultaneously suppress their empathy toward the enemy more strongly through *self-deception*, compared to their counterfactual higher capacity for A.M. empathy. They also imply that, if social preferences influence enlistment decisions: a. Joiners should exhibit greater outcome prosociality toward the District and weaker violence aversion toward the enemy; b. This selection pattern should be stronger within households than within villages; c. This selection pattern is driven by the selective suppression of *cognitive empathy*. We assess these implications in the next section.

[15]Figures A13, A14 confirm robustness to sampling weights and alternative measures (violence inclination).

C. *Testing the Implications: Selection into Enlisting*

We now analyze the selection pattern implied by the *key relationship*: are joiners more *outcome prosocial* toward the District and less *violence averse* toward the enemy (Case 3)?

Figure A12, marking joiners in red and non-joiners in blue, suggests this selection pattern is present in the data, especially *within households* and for *cognitive empathy*. To formally test for the predicted selection pattern, we estimate the equation:

$$y_i = \alpha + \beta J_i + \epsilon_i, \tag{1}$$

where i indexes individuals, y_i is a component of social preferences on a 0 to 10 scale, and J_i is an indicator equal to one if individual i is a joiner, and zero otherwise. We estimate this equation *within joiners' households* and *within joiners' villages*—that is, restricting the sample to joiners and, separately, other randomly selected male adults from: a. their households; b. their villages. The coefficient β captures the difference in y_i between joiners and the average young adult male in their household or village, respectively. Estimates are in Figure 5 and Table A4.

a. Results: Joiners are the "Zealots"

The data reflects the three selection patterns implied by the *key relationship*. First, within both *joiners' households* and *joiners' villages*, joiners exhibit greater outcome prosociality toward the District and simultaneously lower violence aversion toward enemy members compared to non-joiners (Panel A, Figure 5). Joiners' greater outcome prosociality toward District members is observed across measures that are both sensitive (hypothetical donations, the dictator game, and the time dictator game when the opportunity cost of time is endogenous) and insensitive to income effects (the time dictator game with zero opportunity cost of time and declared prosociality).

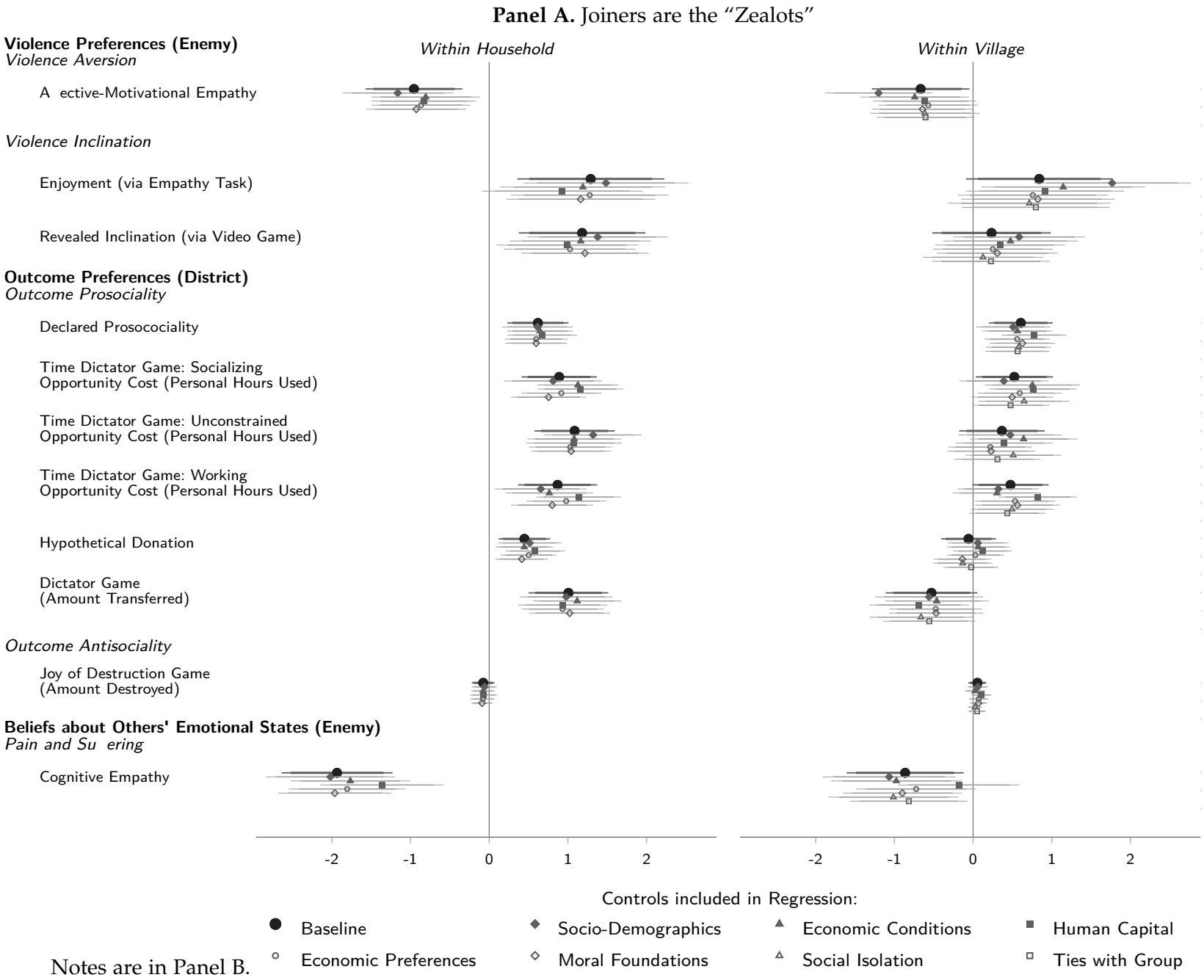
Second, this pattern of selection is more pronounced *within joiners households* than *within joiners villages*, revealing that joiners are more outcome prosocial toward the District and less violence averse toward the enemy especially compared to their young male household members.

Third, the pattern is also more pronounced for *cognitive empathy* than for violence aversion (A.M. empathy) toward the enemy. Furthermore, the joiners' lower violence aversion toward the enemy is *fully* explained by their selectively lower *cognitive empathy* (Table A5).[16]

In sum, the selection patterns are consistent with social preferences influencing enlisting decisions and underscore the role of *self-deception* by the most outcome-prosocial in selection.

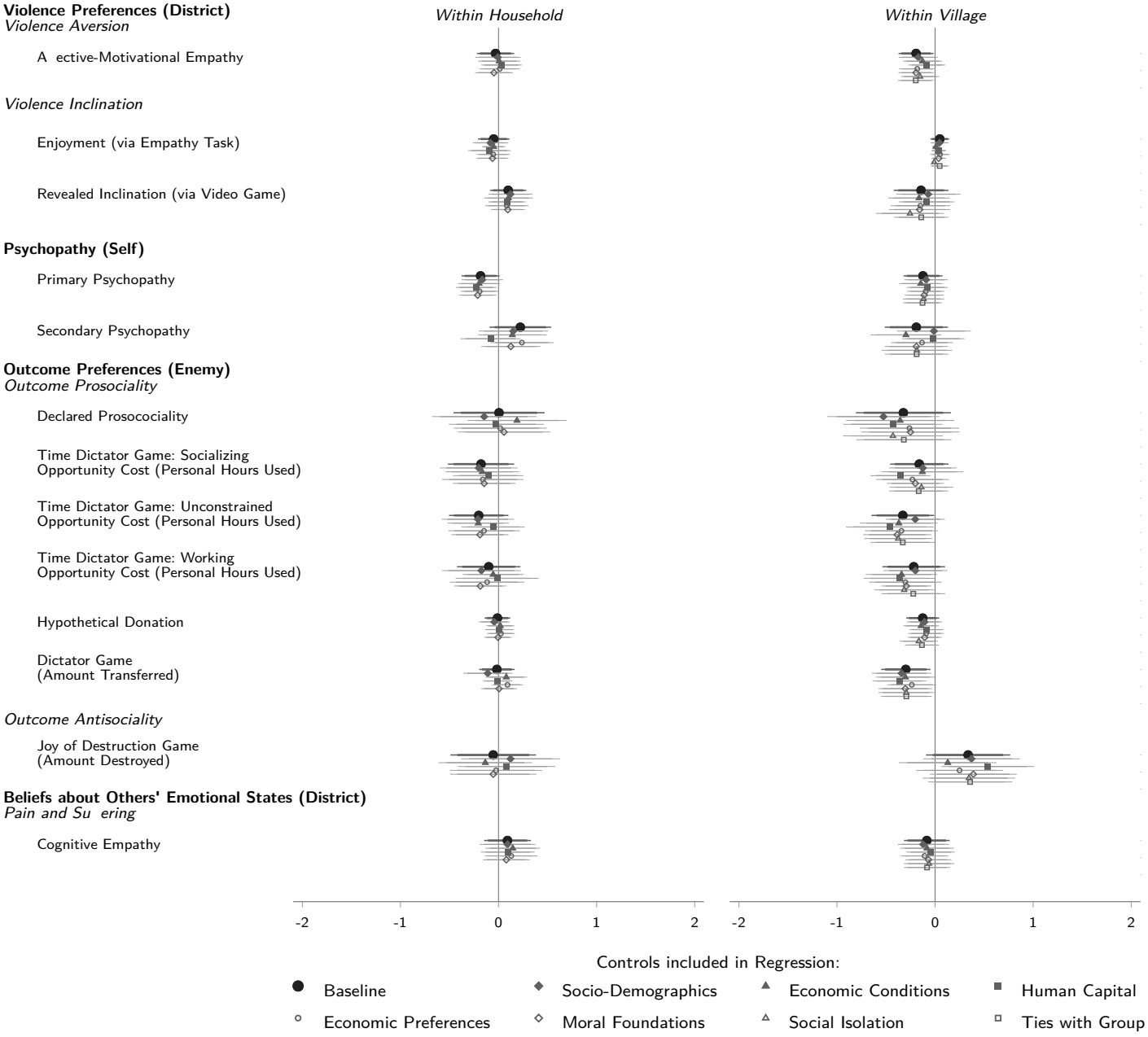
[16]The estimates are robust to alternative specifications, samples and functional forms (Figure A19).

Figure 5: Testing the Implication: Selection into Enlisting



Notes are in Panel B.

Panel B. Isolating the Social Preferences That Drive Enlistment: The Other (Unbundled) Dimensions



Notes. This figure presents the estimates of Equation 1. The coefficients and confidence intervals are presented without additional controls (Baseline) and including the controls indicated in the legend. Those include: Socio-Demographics (age, gender, literacy, ethnicity, religion, language, family); Economic Conditions (occupation in the last five years, employment status past month, labor supply (hours) past month, consumption of the head of household, beliefs about future income inside and outside the group); Human Capital (education, cognitive ability, physical health, mental health, smoking, alcohol consumption, noncognitive skills measured in big 6 HEXACO personality traits); Economic Preferences (risk preferences from risk lotteries and from hypothetical life-threatening choices, time preferences, social preferences in trust game, amount reported in cheating game, justice sensitivity inventory (JSI) from Schmitt, Baumert, Gollwitzer, and Maes (2010)), Moral Foundations (measured in Oliver Scott Curry and Whitehouse (2019)'s moral foundations and Geert Hofstede and Minkov (2010)'s horizontal and vertical collectivism), Social isolation (measured as the standardized [0-10] average of the amount other villagers expect to receive from the individual in question in a trust game and of the amount other villagers transfer in the first stage of a first game to the individual in question), and pre-existing social ties with group members. The specific way in which each of the controls is coded is described in Tables A9-A6 for demographics, economic conditions, human capital, preferences, Figure A15 for social isolation and Figure A17 for social ties with the group.

b. The Distinctive Role of Social Preferences in the Broader Structure of Selection

Unlike conventional analyses of treatment effects, selection into enlistment reflects the joint distribution of preferences and other traits—some of which may co-evolve alongside social preferences, even when social preferences influence enlistment decisions.[17] In this section, we aim to characterize the structure of selection by examining covariates with two objectives.

First, we analyze the distinctive profiles of joiners compared to non-joiners, incorporating all available data on both groups. Joiners are younger, more likely to belong to the dominant ethnic group, more often Protestant, have fewer siblings, and are later in the birth order (Table A6).

Economically, joiners are more vulnerable: they are less likely to be salaried professionals, less likely to have earned income in the past month, and more likely to be unemployed. Yet, they are more optimistic about future income within the group, despite enlistment being perceived as reducing lifetime income for both joiners and non-joiners (Table A7).

In terms of human capital and psychological traits, joiners have *higher* cognitive skills (IQ), are more likely to be secondary school dropouts, have worse mental health, and report higher substance use (Table A8). Behaviorally, joiners are more risk-averse, less patient, and exhibit slightly higher propensity to cheat and sensitivity to injustice (Table A9). Moral foundations and collectivism measures do not predict enlistment (Table A10).[18]

Socially, joiners are well-integrated and trusted within their villages (Figure A15). Furthermore, joiners are no more likely than non-joiners to have a sibling in the group (Figure A17).

Thus, joiners tend to be younger and smart but economically vulnerable and in worse mental health. Their enlistment is not related to social exclusion or unique moral values—both of which have been commonly found for joining gangs, crime or riots in other settings (Scacco, 2016).

Second, we assess the relationship between social preferences and other individual traits associated with enlistment by including these covariates as controls in Equation 1. The inclusion of these controls leaves the estimate of β stable (Figure 5).[19] This indicates that, compared to each of these clusters separately, social preferences are *independently* associated with selection.

[17] For instance, individuals from larger households may develop stronger social preferences that motivate enlistment.

[18] Figure A16 shows the correlation between Oliver Scott Curry and Little (2021)'s measures of moral foundations, which represent the most recent articulation of moral dimensions derived from evolutionary game theory, and Geert Hofstede and Minkov (2010)'s measures of horizontal and vertical collectivism. These dimensions capture traits linked to Graham, Haidt, and Nosek (2009)'s communitarianism/universalism.

[19] Figure A18 presents the variables selected in LASSO as predictors of social preferences: patience, emotionality (measured in HEXACO), being a street vendor, IQ, unemployment, and being a Jehovah's Witness.

c. Isolating the Social Preferences That Drive Enlistment

We now characterize the social preference components driving this selection pattern.

First, we assess whether joiners have a lower *capacity for empathy*. Panel B of Figure 5 shows that joiners exhibit similar violence aversion toward District members as non-joiners. They are also no more likely—if anything, less likely—to exhibit psychopathy, as measured by the Levenson, Kiehl, and Fitzpatrick (1995) scale.[20] Furthermore, the selection pattern remains unchanged when controlling for violence aversion toward District members (Figure A20).[21] This suggests that joiners have the same capacity for affective-motivational empathy as non-joiners but instead selectively suppress it toward enemies through *cognitive empathy* suppression.

Second, we assess whether joiners are more parochially altruistic. While joiners display stronger outcome prosocial preferences toward the District, their outcome preferences toward the enemy are similar to those of non-joiners (Panel B of Figure 5). Furthermore, when controlling for violence aversion toward the District, their outcome prosociality toward the District remains larger (Figure A20). Thus, the selection of joiners reflects a higher average level of outcome prosociality, while suppressing their affective-motivational empathy toward the enemy.

Third, Figure A21, Panel A, shows that joiners do not exhibit a higher inclination for violence against members of the District's armed group compared to non-joiners. This suggests that joiners' lower violence aversion toward enemy members is not merely a response to their association with weapons, but is specific to their enemy identity.[22]

In sum, those who enlist this violent organization exhibit greater outcome prosociality toward their collective and lower violence aversion toward the enemy—corresponding to the "zealots" of society. Together with the *key relationship*, this suggests that social preferences influence enlistment decisions. Furthermore, these patterns of selection are not merely explained by economic or social selection forces, a general deficit in empathy, generic parochial altruism, or incidental associations with the enemy being armed. In the next section, we take a step back to examine the societal moral roots of these two social preference components that are central in selection: outcome prosociality towards the District, and violence aversion towards the enemy.

[20] *"In general, psychopathy is thought to consist of two broad, somewhat distinct sets of traits: one involving emotional coldness, manipulation, and lack of empathy [primary psychopathy], and the other involving impulsivity, hotheadedness, and an antisocial lifestyle [secondary psychopathy]."* (Levenson, Kiehl, and Fitzpatrick, 1995).

[21] The use of terms like "suppressing" assumes that humans are naturally disinclined toward interpersonal violence, a view supported by research in moral psychology and neuroscience (Workman, Yoder, and Decety, 2020).

[22] Furthermore, joiners exhibit stronger outcome prosociality toward members of the District's armed group than non-joiners, reinforcing the conclusion that they are more outcome prosocial on average.

5. Interpretation: A *Morality of War* Shapes the Expression of Social Preferences

In this section, we document the moral foundations shaping the expression of the two central components of social preferences that influence the decision to enlist: the moralization of both the act of enlisting and the act of violence toward the enemy, anticipated at enlistment and bundled within it. We find evidence of a *morality of war* widespread among civilians and joiners alike, which moralizes enlistment as a consequentially *moral service* and the anticipated act of violence as a *moral duty*—henceforth, the *prevailing morality of war*.

A. *The Act of Enlisting as a Moral Service to the District*

The act of enlisting is best understood through a prevailing morality that defines enlistment as a service to the District, grounded in moral consequentialism, defining it as *outcome* prosocial toward the District. To establish this, we analyze civilians' moral convictions *about the act of enlisting*, measured through the choices the *villagers* in a moral dilemma experiment.

Enlistment Moral Dilemma Experiment. Respondents were asked to choose whether they *should* enlist if given the opportunity. In the simplest version, enlisting carries no implied trade-off.[23] The choices made by the villagers reveal that 65% of villagers consider enlisting morally right, even though they themselves have not enlisted (Figure 6, Panel A, upper panel).

Extension to measure consequentialism. Next, we assess the component of this moral conviction that arises exclusively through the intended *consequences* of enlisting (henceforth, the *consequentialist component*), by introducing a trade-off: enlisting prevents n others from doing so (see Figure A22). Respondents are presented with an increasing *trade-off*: 'If you enlist, n other young men cannot. Should you enlist in that case?'. We extend the outcome prosocial payoff presented in Section 2, s_C, by further unbundling it into the following two components: a *consequentialist* component, s_C^c, driven by the consequence of enlisting), and a *deontological* component, s_C^d, driven by the act of enlisting. Appendix A. demonstrates that, with this unbundling, variation in the *trade-off* introduced by n allows to pin down the distribution of the consequentialist component, s_C relative to the bundle of motivations that are not due to self-interest, henceforth denoted d (which includes economic incentives, violence aversion, as well as the deontological value of joining s_C^d). We identify three primary profiles based on their choices.[24]

[23] Responses were recorded privately to ensure anonymity, preventing enumerators from knowing the answers

[24] Strategies played are reported in Figure A24, Panel A.

The strategies played imply the following partition of the 64% who consider enlisting morally right. First, 15.5% are "*pure prosocial consequentialists*" ($s_C^c > -d \geq 0$). These individuals deem enlisting morally right solely due to its consequences but experience disutility from enlisting. Second, 23% are "*mixed prosocial consequentialists*" ($s_C^c > d > 0$). They prioritize the consequences of enlisting over self-interest but still derive positive utility from enlisting—other than through its consequences. Third, 8.4% are "*weak prosocial consequentialists*" ($50 s_C^c > d > s_C^c > 0$). While they consider collective consequences as morally righteous, their decision-making suggests that they assign greater weight to the payoff to the self from enlisting. The remaining 2% adopt nonmonotonic strategies, which likely indicates a misunderstanding of the game.[25]

Taken together, these results show that, of the 64% of civilians who view enlisting as morally righteous, 47% view it as moral through the contribution of enlisting to a perceived consequence—thus the moralization of the act of enlisting aligns with *moral consequentialism*.

But what exactly is that intended *consequence*? To explore the perceived outcome affected by enlisting, we analyze the joiners' *beliefs* about their and others' motivations. Joiners were asked: "Why did you enlist?" (henceforth, *first-order beliefs*, regarding their own motivations), "Why did other joiners enlist?" (*henceforth, second-order beliefs*), and "What do other joiners believe is your motivation for enlisting?" (henceforth, *third-order beliefs*).[26] These questions were also administered 8 months after enlistment (henceforth, *endline*; see Section 7 for more details).

Across these beliefs, "contributing to the defense of territorial integrity" (of the District) predominates (Figure 6, Panel A, lower panel). All other motivations fall below 20%, with income, personal safety, and revenge as the most prominent. This clarifies that the primary *consequence* driving the moralization of enlistment is the defense of the District.

The act of enlistment is thus *moralized* through its impact on the defense of the District and, by extension, its welfare, framed as a *moral service* to the District. This reinforces the potential rewards for individuals with high outcome-prosociality toward the District. Furthermore, this finding supports the interpretation that the selection of outcome-prosocial individuals toward the District is not incidental, but tied to the outcome-prosocial reward inherent in joining, through its perceived contribution to the District's defense and welfare.

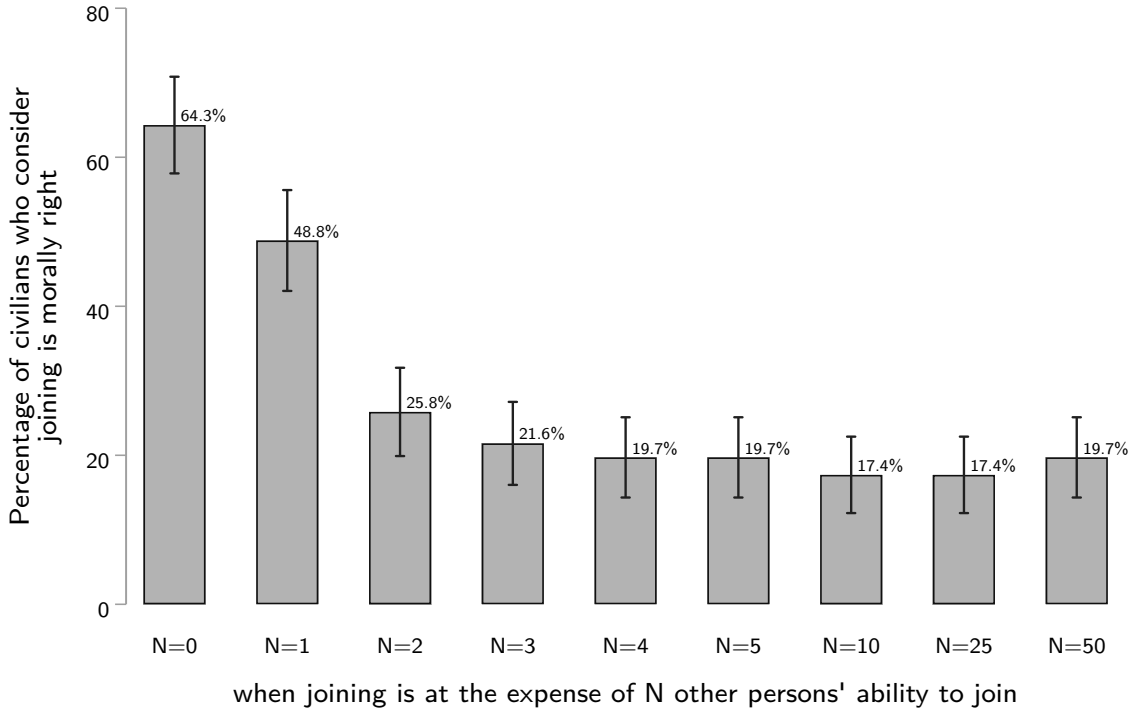
[25] A single crossing/monotonicity underlies the parsimonious interpretation above. If utility is additive in n, respondents' choices should exhibit monotonicity. Namely, those who initially choose enlisting when $n = 0$, should opt out of enlisting when n becomes large enough that collective costs outweigh individual benefits.

[26] To avoid priming the respondents, these questions were open-ended, with no prompts. The enumerator classified the responses into pre-existing categories or labeled them as "other," followed by an open-ended question. For second-order beliefs, respondents allocated 100 hypothetical joiners into bins representing different motives for enlistment.

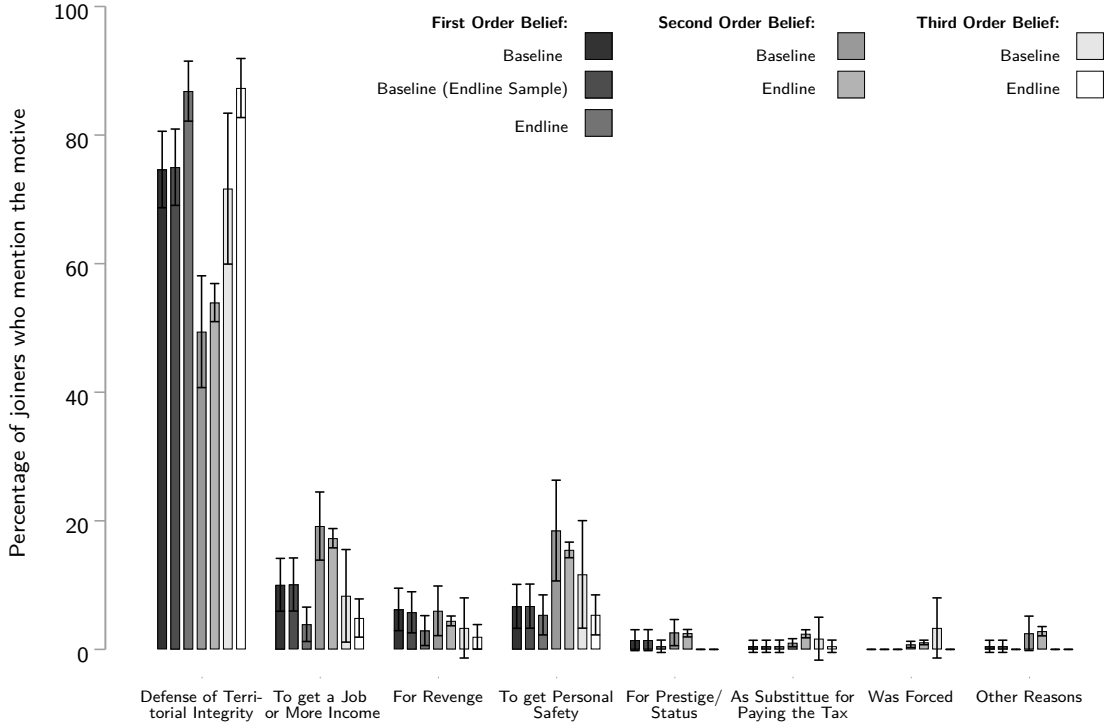
Figure 6: Interpretation: A *Morality of War* Shapes the Expression of Social Preferences

Panel A. The Act of Enlisting as a *Moral Service* to the District

The Act of Enlisting as a Moral Service, Through its Consequences



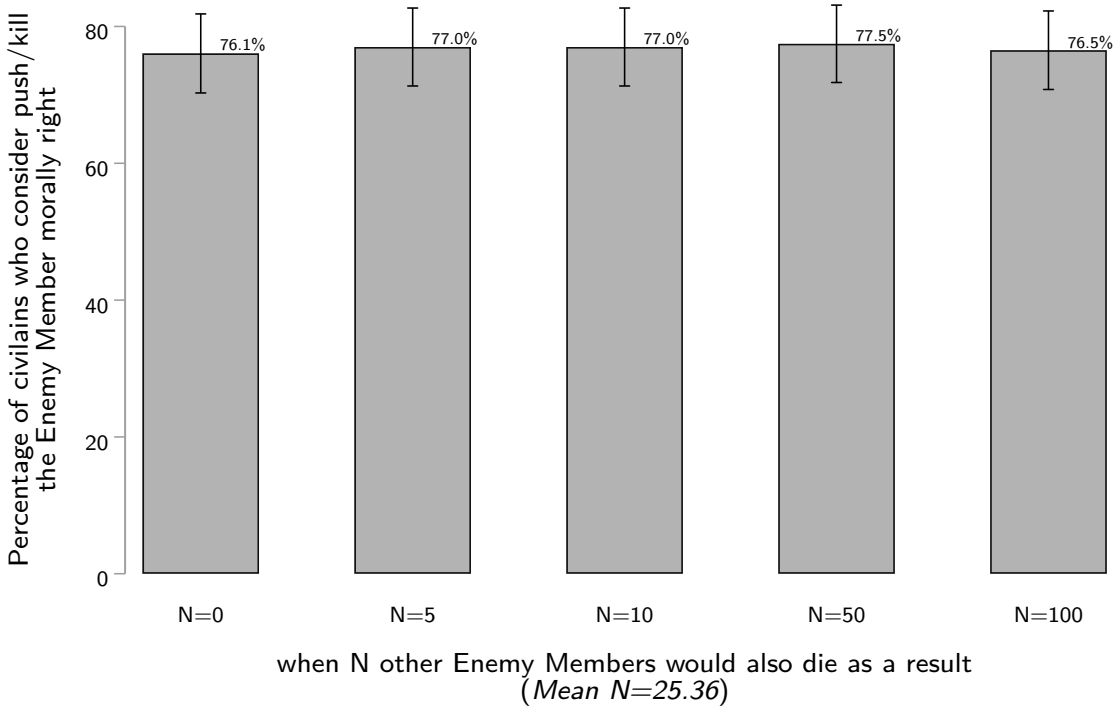
Service to District Defense—Joiners' First, Second, and Third-Order Beliefs



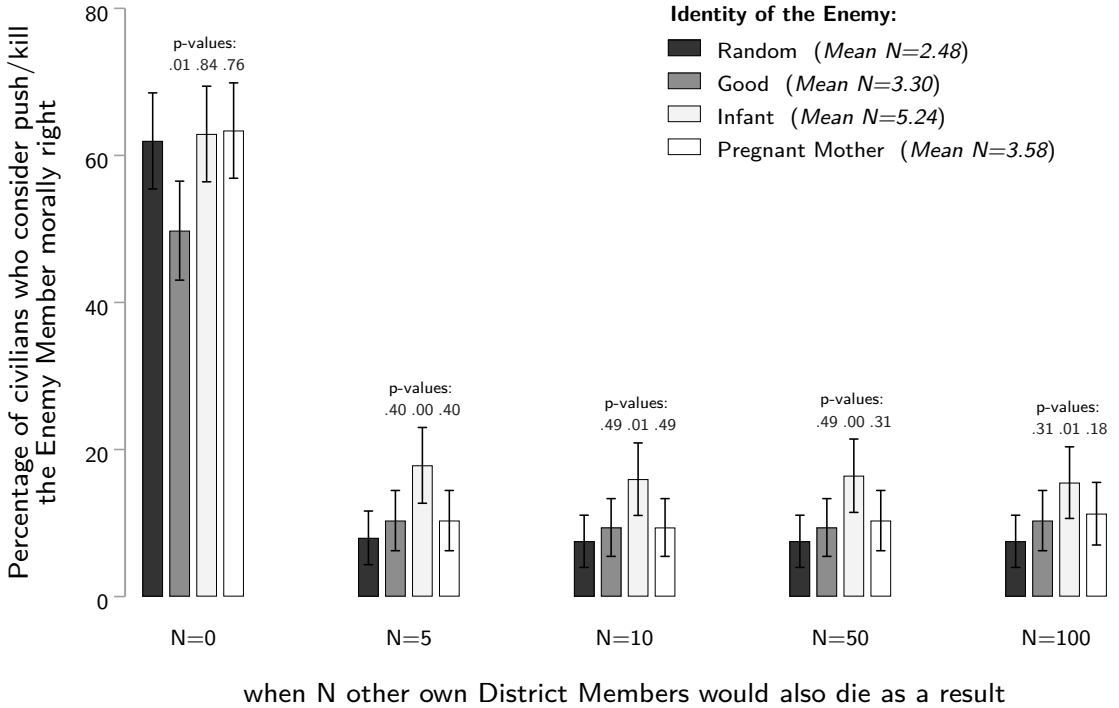
Notes. The upper part of Panel A presents the mean behavioral strategies for the enlisting dilemma, asking respondents whether they think it is morally right to enlist in the group, when their decision to enlist would imply that $N \in \{0, 1, 2, 3, 4, 5, 10, 25, 50\}$ other equally qualified members would not be able to enlist. The lower part of Panel A presents the distribution of first-, second-, and third-order beliefs reported by the joiners in the interview before enlisting, as well as the same measures collected in a survey 8 months after enlisting in the group (which we sometimes refer to as the endline survey). First-order beliefs are the respondent's beliefs about the motives why the respondent enlisted in the group; second-order beliefs are the respondent's beliefs about the distribution of motives why all other joiners enlisted; third-order beliefs are the respondent's beliefs about the distribution of other joiner's beliefs about why the respondent enlisted.

Panel B. The Act of (Lethal) Violence Toward the Enemy a Moral Duty

Moral Dilemma: Killing an Enemy Member, at the Expense of Additional Enemy Members' Lives



Moral Dilemma: Killing an Enemy Member, at the Expense of own District Civilians' Lives



Notes. Panel B presents the mean choices for the killing moral dilemma. The upper part presents it for the dilemma in which respondents are asked whether pushing an enemy member off the bridge is morally right, when doing so results in the death of $N \in \{0, 5, 10, 50, 100\}$ additional enemy members. Vertical brackets indicate confidence intervals. The lower part presents it for dilemma in which respondents are asked whether pushing an enemy member off the bridge—resulting in their death—is morally right, when doing so also results in the death of $N \in \{0, 5, 10, 50, 100\}$ civilians from the District. Various versions of this dilemma were implemented, with the identities of the member of the enemy in the following set: {Random enemy, *Harmless* enemy civilian, Infant enemy, Pregnant Mother enemy}. The choice is elicited through the strategy method, in the following increments for the number of the [other] sacrificed: [0; 5; 10; 25; 50; 100]; after the last choice, the respondent is asked to make their strategy explicit. The specific script for this dilemma is: "*You are on this bridge [see the screen]. Under you, there is a tram, going toward a group of people from the District tied to the rails. Next to you, there is Rwandan Hutu. Choice A: if you do nothing, the tram stops, because it is not going fast, and none will die; Choice B: if you push the member of the enemy off the bridge, the person dies, but this also creates a short-circuit that results in the death of the members tied to the rails. Regardless of what you would do, what should you do? Regardless of your choice, it is important to remind that your choice would carry no consequence (e.g., legal) and that none, in the scenario or in the real world, will be able to identify that the choice was made by you.*" The two-digit numbers indicated above the bars are the p-value of the test of the following null Hypothesis, H_0: the mean in the corresponding bar is the same as the mean in the previous bar.

B. *The Act of Violence Toward Enemy Members as Moral Duty of Collective Retaliation*

The act of violence toward the enemy, anticipated at enlistment and bundled within that service, is best understood through a prevailing morality that defines *killing enemy members* as a *moral duty*. To establish this, we analyze civilians' moral convictions about violence toward enemy members, measured through the choices of the *villagers* in a moral dilemma experiment about violence.

(Mass) Killing the Enemy as a Moral Act: Respondents face the following dilemma (Figure A23, Panel C): if they do nothing, no one dies; if they push an enemy member off a bridge (resulting in their death), this action may, in some scenarios, lead to additional enemy casualties. Respondents choose in private: What *should* you do? (regardless of what you *would* do).[27]

The choices made by civilians reveal that the enemy member mass killings are widely considered *moral*. The majority (76%) consider it moral when there are no other casualties (Figure 6, Panel B, upper level). This fraction remains stable even when the action would result in 100 casualties, suggesting a widespread perception of mass killings of enemy members as moral.

Killing the Enemy as a Moral Duty. In this version of the dilemma, the enemy target is a harmless enemy civilian who has been integrated into society, neutralizing the potential instrumental value of their killing for the District's security. The choices made by the *villagers* reveal that killing enemy members is *mostly* perceived as a moral *duty* rather than morally justified as a means to an end. Among the 60% who initially endorsed killing (Figure 6, Panel B, lower level), 10% no longer endorsed it if the target was harmless—a statistically significant reduction—suggesting these individuals were originally motivated by *consequentialist* reasoning. The remaining 50% continued to endorse killing, indicating that the majority of civilians view killing the enemy as a *moral duty*, aligning with *deontological* morality.[28]

Killing infants and pregnant women (subgroups systematically targeted in genocidal violence) as a moral duty.[29] The fraction considering killing the enemy member rises from 50% to 62% when the enemy civilian is an infant or pregnant woman (Figure 6, Panel B, upper level) consistent with a morality that considers the *elimination of the enemy social group a duty*.[30]

[27]This is an adaptation of the trolley bridge dilemma from survey experiments (Thomson, 1984, Greene, 2013, Bénabou, Falk, and Henkel, 2024), with modified trade-offs.

[28]Figure A24, Panel B, presents the distribution of strategies.

[29]Infants and pregnant women are often targeted in genocide (Graça Machel, 1996, Ndahayo and Dufatanye, 2015). As Graça Machel noted in her 1996 UN report on children in war (Graça Machel, 1996): "The escalation from ethnic superiority to ethnic cleansing to genocide becomes an irresistible process. Killing adults is then not enough; future generations of the enemy—their children—must also be eliminated." Pregnant women are similarly targeted due to their reproductive capacity.

[30]While our design cannot rule out that respondents perceive these killings as indirectly benefiting.

Collective narcissism. A smaller subset of civilians exhibit moral values consistent with *collective narcissism*, defined as "a belief that the exaggerated greatness of one's group is not sufficiently recognized by others" (De Zavala, 2023).[31] Specifically, 10% believe it is morally right to kill an enemy member regardless of the number of District civilians who also die. This fraction doubles (20%) when the target is an enemy subgroup typically targeted in genocide.[32]

This, for a majority of civilians, violence toward the enemy is moralized as a *moral duty*—aligning with a *deontological*, potentially genocidal, morality of violence.

What exactly is the meaning of such *duty*? Respondents were shown images of enemy civilians—an infant, and a pregnant woman—visibly wounded, presented as victimized by lethal violence by a Congolese militia from the District.[33] The answers provided by the *villagers* reveal that these moral convictions reflect a morality of collective retaliation and, to a lesser extent, liberation. Among those who responded, 40% reported feeling *good* about these images (Table A12) and 55% morally *justified* the perpetration of violence that caused it (Table A13); of those, 56% cited collective retaliation for past massacres by the enemy, and 10% cited violence as a mean for collective liberation.[34] Consistent with this the interpretation of these moral convictions as collective retaliation, past family victimization and negative reciprocity are associated with choosing kill in the moral dilemma and in providing justifications (Table 2, Panel A).[35]

However, holding such morality likely generates discomfort by conflicting with standard moral principles of no harm. We find that holding such values is linked *moral disengagement* (Bandura, 2011).[36] Moral disengagement strongly predicts holding violent moral values (Table 2), yet negative reciprocity remains central to explain holding such moral convictions.

[31] We thank Jean-François Bonnefon for this suggestion.

[32] Figure A25 presents the distribution of behavioral strategies played in the first dilemma. Figure A26, Panel A, confirms that this finding is specific to enemy members rather than a general moral permissiveness toward killing. Panels B and C present results from standard bridge dilemmas.

[33] To elicit moral justification, a statement condemning such violence was introduced before the questions.

[34] Table A12 provides examples Reports of enemy-perpetrated massacres are well-documented in historical records and supported by the data: the median household experienced two attacks (Table A11). Examples of justifications invoking liberation include: "Not to mistreat but to push back for peace," "They should leave our country," "Feel good because Mai-Mai help chase them away," "Hutu massacres here are the result of the invasion of our ancestral land," "We beat them because they came to kill us; we defended ourselves," "Hutus must stop coming to Walikale. Mai-Mai should multiply attacks."

[35] Figure A27 presents a text analysis of responses to a module on emotions, in which respondents described the last time they felt anger. Most responses referenced war, killing, or invasion.

[36] Moral disengagement includes dehumanization, moral justification, euphemistic language, advantageous comparison, diffusion of responsibility, displacement of responsibility, distortion of consequences, and victim-blaming. The most prevalent mechanisms among civilians are moral justification—e.g., "*It is okay to beat a Rwandan Hutu because they hurt [District]'s population*"—victim-blaming (68%), exemplified by statements such as "*Rwandan Hutus who have been abused usually did things to deserve it*," and *dehumanization* (Table A14).

Table 2: Roots of the Prevailing Morality of Killing

	Any Justification			Retaliatory Justification			Chooses Kill in Killing Dilemma		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Negative Reciprocity toward the Enemy	0.029**		0.000	-0.005		-0.026**	0.034***		0.016*
	(0.014)		(0.012)	(0.013)		(0.012)	(0.009)		(0.009)
Family Victimized by the Enemy	0.135*		0.128*	0.229***		0.223***	0.050		0.035
	(0.079)		(0.068)	(0.073)		(0.068)	(0.048)		(0.045)
Moral Disengagement (Bandura 2011)		0.115***	0.115***		0.075***	0.085***		0.081***	0.076***
		(0.013)	(0.014)		(0.014)	(0.014)		(0.010)	(0.011)
Joiner ID FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean Dependent Variable	0.55	0.55	0.55	0.32	0.32	0.32	0.61	0.61	0.61
R^2	0.37	0.50	0.51	0.33	0.37	0.42	0.37	0.43	0.44
Observations	290	290	290	290	290	290	631	632	630

Notes. This table analyzes the role of motivated reasoning ("moral disengagement" (Bandura, 2011)) and of negative reciprocity against past victimization on the emergence of violent moral convictions. To do so, we estimate the following baseline regression in each panel, where the dependent variables differ across panels and where we sequentially add controls as indicated in each panel: $y_i = \alpha + \beta_1 NR_i + \beta_2 V_i + \beta_3 MD_i + \epsilon_i$ where NR_i, V_i, MD_i are the measures of negative reciprocity toward the Enemy, an indicator taking value 1 if the family of the respondent was previously victimized by the Enemy (as shown in Table A11, the median value is 1, which is true for 53% of respondents/observations), and Bandura (2011)'s Moral Disengagement scale about physically hurting enemy members. The dependent variable y_i is an indicator taking value 1 if the respondent observing a photo of an enemy civilian victimized by the group provides any justification for such violence and zero otherwise (Columns 1-3), an indicator taking value 1 if the respondent observing a photo of a Rwandan Hutu victimized by the group provides a *retaliatory* justification for such violence and zero otherwise (Columns 4-6), and an indicator taking value 1 if the respondent chooses "Push/Kill" the enemy in the Bridge Killing dilemma in which there are no other casualties ensuing this action (Columns 7-9).

C. *Implications for Interpreting Enlistment and the Key Relationship*

This analysis of moral convictions helps interpret the empirical findings on social preferences and on selection, namely, the structure of unbundled social preferences in the population, the pattern of enlistment, and the key relationship.

First, consistent with recent findings in neuroscience, our findings suggest that "moral conviction about sociopolitical issues serves to increase their subjective value, overriding natural aversion to interpersonal harm" (Workman, Yoder, and Decety, 2020), thus the morality of killing may shape the selective suppression of violence aversion toward the enemy.

Second, the morality of enlisting shapes the outcome-prosocial reward inherent in joining, helping explain the selection of joiners. Does this reflect idiosyncratic variation in social preferences or instead on moral convictions? Table 3 shows that while joiners share similar violent moral values and dehumanize more, neither moral disengagement nor reciprocity explains why they are more prosocial toward the District and less empathetic toward the enemy. This underscores the interpretation of the role of the prevailing morality.

Third, we found that this pattern of selection hinged on a weakly negative correlation between outcome-prosociality toward the District and violence aversion toward the enemy. Given the higher levels of A.M. empathy displayed by outcome-prosocial individuals toward the District, these findings offer suggestive evidence for the existence of the following mechanism: those most committed to the District (the outcome-prosocial toward the District) uphold the *prevailing morality of killing* more strongly, suppressing cognitive empathy toward the enemy and thereby suppressing A.M. empathy. Section Appendix C formalizes this mechanism.

In sum, our finding of a prevailing morality of war—which moralizes both enlisting and violence against enemy members as virtuous actions—shapes the expression of social preferences. Specifically, it defines enlisting as yielding an outcome-prosocial reward while simultaneously providing a violence inclination reward of violence toward the enemy. This, in turn, helps explain both the selection of enlisting behavior and the correlation underpinning the key relationship.[37]

[37]Controlling for moral disengagement, family victimization, and negative reciprocity (Table A15) leaves the weakly negative or absent relationship between outcome-prosocial preferences toward the District and aversion to violence against the enemy qualitatively unchanged. This rules out the possibility that the key relationship is driven by variation in moral values, preferences for reciprocity toward the enemy, or moral disengagement. Figure A28 replicates the original scatter plots and calculations on residualized data, showing that the inclusion of these three variables leaves the pattern of selection unchanged. Given the higher capacity for violence aversion by outcome-prosocial individuals toward the District, these individual could show more, not less, empathy toward enemy members as well. The fact that we instead observe a flat or weakly negative relationship suggests that these prosocial individuals must be actively suppressing their cognitive empathy toward the enemy even more strongly than others, relative to their generally higher capacity for empathy.

Table 3: Outcome Prosociality toward the District, not Differences in Individual-Held Morality, is Associated with Enlisting

	Dependent Variable: Enlists				
	(1)	(2)	(3)	(4)	(5)
Prosociality toward the District	0.063***	0.056***	0.049***	0.043**	0.038**
	(0.019)	(0.019)	(0.018)	(0.019)	(0.019)
A.M. Empathy toward the Enemy	-0.025***	-0.005	0.008	0.025**	0.032***
	(0.008)	(0.010)	(0.009)	(0.012)	(0.011)
Cognitive Empathy toward the Enemy		-0.027***	-0.023***	-0.029***	-0.025***
		(0.009)	(0.008)	(0.009)	(0.008)
Moral to Join			0.395***		0.364***
			(0.052)		(0.055)
Chooses to Kill in Dilemma				0.005	-0.018
				(0.064)	(0.060)
Moral Disengagement (Bandura 2011)				0.077***	0.059***
				(0.015)	(0.015)
Negative Reciprocity toward the Enemy				-0.009	0.005
				(0.012)	(0.011)
Family Victimized by the Enemy				-0.127**	-0.118**
				(0.053)	(0.051)
Joiner ID FEs	✓	✓	✓	✓	✓
Mean Dependent Variable	0.33	0.33	0.33	0.33	0.33
R^2	0.05	0.07	0.18	0.13	0.21
Observations	635	635	635	630	630

Notes. This table presents the estimates from a regression in which the dependent variable is an indicator taking value 1 if the respondent belongs to the set of joiners and zero otherwise. The sample in all regressions includes joiners, the *household members* (a randomly selected young male adult within each joiner's household), and the *villagers* (a randomly selected young male adult within a randomly selected household in each joiner's village), in which case the sample is 635. Regressions include fixed effects for joiner ID, which generate groups for each joiner, his corresponding *household member*, and his corresponding *villager*. Standard errors are not clustered.

6. Violence Preferences after Enlisting—Selective Desensitization

In this section, we analyze the change of *violence preferences* after enlisting, shedding light both on the dynamics of the relationship between violence preferences and violence participation, and investigating the malleability of such preferences.

To measure violence preferences following enlisting, we conducted a follow-up of the joiners eight months after they enlisted, between September 2022 and May 2023 (Figure A7) (henceforth, *endline*) and collected (again) data on *all* the measures of violence preferences and a a sub-set of the measures of outcome preferences (the two-sided dictator game).[38] The measures of violence preferences collected eight months after joining reveal that violence preferences are *selectively* malleable and that enlisting shapes them through *selective desensitization*.

Considering violence preferences toward the District, the levels of violence aversion (A.M. empathy) as well as cognitive empathy toward district members are comparable to enlistment; they are even statistically significantly larger (Figure 7). Thus, joiners' violence preferences and beliefs about others' pain and suffering over District members remain intact after enlisting.

In contrast, joiners exhibit substantial *desensitization* to to the pain and suffering of the enemy: while their violence aversion is almost halved and joiners have greater enjoyment of enemy suffering and are more to choose kill in the video game (revealed inclination), their cognitive empathy rises from 4.98 to 6.92—a 38% increase.[39] Taken together, these changes imply increased tolerance (henceforth, *desensitization*) to knowledge about suffering of the enemy.

To test whether selective desensitization is statistically significant, we examine the following ratio $\frac{\text{A.M. empathy}}{1 + \text{cognitive empathy}}$, which increases monotonically in sensitivity to pain and suffering that is cognitively recognized. The ratio is constant and close to one toward District members but drops from .4 to .24 (statistically significant), underscoring the increased gap in cognitive empathy vs latent violence aversion (Figure 7, Panel C).[40]

In sum, enlisting is followed by significant *desensitization* toward enemy suffering. This mechanism differs from the *self-deception* that differentiated joiners from civilians.

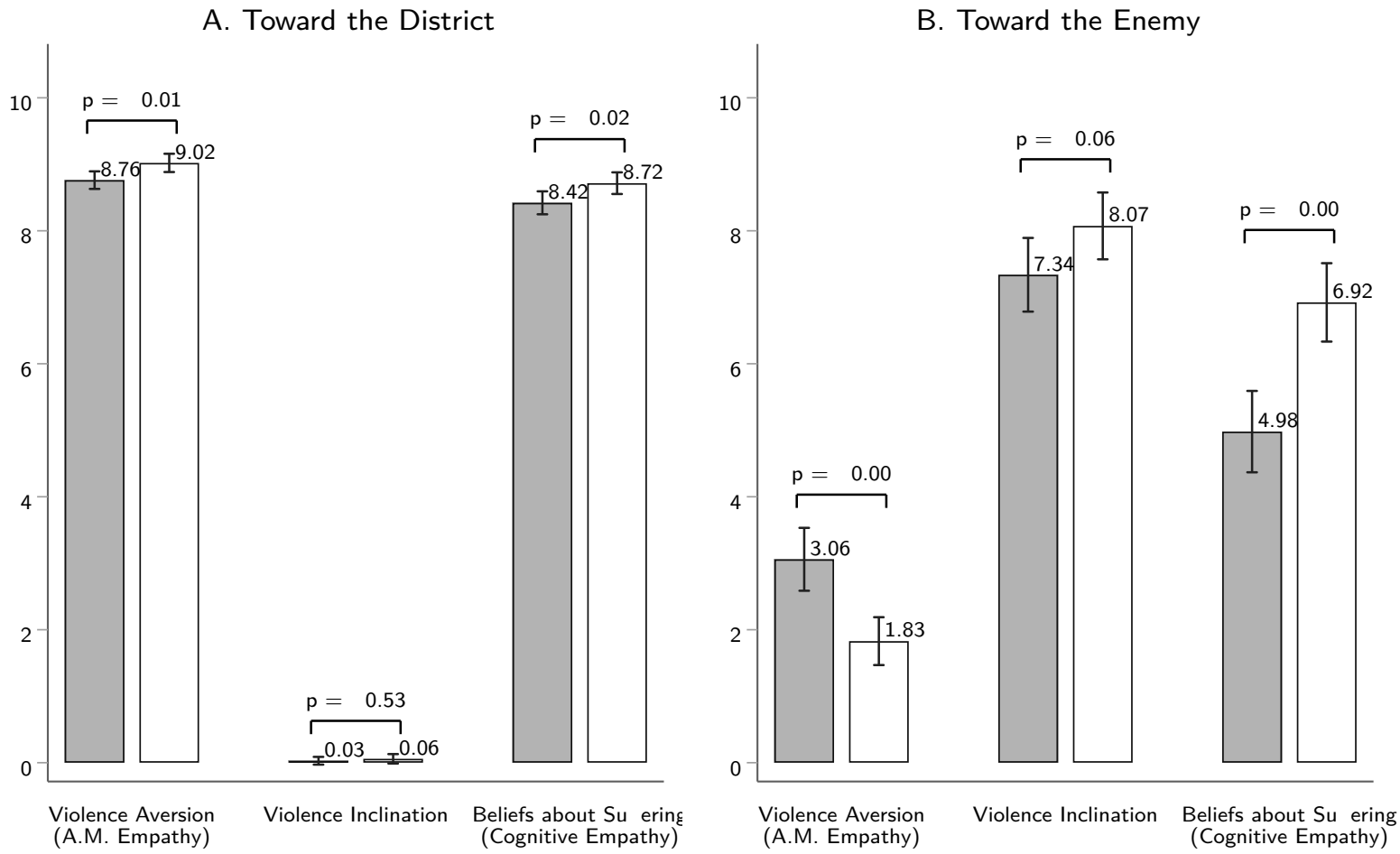
[38]The latter focus reflects time and budgetary constraints to collect data on active participants. The conditions of the interview were identical to those of the survey administered prior to enlisting.

[39]This shift reflects *homogenization* of violence aversion toward enemy members within the group (Figure A29).

[40]Outcome preferences become less prosocial generally and more anti-social toward enemy members (Figure A30). Figure A31 shows that aversion to violence against the District's group members increases, consistent with a restructuring of empathy towards their group—while keeping constant and even increasing aversion to violence against District members.

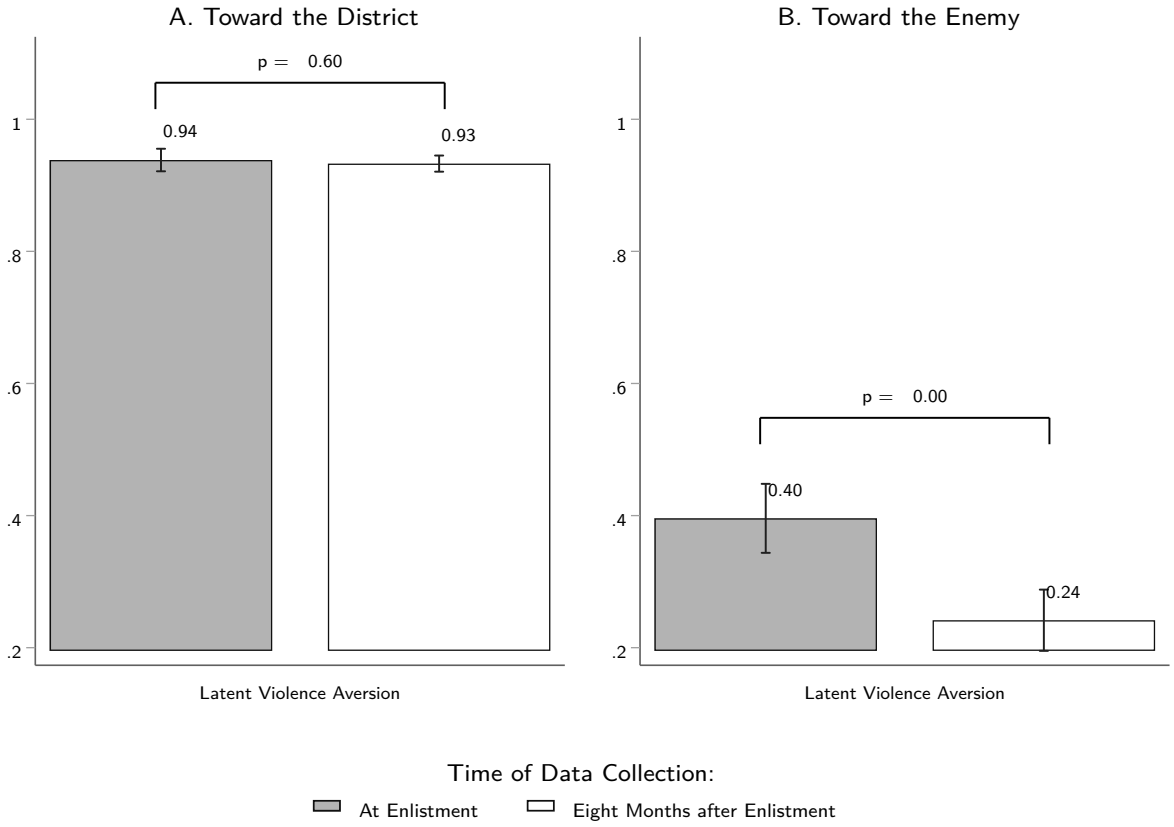
Figure 7: Violence Preferences and Beliefs after Enlisting—*Selective Desensitization*

Panel A. Violence Preferences and Beliefs after Enlisting



Notes. This figure presents the analysis of violence preferences after enlistment. The figure presents the means of each dimension towards the District (left panel) and towards the enemy (right panel), at time of enlisting (as measured in the baseline) and eight months after enlisting (as measured in the endline). The p-values in those quadrants are the p-values of the two-sided test of equality of means of baseline vs. endline for each corresponding variable. Outcomes are rated from 0 to 10. *Sample*. We use n=168 joiners in this figure. This sample size is obtained as follows. Of 213 phone calls received about individuals who enlisted, 210 were interviewed at baseline (the remaining 3 had the civilian surveys done, but not the joiners), of which 173 are still present at endline, 18 are dead, and 22 are evaded, hence the remaining sample is 173. A remaining n=5 have missing data: 2 have no endline data, one lacks a self-assessment survey and the other the survey altogether. All comparisons include controls for treatment indicators of a separate experiment conducted among new joiners in which the treatments (perspective taking designed to promote empathy toward civilians, international humanitarian law and norms training) were administered *after the baseline*.

Panel B. Selective Desensitization—*Sensitization Ratio*



Notes. To test whether selective desensitization is statistically significant, this panel presents the levels of the following ratio $\frac{\text{A.M. empathy}}{1 + \text{cognitive empathy}}$, which increases monotonically in sensitivity to pain and suffering that is cognitively recognized. The p-value is the p-value of the test of equality of means between the ratio at enlistment vs. eight months after enlisting.

7. Validation of the Concept and its Measurement

In this section, we investigate whether the measures of violence preferences that we have introduced in our analysis are related to outcomes of interest for social science.

Validation with Psychophysiological Measures of Emotional Arousal

Research in psychophysiology has shown that affective-motivational empathy is associated with emotional arousal in response to witnessing another person's pain, as measured by Electrodermal Activity (EDA), a psychophysiological measure that reflects emotional arousal and is hard to

manipulate (Avenanti, Sirigu, and Aglioti, 2010).[41] Existing research has elicited a link between *empathic responses* and emotional arousal upon observing someone else in pain. Figure A32, Panel A, shows that violence preferences and beliefs are associated to true emotional responses.

The affective-motivational components of empathy, as well as cognitive empathy, towards a target are associated with higher psychophysiological emotional arousal upon witnessing pain inflicted *on that target*. When the target is a *District member*, this holds only for cognitive empathy—reflecting the low variation of empathy in these dimensions since, as we have shown, the affective-motivational components of empathy are high for most respondents. Violence *inclination* towards a target is associated with significantly *lower* psychophysiological emotional arousal upon witnessing pain inflicted *on that target*. Indeed, enjoyment of violence as well as *revealed inclination* are both associated with *less* psychophysiological emotional arousal—consistent with violence inclination correctly reflecting the *absence* of an empathic response.

In contrast, *pro- and anti-social preferences* (two-sided dictator game) are not significantly associated with psychophysiological arousal when observing pain inflicted on the target.[42] These associations are comparable after controlling for social preferences, suggesting that violence preference is a better measure of emotional responses associated to violence perpetration.

Validation with Third-Party Assessments of Actual Behavior

To measure violent behavior, we collected assessments eight months after joining from interviews with: joiner's supervisor (his direct commander), a randomly selected peer (a combatant) who knew the joiner, and the joiners themselves. These assessments were conducted independently,

[41] Electrodermal Activity (EDA) measures the activity of sweat glands, which are controlled by the sympathetic nervous system, and is a reliable indicator of emotional or stress responses. To measure this, respondents wore Empatica E4 bracelets designed to collect EDA data while completing the *suffering empathy task*. These measurements enabled us to detect emotional arousal in response to witnessing others' pain. Due to the complexity and cost of implementing the bracelets, data were collected from a non-random subset of the sample, primarily focusing on the joiners. This resulted in successful data collection from 44 randomly selected youth (out of 213), 59 youth from joiners' households (out of 213), and 125 joiners (out of 213). SCR (Skin Conductance Response) is the intensity of electrical activity (measured in MicroSiemens) after filtering out the tonic response. A detailed explanation is in Appendix E.

[42] Using psychophysiological emotional arousal instead of violence preferences, we also replicate the finding of a strong anti-enemy bias in aversion to violence. Figure A34 shows that witnessing pain inflicted upon a District member induces more emotional arousal than witnessing the pain of an enemy member. It also confirms the distinctive violence aversion of joiners versus non-joiners. Figure A35 illustrates that non-joiners experience almost twice the psychophysiological emotional arousal of joiners when observing an enemy member suffering; however, joiners and non-joiners exhibit similar levels of psychophysiological emotional arousal when witnessing pain inflicted on a District member. Figure A33 demonstrates that the association with the measures of violence aversion, collected concurrently, holds across various windows and respondent samples after the *suffering empathy task* prime.

with no shared knowledge between them.[43] These interviews allow us to measure the relationship between the joiners' behavior, measured eight months after enlisting, to the joiners' contemporaneous violence preferences measured eight months after enlisting.

Figure A32, Panel B, shows that the measures of violence preferences are correlated with contemporaneous violent behavior. First, the *affective-motivational* components of empathy towards a target are associated to less violent behavior against that target. For cognitive empathy, this relationship vanishes for the enemy, consistent with the *desensitization* we just documented. Second, *violence inclination* toward the enemy is associated with violent behavior against the enemy.[44] These findings are unaffected by controlling for outcome preferences.

In contrast, endline outcome preferences are only related to violent behavior when directed toward enemy members. This association vanishes after controlling for violence preferences, suggesting that it was driven by violence preferences.

This empirical validation underscores the contribution of distinguishing violence preferences and outcome preferences for the analysis of violence.

8. Conclusion

The conventional view in economics holds that prosocial preferences primarily promote cooperation. This study finds that, when structured by moral values that define violence as prosocial and enabled by humans' capacity for empathy suppression, they can also drive violence.

At a fundamental level, our findings challenge simple views of generalized prosociality. Rather than individuals being inherently "good" or "bad," those most committed to their collective selectively suppress empathy toward certain groups while maintaining compassion toward others—upholding a societal "morality of war" that diverges from conventional peacetime moral principles by morally sanctioning violence. The finding of malleability of affective-motivational empathy, both through *self-deception* and through *desensitization*, echoes Browning (2017), who argued that genocidal violence is shaped not only by pre-existing *dispositions* but also by situational and organizational forces. It also aligns with Solzhenitsyn (2007)'s insight that "the line dividing good and evil cuts through the heart of every human being," reflected in the chilling normalcy of genocide alongside everyday life, as portrayed in the context of the Holocaust (Glazer, 2023).

[43] After consulting with the supervisor, we identified a subset of peers who knew the joiner well enough to provide an assessment. A random selection tool was then used through ODK Collect to choose the peer for the assessment.

[44] For civilians, this analysis is irrelevant, as the sample of joiners who exhibit enjoyment of violence or revealed violence inclination toward District members is virtually zero (n=9), making the relationship an invalid test.

These findings expand the dominant view of violence in economics, which emphasizes economic incentives as the *primary* driver of participation (Becker, 1968). While fully consistent with their role—and indeed supporting it—our results show that individuals are motivated not only by economic payoffs but also by prosocial rewards—fulfilling moral commitments to their collective. Moral values do not merely justify violence *ex post*; they shape participation *ex ante*. This provides a rationale for why getting to peace agreements is hard, as people's preferences can be over *acts* and moral values can make compromises much harder. Furthermore, violent groups could be seen not as unitary rational actors but as coalitions of (potentially) morally motivated combatants and (potentially) strategically calculating elites. Selfish economic incentives and moral commitments are not mutually exclusive, but can operate on different layers.

Our findings also question the view of conflict which posits its analysis merely as a "bargaining failure." Economists have assumed that violence is inefficient because it destroys resources (Fearon, 1995, Besley and Mueller, 2012). However, if moral values are embedded in preferences, conflict cannot always be reduced to failed bargaining. For some, perpetrating violence generates non-transferable utility shaped by moral values that ascribe meaning and utility to violence.

Our findings reveal that when effort is grounded in moral *duty*, rooted in *deontological* morality, it disrupts the foundations of the utilitarian bargaining framework, with implications beyond conflict to welfare economics. Actions driven by duty lack a price and cannot be traded, rendering utilitarian notions of social optimality inadequate—an issue distinct from classic problems of cardinality. This connects to an underexplored frontier in rationalist models of bargaining failure: "indivisibilities," when they arise not from technological constraints, but from the non-tradeability of subjective moral evaluations. Moral duties may thus represent a fundamental form of indivisibility, re-framing conflict as a reflection of moral commitments at the boundaries of rationalist bargaining failures and subjective explanations for conflict. Future research should investigate how these moral indivisibilities are psychologically *structured*, embedding evaluations of social welfare in the participants' moral convictions.

The significance of our discovery of a "morality of war" extends beyond militias. Mass mobilization—whether in state militaries, insurgencies, or revolutionary movements—has historically relied on moral justifications, framing violence as a duty or a righteous contribution to the collective (Besley and Persson, 2010). More broadly, colonial institutions, apartheid regimes, and authoritarian states have sustained inequality and extractive systems not merely through coercion, but through ideological frameworks that confer moral legitimacy, systematically suppressing

empathy toward those who bear their costs. Beyond providing intrinsic value to institutions, ideology may function by overriding the natural aversion rooted in human empathy, as evidenced by the formation of ideologies dehumanizing persons of African descent coincidentally with the spread and profitability institutions of based on their enslavement (Williams, 1944).

Our findings also enrich policy approaches that rely on economic incentives in violence prevention. If participation is driven by moral commitments and social preferences, demobilization and reintegration must address moral convictions. In ongoing research, for example, we experimentally administered a year-long perspective program among combatants, only to find that this accelerates the rate of exit by triplicating the probability that they exit the group after one year—consistent with the role of violence aversion in violence participation.

Studying the question of violence remains an extraordinary empirical challenge. Our findings reflect an unprecedented effort, involving ten years of building a personal relationship with a militia, assessing the feasibility of conducting research within its boundaries, and leveraging its HR organizational structure to collect data. Yet, they capture only a snapshot of a dynamic process. Unlocking these constraints could not only serve as a check on external validity but also open the exploration of microfoundations of violence and coercive institutions.

References

Abbink, K., and A. Sadrieh (2009): "The pleasure of being nasty," *Economics Letters*, 105(3), 306–308.

Arendt, H., and A. Elon (2006): *Eichmann in Jerusalem: A Report on the Banality of Evil*, Penguin Classics. Penguin Publishing Group.

Athias, L. (2025): "On the Normative Motivation of States: State Universalism in Western Post-Enlightenment Societies," Working paper.

Avenanti, A., A. Sirigu, and S. M. Aglioti (2010): "Racial Bias Reduces Empathic Sensorimotor Resonance with Other-Race Pain," *Current Biology*, 20(11), 1018–1022.

Bandura, A. (2011): *Moral Disengagement: How People Do Harm and Live with Themselves*. Worth Publishers.

Batson, C. D., J. Fultz, and P. A. Schoenrade (1987): "Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences," *Journal of personality*, 55(1), 19–39.

Bauer, M., C. Blattman, J. Chytilová, J. Henrich, E. Miguel, and T. Mitts (2016): "Can War Foster Cooperation?," *Journal of Economic Perspectives*, 30(3), 249–274.

Becker, G. S. (1968): "Crime and Punishment: An Economic Approach," *Journal of Political Economy*, 76(2), 169–217.

Benmelech, E., and E. F. Klor (2020): "What Explains the Flow of Foreign Fighters to ISIS?," *Terrorism and Political Violence*, 32(7), 1458–1481.

Bernini, A., G. Facchini, M. Tabellini, and C. Testa (Forthcoming): "Black Empowerment and White Mobilization: The Effects of the Voting Rights Act," *Journal of Political Economy*.

Besley, T., and M. Ghatak (2005a): "Competition and Incentives with Motivated Agents," *American Economic Review*, 95(3), 616–636.

——— (2005b): "Competition and Incentives with Motivated Agents," *American Economic Review*, 95(3), 616–636.

Besley, T., and H. Mueller (2012): "Estimating the Peace Dividend: The Impact of Violence on House Prices in Northern Ireland," *American Economic Review*, 102(2), 810–833.

Besley, T., and T. Persson (2009): "The Origins of State Capacity: Property Rights, Taxation, and Politics," *American Economic Review*, 99(4), 1218–1244.

——— (2010): "State Capacity, Conflict, and Development," *Econometrica*, 78(1), 1–34.

——— (2011): "The Logic of Political Violence," *The Quarterly Journal of Economics*, 126(3), 1411–1445.

Bowles, S., and H. Gintis (2011): *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton University Press, Princeton, NJ.

Browning, C. R. (2017): *Ordinary Men: Reserve Police Battalion 101 and the Final Solution in Poland*. Harper Perennial.

Butinda, L. (2021): *D'Enfant a la Prise de l'Arme: Un Voyage Aller-Retour*. Unpublished Autobiography.

Bénabou, R., A. Falk, and L. Henkel (2024): "Ends Versus Means: Kantians, Utilitarians, and Moral Decisions," *SSRN Electronic Journal*.

Bénabou, R., and J. Tirole (2016): "Mindful Economics: The Production, Consumption, and Value of Beliefs," *Journal of Economic Perspectives*, 30(3), 141–64.

Cagé, J., et al. (2023): "Political Collaborations and WWII: The Role of Pétain's Networks," *Working Paper*.

Campante, F., and D. Yanagizawa-Drott (2015): "The Intergenerational Transmission of War," *Journal of the European Economic Association*, 13(1), 135–165.

Cerina, R., C. Barrie, N. Ketchley, and A. Y. Zelin (2023): "Explaining Recruitment to Extremism: A Bayesian Hierarchical Case–Control Approach," *Political Analysis*, pp. 1–19.

Choi, J.-K., and S. Bowles (2007): "The Coevolution of Parochial Altruism and War," *Science*, 318(5850), 636–640.

Cikara, M., V. Fouka, and M. Tabellini (2022): "Hate crime towards minoritized groups increases as they increase in size-based rank," *Nature Human Behaviour*, 6(11), 1537–1544.

De Zavala, A. (2023): *The Psychology of Collective Narcissism: Insights from Social Identity Theory*. Taylor & Francis.

Decety, J. (2011): *Empathy: From Bench to Bedside*. The MIT Press.

Decety, J., and P. Jackson (2004a): "The Functional Architecture of Human Empathy," *Behavioral and cognitive neuroscience reviews*, 3, 71–100.

Decety, J., and P. L. Jackson (2004b): "The functional architecture of human empathy," *Behavioral and Cognitive Neuroscience Reviews*, 3(2), 71–100.

Enke, B., R. Rodriguez-Padilla, and F. Zimmermann (2023): "Moral Universalism and the Structure of Ideology," *Review of Economic Studies*, 90(4), 1934–1962.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde (2018): "Global Evidence on Economic Preferences," *The Quarterly Journal of Economics*, 133(4), 1645–1692.

Falk, A., A. Becker, T. Dohmen, D. Huffman, and U. Sunde (2022): "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences," *Management Science*, 69.

Fearon, J. D. (1995): "Rationalist Explanations for War," *International Organization*, 49(3), 379–414.

Ferguson, A. M., and M. Inzlicht (2022): "Reliability of the empathy selection task, a novel behavioral measure of empathy avoidance," *Behavior Research Methods*, 55, 1–18.

Fouka, V. (2020): "Backlash: The Unintended Effects of Language Prohibition in US Schools after World War I," *The Review of Economic Studies*, 87(1), 204–239.

Fouka, V., and M. Tabellini (2022): "Changing Ingroup Boundaries: The Effect of Immigration on Race Relations in the United States," *American Political Science Review*, 116(3), 968–984.

Geert Hofstede, G. J. H., and M. Minkov (2010): *Cultures and Organizations: Software of the Mind*. McGraw-Hill, 3rd edn.

Glazer, J. (2023): "The Zone of Interest," .

Graham, J., J. Haidt, and B. A. Nosek (2009): "Moral Foundations Dictionary," Version 1.0.

Graça Machel (1996): "Impact of Armed Conflict on Children: Report of the expert of the Secretary-General, Ms. Graça Machel, submitted pursuant to General Assembly resolution 48/157," Discussion paper, Human Rights Watch.

Greene, J. (2013): *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin Publishing Group.

Grosjean, P. (2014): "A History of Violence: The Culture of Honor and Homicide in the US South," *Journal of the European Economic Association*, 12(5), 1285–1316.

Gómez, J., M. Verdú, A. González Megías, and M. Méndez (2016): "The phylogenetic roots of human lethal violence," *Nature*, 538.

Habyarimana, J., M. Humphreys, D. Posner, and J. Weinstein (2007): "Why Does Ethnic Diversity Undermine Public Goods Provision?," *American Political Science Review*, pp. 709–725.

Hartman, A. C., and B. S. Morse (2020): "Violence, Empathy and Altruism: Evidence from the Ivorian Refugee Crisis in Liberia," *British Journal of Political Science*, 50(2), 731–755.

Henn, S., and C. Huff (2024): "The Local Memory of Repression, and Who Fights," *World Politics*, 76, 219–258.

Humphreys, M., and J. M. Weinstein (2008): "Who Fights? The Determinants of Participation in Civil War," *American Journal of Political Science*, 52(2), 436–455.

Jha, S. (2013): "Trade, institutions, and ethnic tolerance: Evidence from South Asia," *American Political Science Review*, 107(4).

Jha, S., and M. Shayo (2019): "Valuing Peace: The Effects of Financial Market Exposure on Votes and Political Attitudes," *Econometrica*, 87(5), 1561–1588.

Lamm, C., C. D. Batson, and J. Decety (2007): "The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal," *Journal of cognitive neuroscience*, 19(1), 42–58.

Levenson, M. R., K. A. Kiehl, and C. M. Fitzpatrick (1995): "Assessing psychopathic attributes in a noninstitutionalized population," *Journal of Personality and Social Psychology*, 68(1), 151–158.

Mugaruka, M., G. Marchais, R. Sánchez de la Sierra, and Q. Wu (2024): "Social Origins of Militias: The Extraordinary Rise of the Outraged Citizens," *Working paper*.

Ndahayo, E., and A. Dufatanye (2015): *La violence politico-militaire contre les femmes au Rwanda – De Ndabaga a Ingabire*. Editions Sources du Nil ed. Lille: Collection Pyramides.

Nunn, N., Y. Cao, B. Enke, A. Falk, and P. Giuliano (2021): "Herding, Warfare, and a Culture of Honor: Global Evidence," *National Bureau of Economic Research Working Paper*, (29250).

Oliver Scott Curry, D. A. M., and H. Whitehouse (2019): "Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies," *Current Anthropology*, 60(1), 47–69.

Oliver Scott Curry, L. D.-W., and A. C. Little (2021): "Cooperation, Trust, and Human Morality: Examining Cross-Cultural Variation in Moral Foundations," *Journal of Cross-Cultural Psychology*, 52(6), 552–572.

Pape, R. A., J. Decety, K. Ruby, A. Albanez Rivas, J. Jessen, and C. Wegner (2017): "The American Face of ISIS: Analysis of ISIS-Related Terrorism in the US, March 2014–August 2016," .

Pinker, S. (2011): *The Better Angels of Our Nature: Why Violence Has Decline*. Viking Books, New York.

Qian, N., and M. Tabellini (2024): "Racial Discrimination and the Social Contract: Evidence from U.S. Army Enlistment during WWII," *Review of Economic Studies*, Forthcoming.

Sánchez de la Sierra, R. (2021): "Whither Formal Contracts?," *Econometrica*, 89(5), 2341–2373.

Scacco, A. (2016): *Anatomy of a Riot Participation in Ethnic Violence in Nigeria*. NYU.

Schmitt, M., A. Baumert, M. Gollwitzer, and J. Maes (2010): "The Justice Sensitivity Inventory: Factorial validity, location in the personality facet space, demographic pattern, and normative data," *Social Justice Research*, 23(2-3), 211–238.

Solzhenitsyn, A. I. (2007): *The Gulag Archipelago 1918–1956: An Experiment in Literary Investigation*. Harper Perennial Modern Classics, New York.

Tabellini, M. (2020): "Gifts of the Immigrants, Woes of the Natives: Lessons from the Age of Mass Migration," *Review of Economic Studies*, 87(1), 454–486, Also Harvard Business School Working Paper, No. 19-005, July 2018. Available from Cato Institute, Microeconomic Insights, VOX, Broadstreet, Cato Institute, and Oxford University Press's Blog.

Thomson, J. J. (1984): "The trolley problem," *Yale LJ*, 94, 1395.

Tilly, C. (1985): "War Making and State Making as Organized Crime," in *Bringing the State Back In*, ed. by T. S. Peter Evans, Dietrich Rueschemeyer, Cambridge. Cambridge University Press.

Vogel, C., G. Salvaggio, P. Boisselet, and J. K. Stearns (2021): "The Landscape of Armed Groups in Eastern Congo: Missed Opportunities, P-rotracted Insecurity and Self-Fulfilling prophecies," Discussion paper, Kivu Security Tracker - Congo Research Group - NYU Center on International Cooperation, New York.

Williams, E. (1944): *Capitalism and Slavery*. University of North Carolina Press, Chapel Hill.

Workman, C., K. Yoder, and J. Decety (2020): "The Dark Side of Morality - Neural Mechanisms Underpinning Moral Convictions and Support for Violence.," *AJOB Neurosci*, (Oct-Dec;11(4):269-284).

Wrangham, R. (2019): *The Goodness Paradox: The Strange Relationship Between Virtue and Violence in Human Evolution*. Pantheon Books, New York.

Zaki, J., and K. Ochsner (2012a): "The neuroscience of empathy: progress, pitfalls and promise," *Nature neuroscience*, 15(5), 675–680.

Zaki, J., and K. N. Ochsner (2012b): "The neuroscience of empathy: progress, pitfalls and promise," *Nature neuroscience*, 15(5), 675—680.

Zimbardo, P. (2008): *The Lucifer Effect: How Good People Turn Evil*. Rider.

Appendix A. Mathematical Appendix

A.. Moral Consequentialism: Framework and Predictions

To interpret these choices, we distinguish a component driven by the expected impact of an individual's participation and a component reflecting the intrinsic moral value of participation: $s_C = s_C^d + s_C^c$ where s_C^d represents deontological motivation, s_C^c captures consequentialist motivation, and n denotes the number of other participants:

$$U_0(n) = w_0 + s_C^c * n$$
$$U_1(n) = w_1 + \ s_C^c * n + \left[s_C^c + s_C^d - \lambda_E\right].$$

The differential payoff from enlisting can be written as the following expression:

$$U_J(n) = \underbrace{w_0 + s_C^c * n}_{\text{baseline}} + J * \left(\underbrace{(w_1 - w_0 - \lambda + s_C^d)}_{\text{d: benefit to self}} + \underbrace{s_C^c}_{\text{benefit to District}} \right),$$

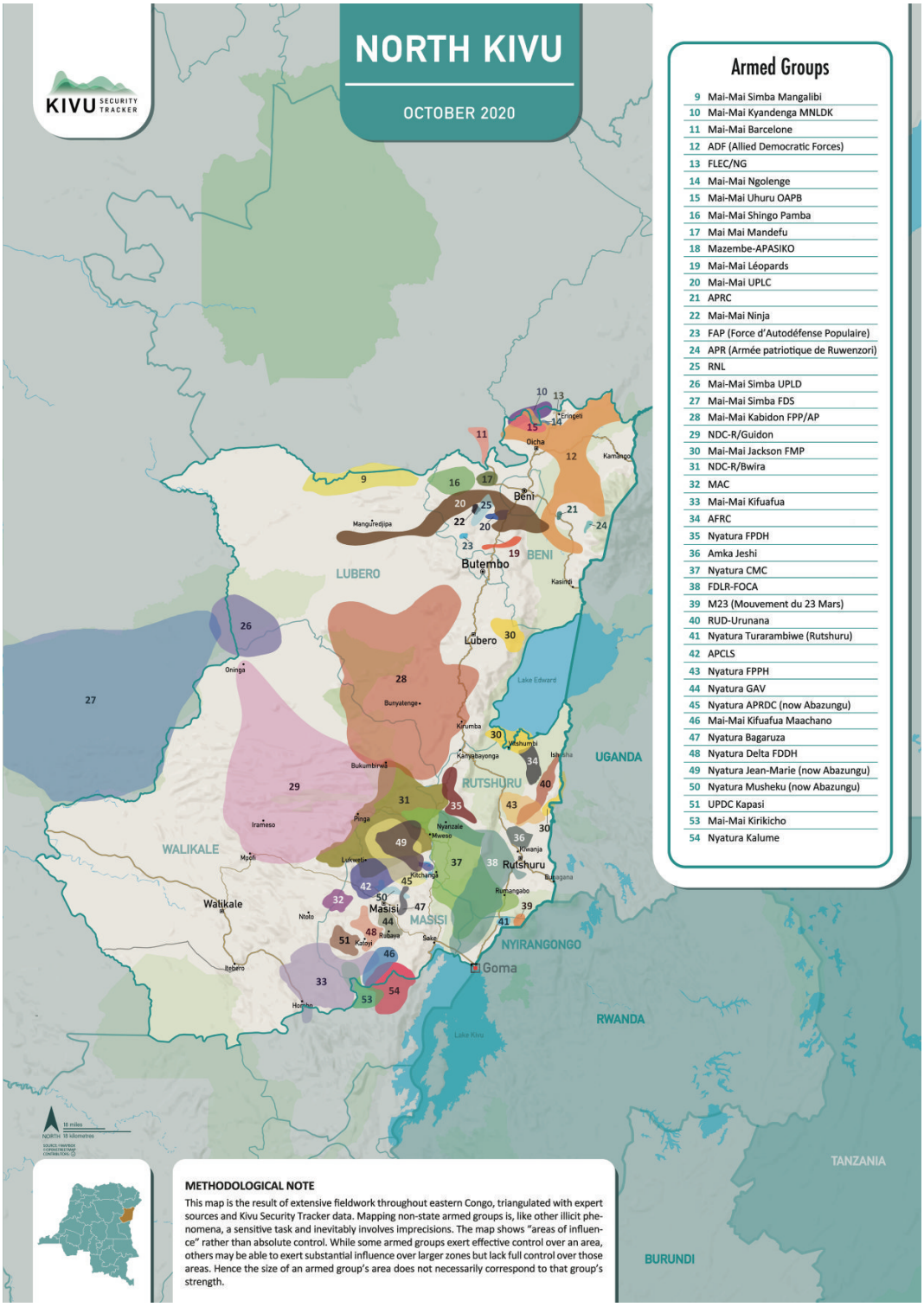
where J is an indicator taking value 1 if the individual enlists and zero otherwise. Enlisting switches on two terms: the benefit to *self* and to the District. The former bundles the economic benefit (the wage differential $w_1 - w_0$), the (dis)utility from hurting the enemy, and the deontological motivation for enlisting. The latter *isolates* the consequentialist value of enlisting: if it is positive, the person considers enlisting right because of its consequence for the District.

Consider first an individual who believes enlisting is the right choice when it comes at no expense to others. This implies $U_1(0) \geq U_0(0)$, or equivalently, $d + s_C^c \geq 0$. For these individuals, the combined effect of "selfish" (economic, aversion to violence against the enemy, and duty) and consequentialist motivations is strong enough to make enlisting preferable. Now, among these individuals, consider the decision to enlist when enlisting prevents another person from doing so. If they still choose to enlist, then $U_1(0) \geq U_0(1)$, which implies $d \geq 0$: they have a net positive bundle of economic and moral motivations. However, those who switch at this point—allowing another person to enlist instead—satisfy $U_1(0) \geq U_0(1)$ and $U_1(0) \geq U_0(0)$, implying $-s_C^c < d \leq 0$. In other words, they enlist when no trade-off exists but concede as soon as someone else can enlist, indicating their prosocial consequentialist motivation outweighs their "selfish" motivations. Similarly, someone who enlists when no one else is affected, when one person is affected, and when two people are affected must satisfy $U_1(0) \geq U_0(2)$, implying $d \geq s_C^c$. If they switch at two, then $0 < d \leq s_C^c$. Using this logic, we can categorize individuals

by the relative size of their prosocial consequentialist parameter, tracing its distribution down to insignificance.

Appendix B. Appendix Tables and Figures

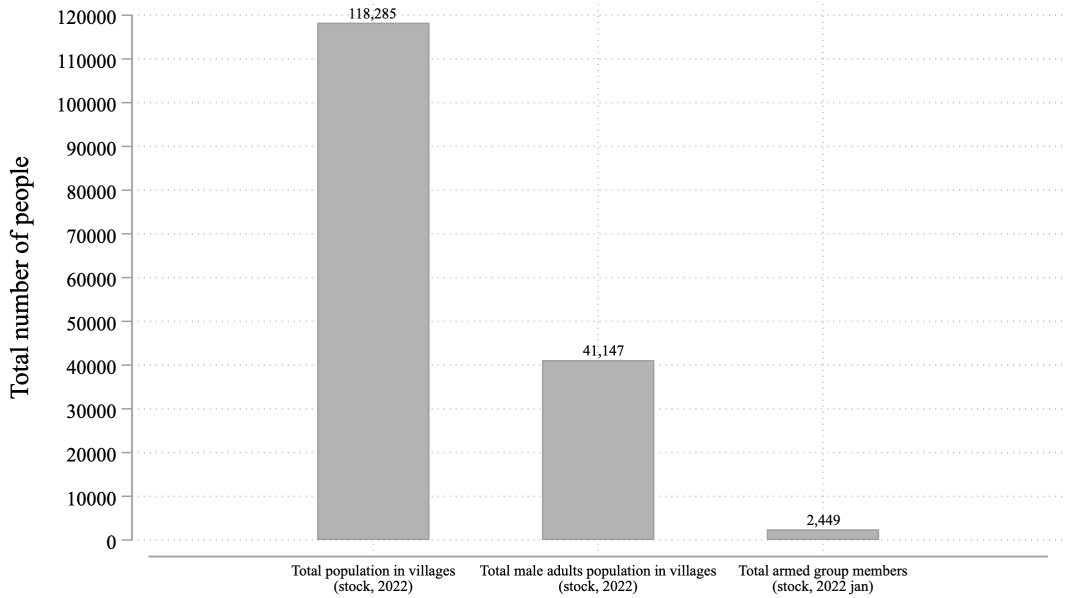
Figure A1: Armed Groups' Territorial Control in 2021



Source: KST (2021).

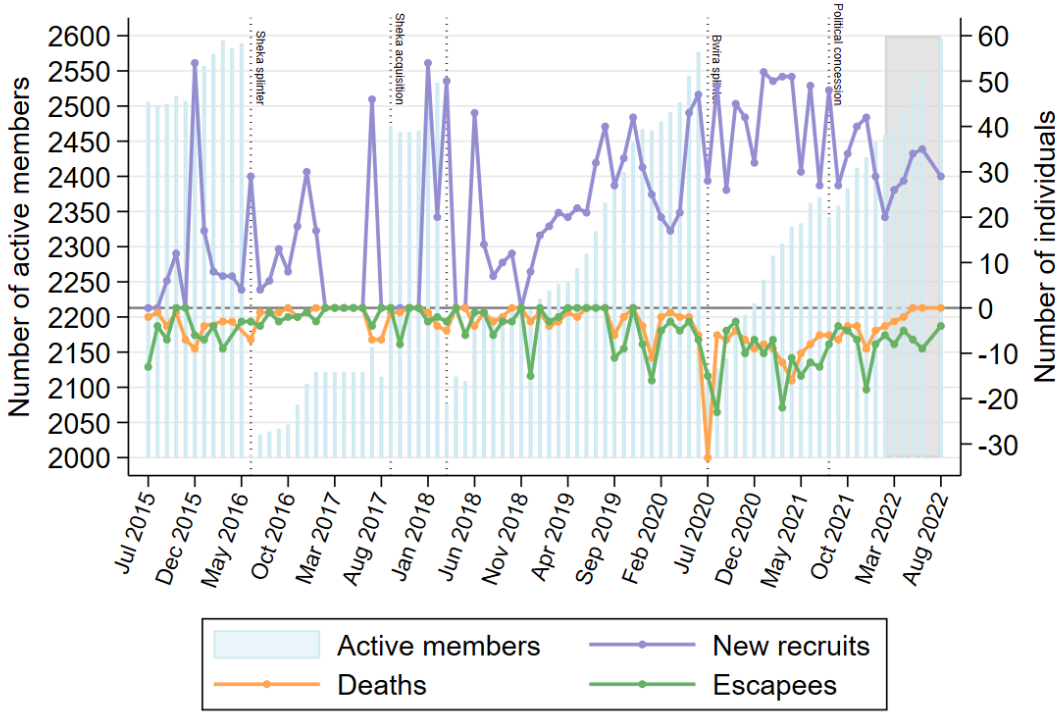
Figure A2: The Group from its Administrative Records

Panel A. Fraction of the Population in the Catchment Area that is a group member



Fraction who are in militia among male adults: 5%
Fraction who are in militia among total population: 2%

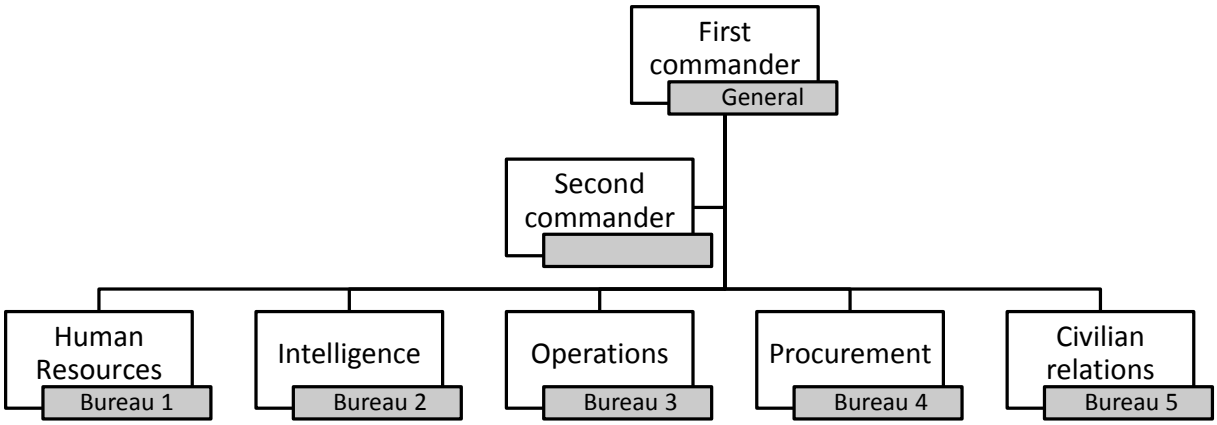
Panel B. Joiners and Stock of Combatants by Month



Number of individuals per month from July 2015 to August 2022

Sources. Panel A: the group's administrative records obtained by the authors and authors' calculations. Panel B: the group's administrative records obtained by the authors, data from survey obtained in this paper among the civilians in the catchment area of the group.

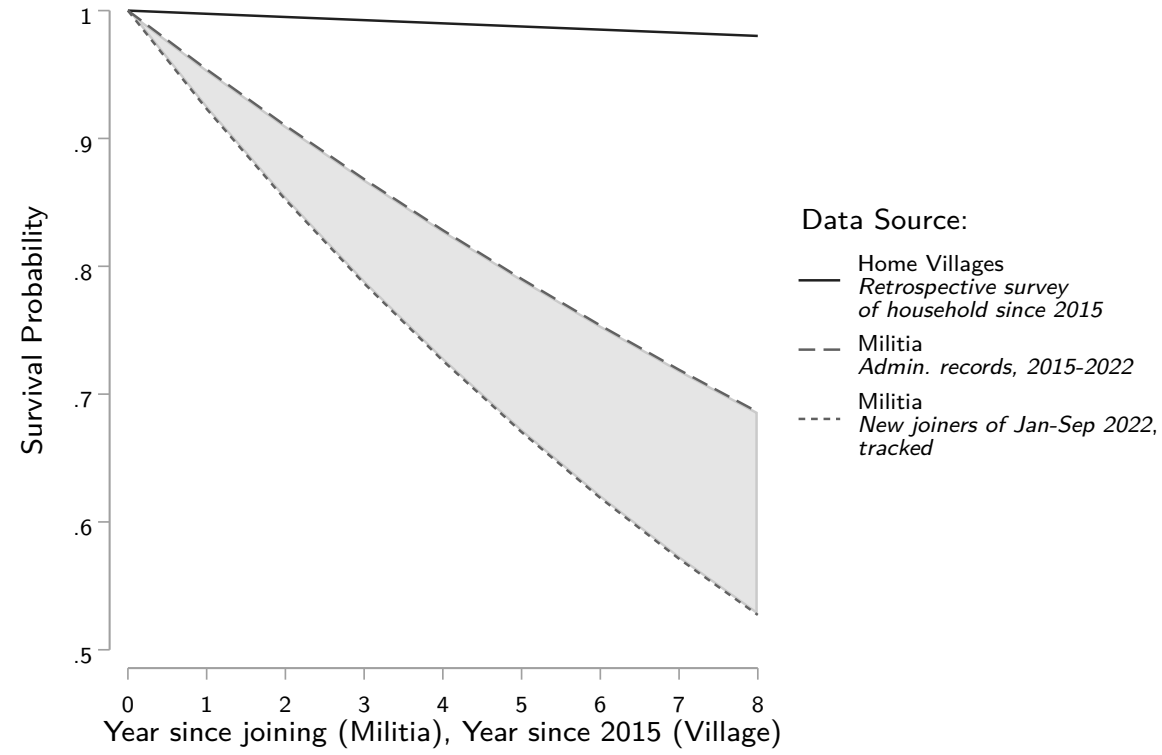
Figure A3: The Group's Functional Organization



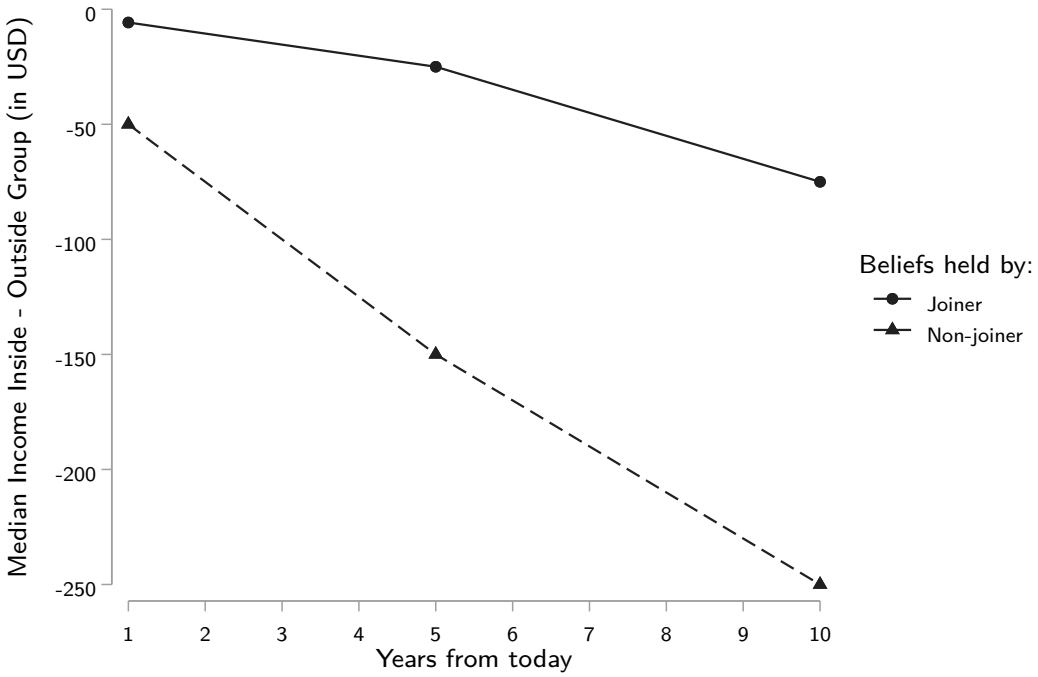
Source: Group's own documents, transcribed using computer by the authors for presentation

Figure A4: Enlisting is Not Privately Economically Profitable

Panel A. Enlisting is Life-Threatening

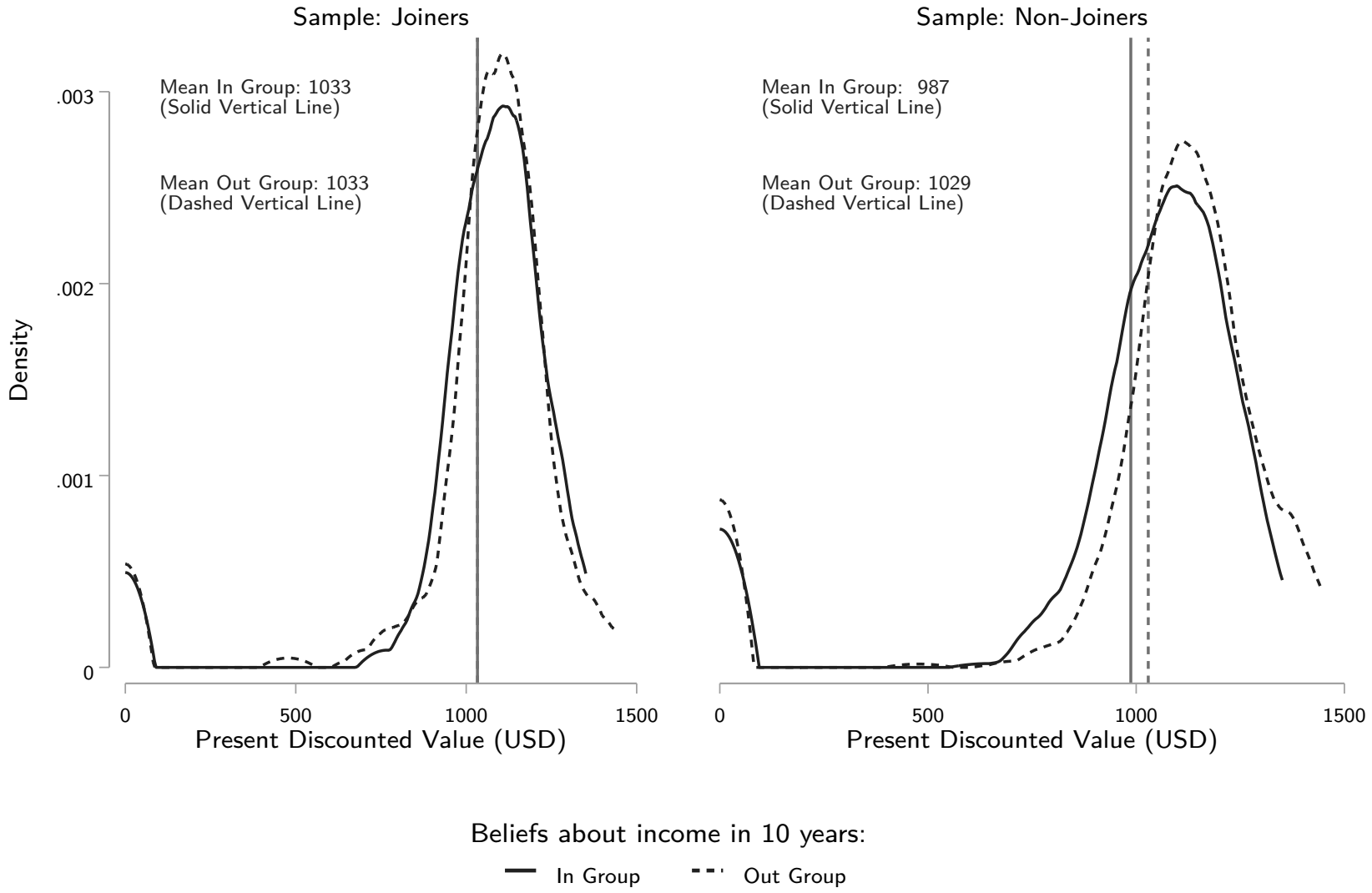


Panel B. Enlisting is not Expected to Pay Off



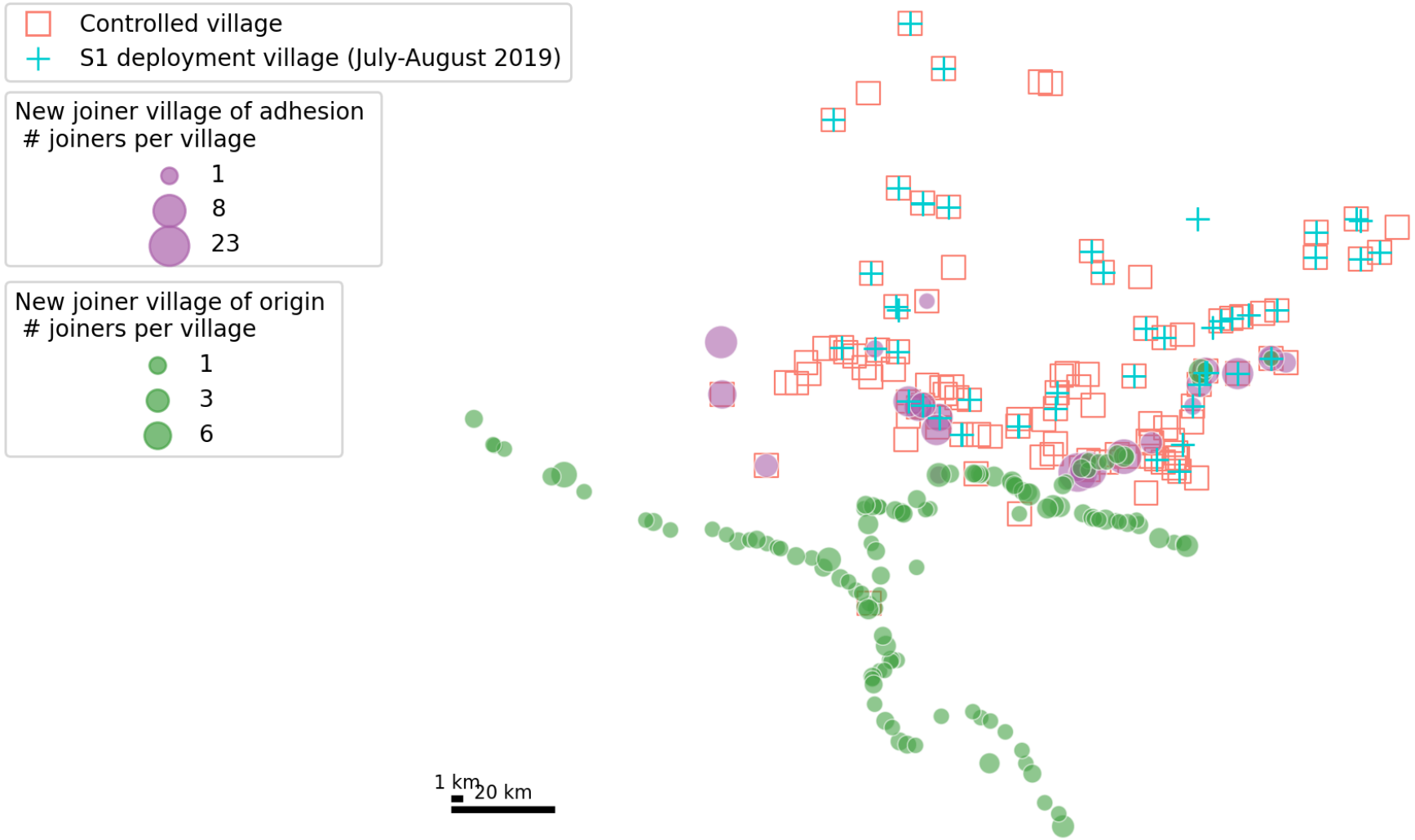
Notes. Panel A presents the prediction derived from an exponential hazard model to estimate the probability to remain alive after 8 years. The source for the data of survival used to estimate the model is listed on the right. The "Home Villages" data source is a survey administered in one randomly selected household in the village of each of 213 new joiners. The survey gathered a household roster including everyone who had lived in the survey since its creation or back to 1990 if it was created earlier, we used the 8 years prior to the survey. The rosters allow us to know who perished and when, in the set of the original household members since its creation. The "Militia, Admin. records" are monthly administrative records of the militia, which we have obtained since 2015. The records indicate, every month, the stock of combatants, the flow of deaths, and the flow of desertions. The "Militia, New joiners" is a survey conducted among the universe of those who enlisted between January and September 2022, for each joiner, eight months after they had enlisted. The survey contains information of whether the joiner was still alive eight months after enlisting. Note that all but one deaths were due to illness. Panel B, shows the median beliefs held by civilians and by joiners alike, the median expected income ten years after enlisting. The beliefs were gathered in a survey in one randomly selected household in the village of each of 213 new joiners, in one randomly selected young male in the household of each joiner, as well as among each joiner before they enlisted. The question asked about their monthly income expectations in 1, 5, 10 years from now: a. if they enlisted the group; b. if they stayed outside the group. Naturally, the counterfactual corresponding to each respondent was presented as a counterfactual "if you had..."

Figure A5: Enlisting is not Expected to Pay Off



Notes: This figure shows the smoothed kernel distribution of the present discounted value (calculated by the authors) of future expected real income and future expected counterfacutal income (provided by the respondents), separately as provided by the joiners and by the pooled sample of non joiners. These beliefs about future real and counterfactual income were gathered in a survey in one randomly selected household in the village of each of 213 new joiners, in one randomly selected young male in the household of each joiner, as well as among each joiner before they joined the group. The question asked about their monthly income expectations in 1, 5, 10 years from now: a. if they joined the group; b. if they stayed outside the group. Naturally, the counterfactual corresponding to each respondent was presented as a counterfactual "if you had..."

Figure A6: Locations of New Joiners and their Villages of Origin



Notes. This figure presents the map of the location of new joiners adhesion into the group (purple dots) against the villages of origin of the new joiners (green dots). The GPS information of the villages of origin and of the villages of adhesion was obtained at the time of the interview conducted in response to the S1 call. These locations are mapped against the backdrop of the locations controlled by the group in 2022 (orange squares) as well as the location of the S1s (blue crosses), which was obtained during our S1 onboarding survey in July 2019. The GPS information of controlled villages is missing for at least one third of the controlled villages, thus the map under-represents territorial control by the group. The source for the controlled villages are the administrative records of the group and various gps gathering exercises in the field conducted to map out the armed group's area of control.

Figure A7: Timing of Data Collection

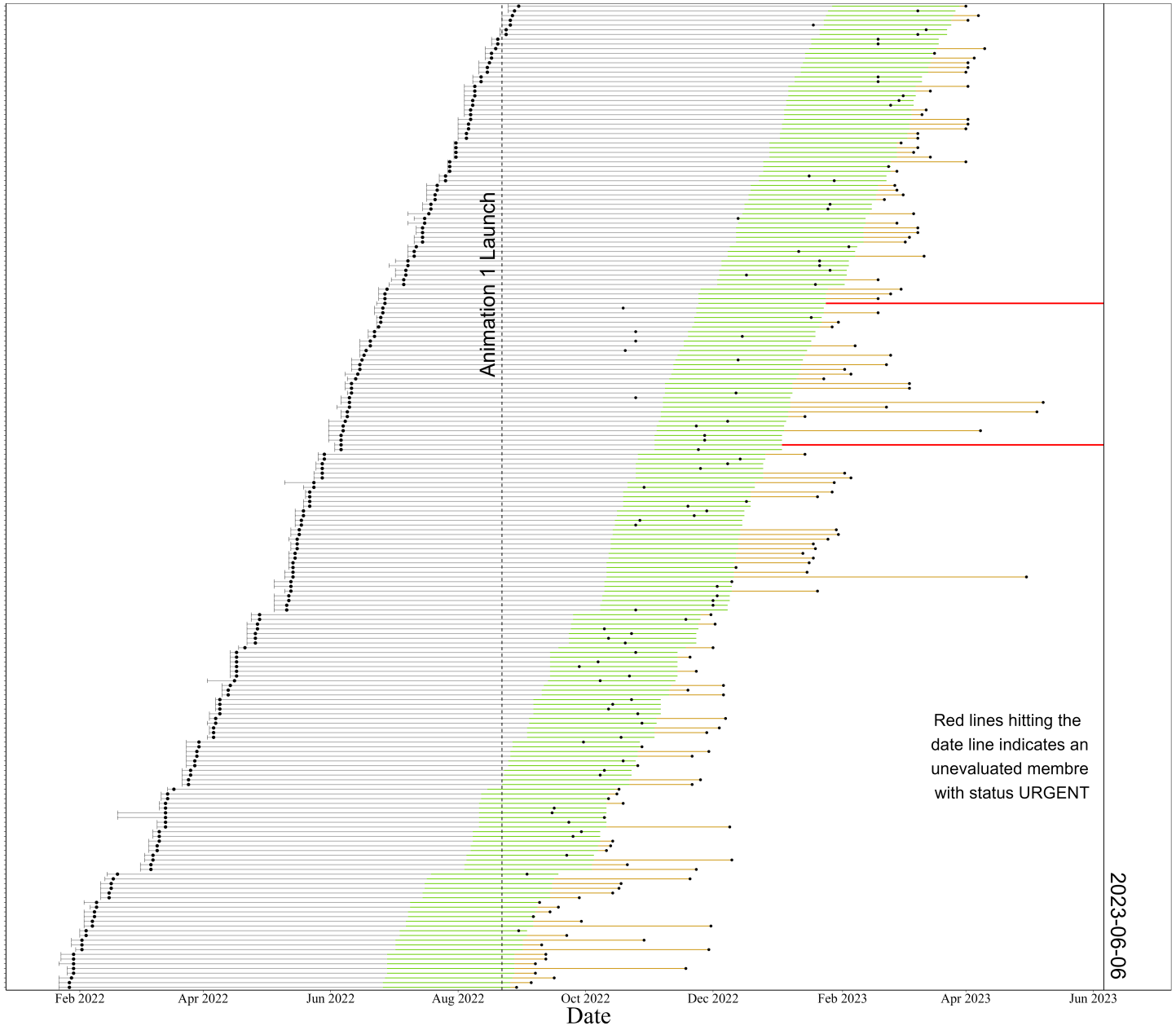
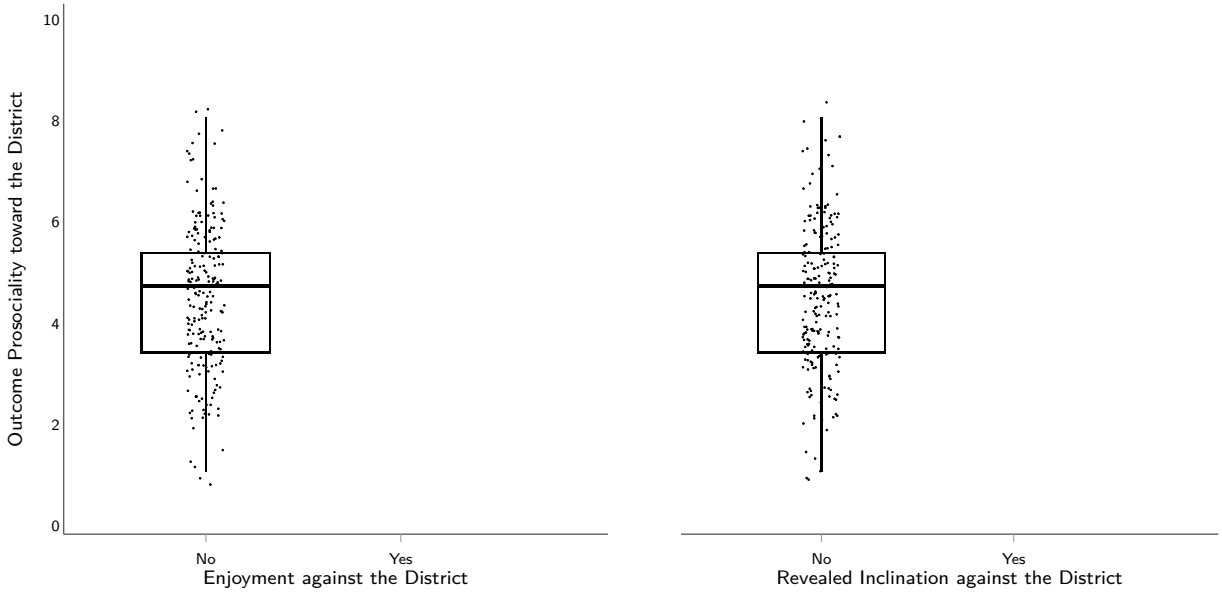
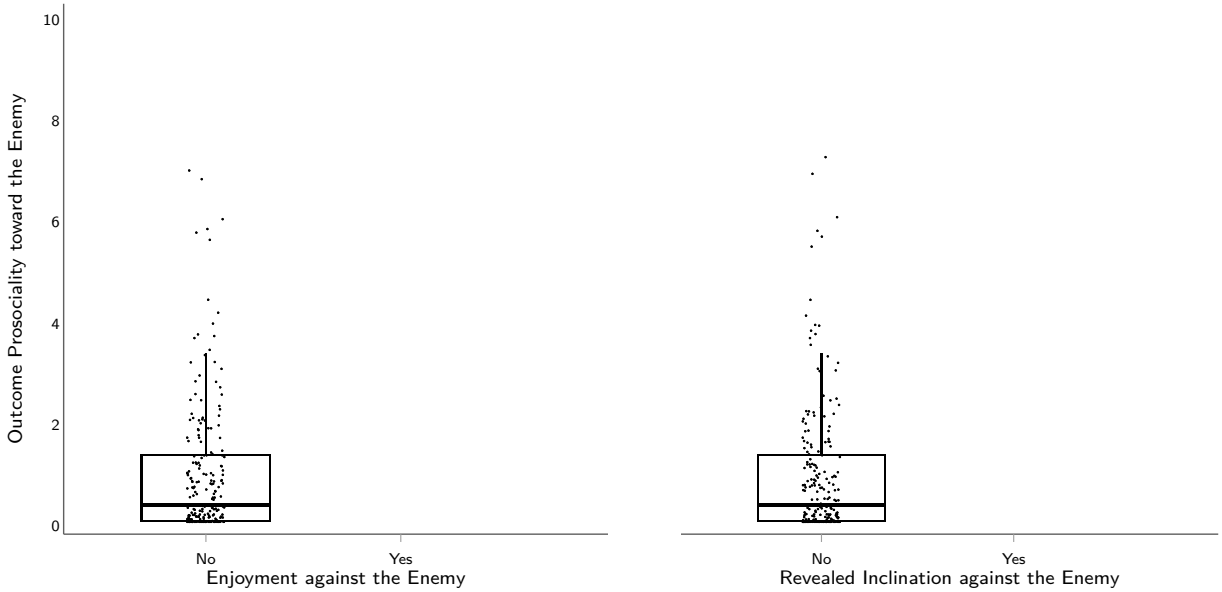


Figure A9: Outcome Preferences and Violence *Inclination*

Panel A. Toward District Members



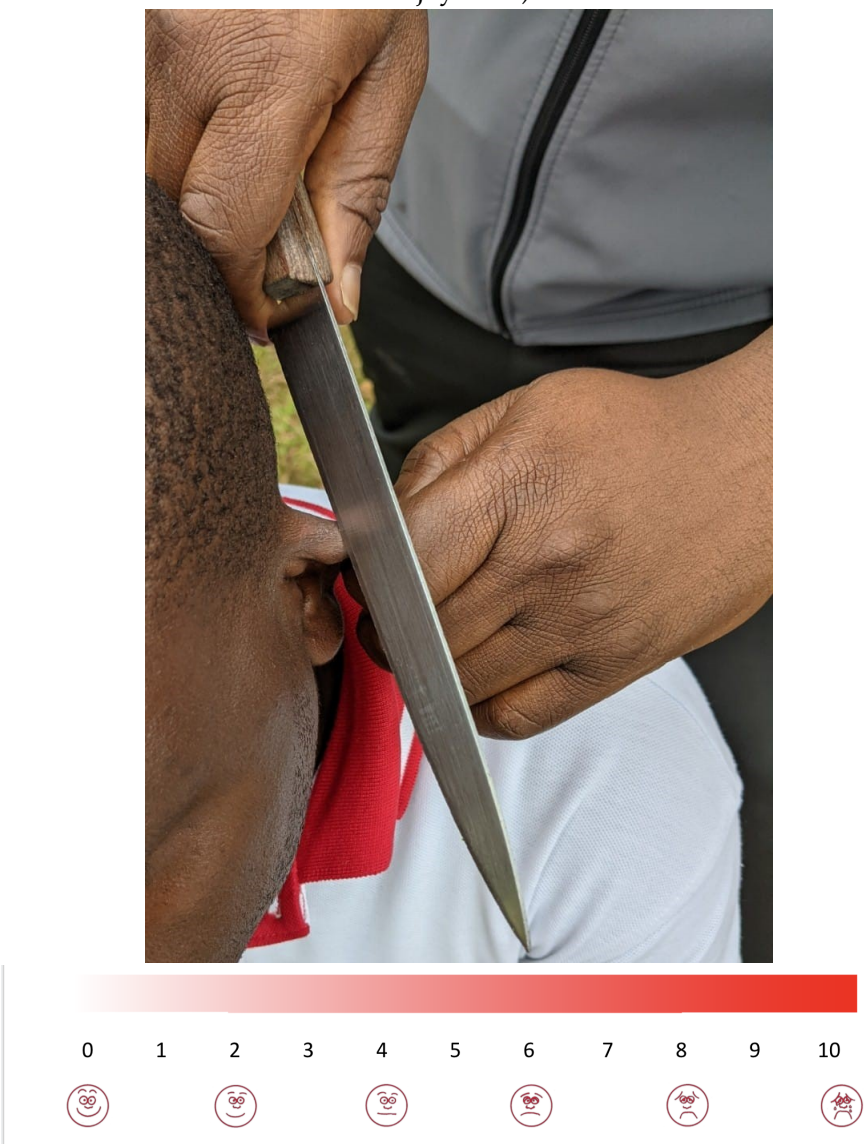
Panel B. Toward Enemy Members



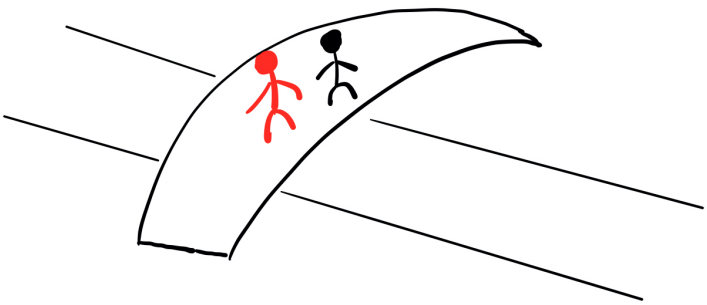
Notes. This figure presents the heatmap of social preferences (an index of declared altruism, hypothetical donation, and dictator game donation, *joy-of-destruction* destroyed amount) toward target of identity X and violence preferences toward a person of identity X, where X is member of the District in Panel A and enemy member in Panel B.

Figure A8: The Survey Tools: Violence Preferences

Panel A. The *Suffering Empathy Task* (Aversion, Beliefs about others' pain and suffering, enjoyment)



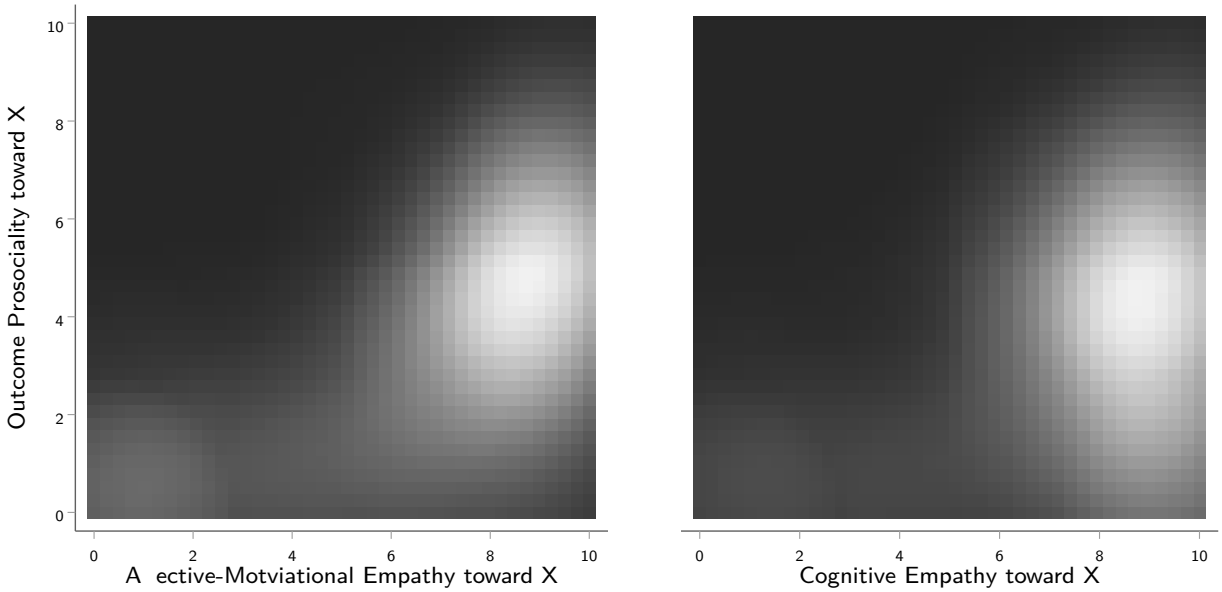
Panel B. Video Game (Revealed Inclination)



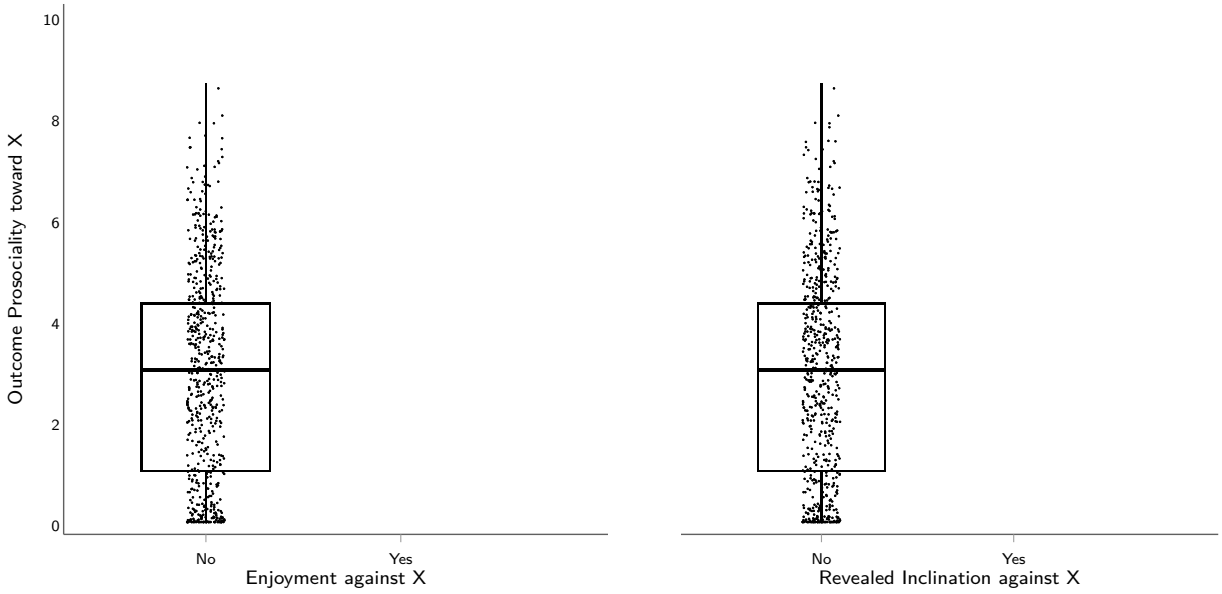
Notes: In Panel A, the photo was staged by members of the team. Panel B shows the basic scenario of the video game, in which animations are added with movements of the subject on the right, alongside with visible consequences and sounds.

Figure A10: Outcome Preferences and Violence Preferences for Identity X

Panel A. Outcome Preferences and Violence *Aversion*

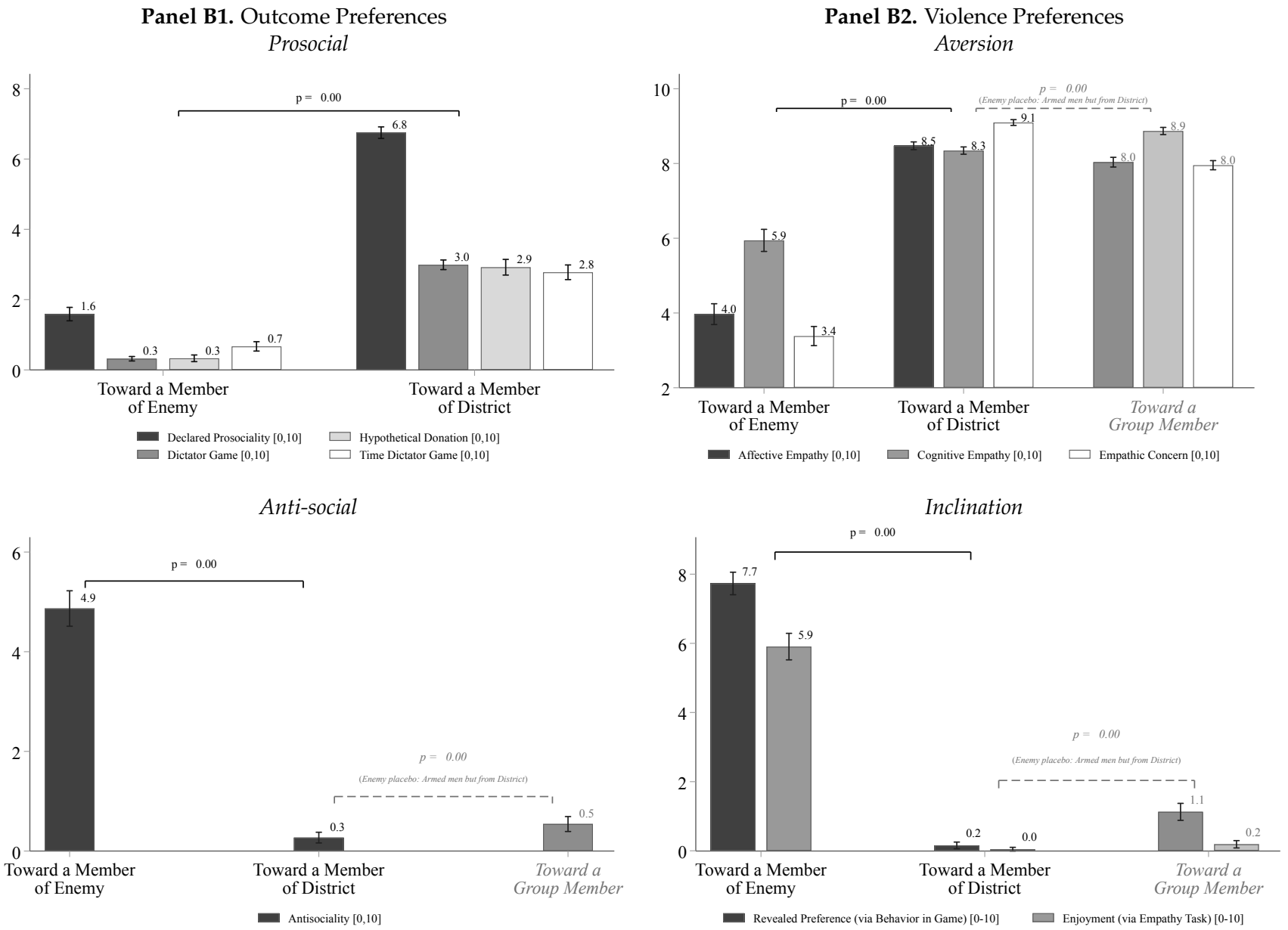


Panel B. Outcome Preferences and Violence *Inclination*



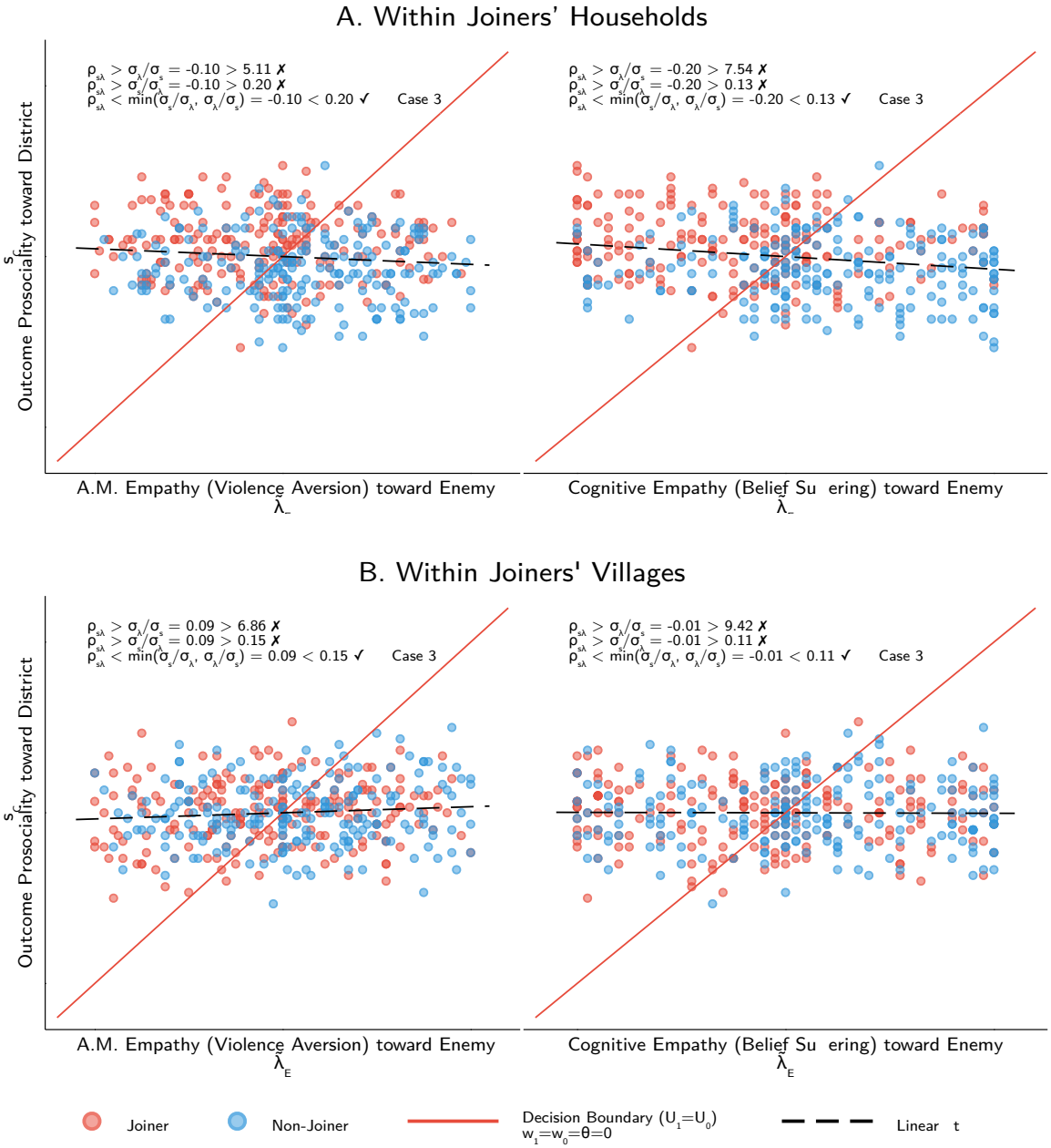
Notes. This figure presents the heatmap of outcome preferences toward a person of identity X on the y-axis and violence preferences (Panel A: aversion, Panel B: inclination) toward a person of identity X on the x-axis, where $X \in \{$a member of the District; an enemy member$)$. Social preferences toward a person of identity X is the average of declared altruism [0-10], hypothetical donation [0-1000] (rescaled to [0-10]), and dictator game donation [0-10] toward a target of identity X. Violence aversion toward a person of identity X is the raw data of cognitive empathy [0-10] (left panels), affective empathy [0-10] (middle panels), empathic concern [0-10] (right panels) toward a person of identity X. Lighter color indicates a larger concentration of observations; darker color indicates a lower concentration of observations.

Figure A11: Social Preferences Unbundled—Including Toward Group Members



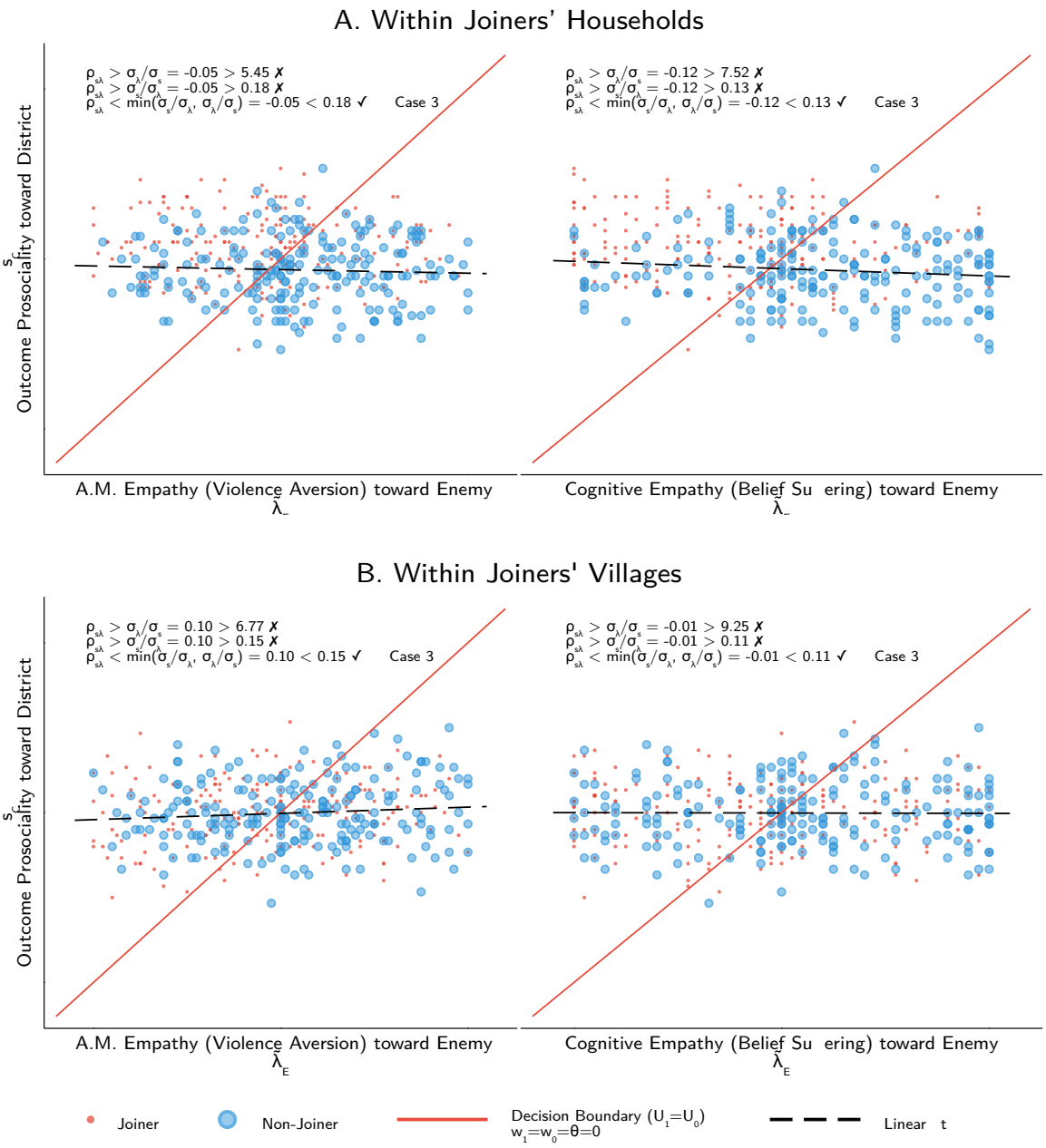
Notes. This figure presents the raw data for the key measures of social preferences (Panel B1), which includes pro- and anti-social preferences, of violence preferences (Panel B2), which violence aversion and violence inclination. In each panel, the measures are presented for when the target is a member of the District, an enemy member, or a member of the District's armed group. All measures are presented in a scale of integers from 0 to 10.

Figure A12: The *Key Relationship*: Outcome Preferences toward the District and Violence Aversion Toward the Enemy



Notes. This figure presents the scatter plots of *residualized* social preferences toward a member of the District on the y-axis on *residualized* violence aversion toward an enemy member on the x-axis. The residuals are obtained in a regression of the outcomes on joiner ID fixed effects. For each joiner, the randomly selected young male household member and the randomly selected young male villager have the same ID. The residuals therefore allow to estimate the relationship for a snapshot of the population at the time of enlisting for each individual joiner, and average for each of the snapshots. This shields from biases in the relationship that might arise from selection of joiners. Social preferences toward a member of the District is the average of declared altruism [0-10], hypothetical donation [0-1000] (rescaled to [0-10]), and dictator game donation [0-10] toward a member of the District and thus amounts to the prosocial component of social preferences (anti-social preferences are only measured in the joy of destruction game, which yields the same qualitative interpretation as prosocial preferences). Violence aversion toward an enemy member is the raw data of cognitive empathy [0-10] (left panels), affective empathy [0-10] (middle panels), empathic concern [0-10] (right panels) toward an enemy member. The linear fit is estimated through OLS.

Figure A13: The *Key Relationship*—With Sampling Weights



The calculations in this figure are conducted weighting each observation by the inverse of their sampling probability.

Figure A14: The *Key Relationship*, Using Violence *Inclination* Toward the Enemy

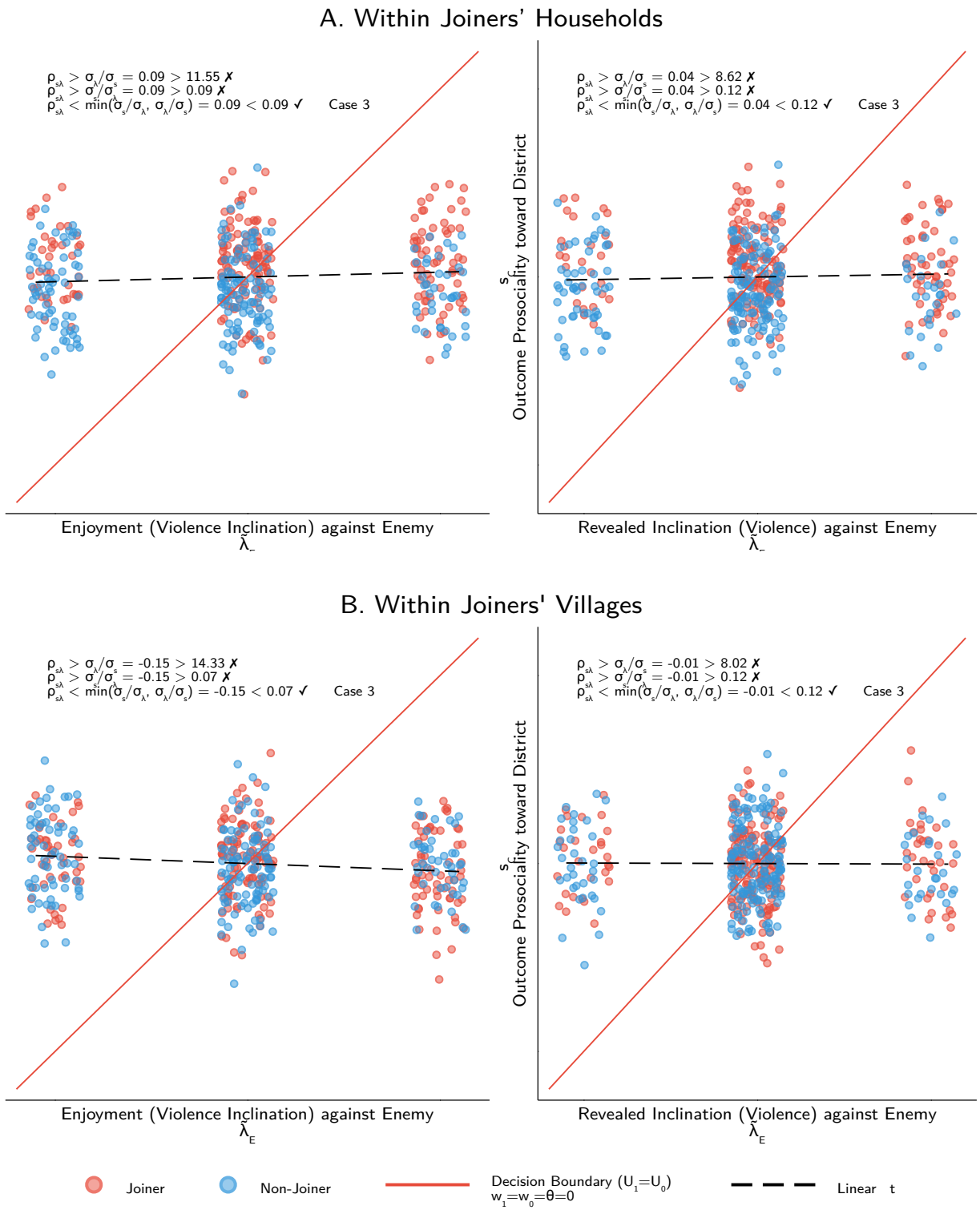
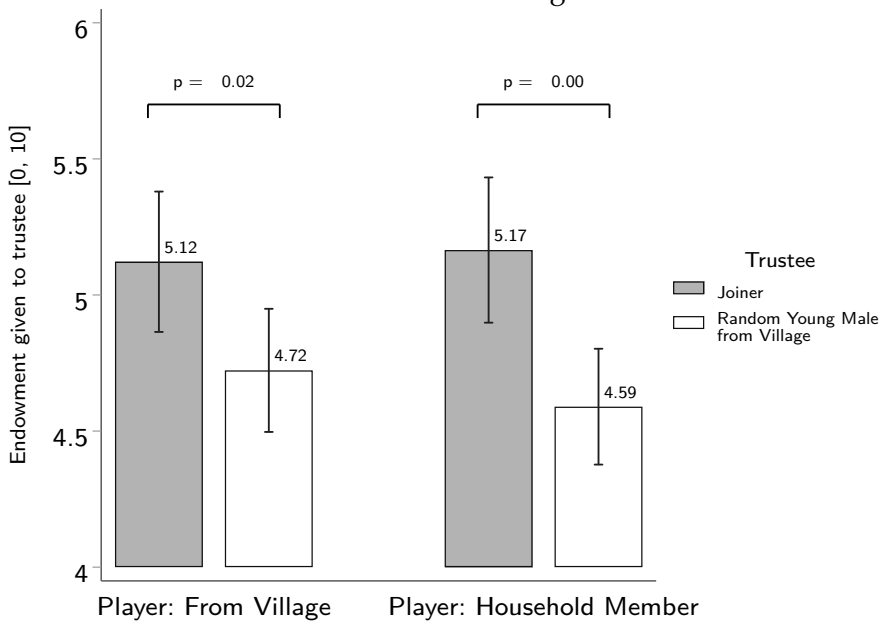
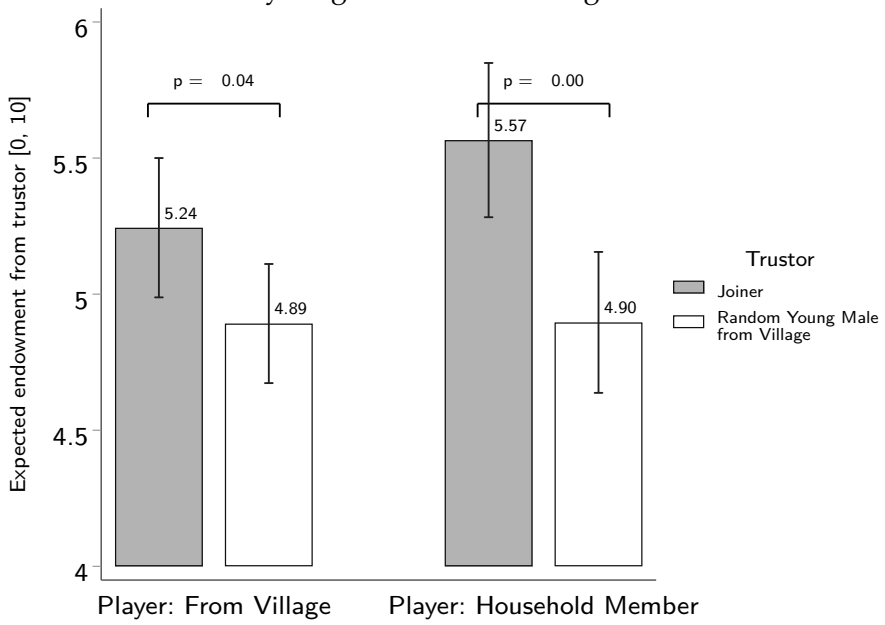


Figure A15: The Joiners are More Trusted, and Expected to be more Trusting

Panel A. Amount Transferred to [name of joiner] vs. Amount Transferred to "random young male" of the Village

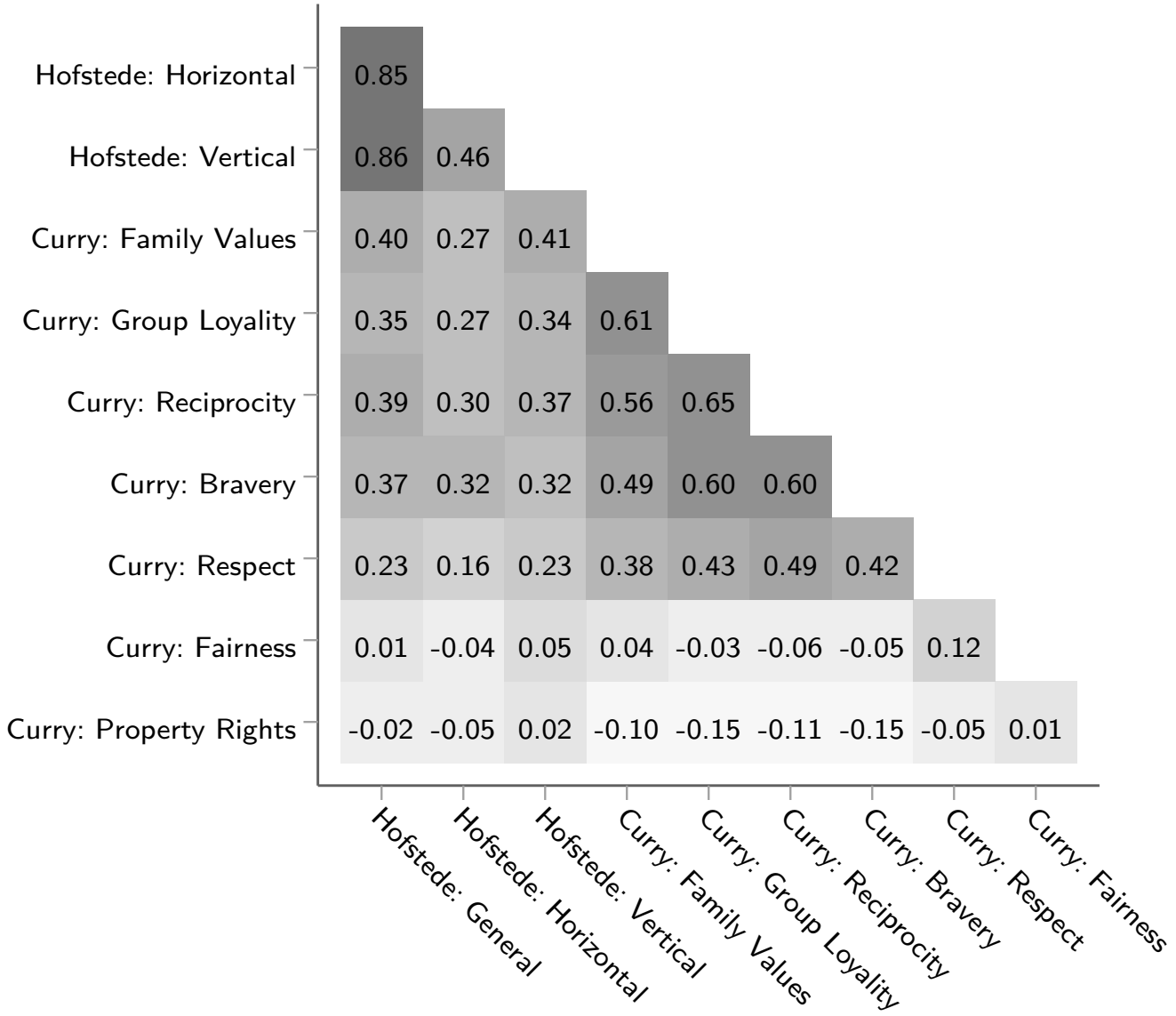


Panel B. Amount that the Player Expects [name of joiner] to Transfer vs. Amount Expects "random young male" of the Village to Transfer



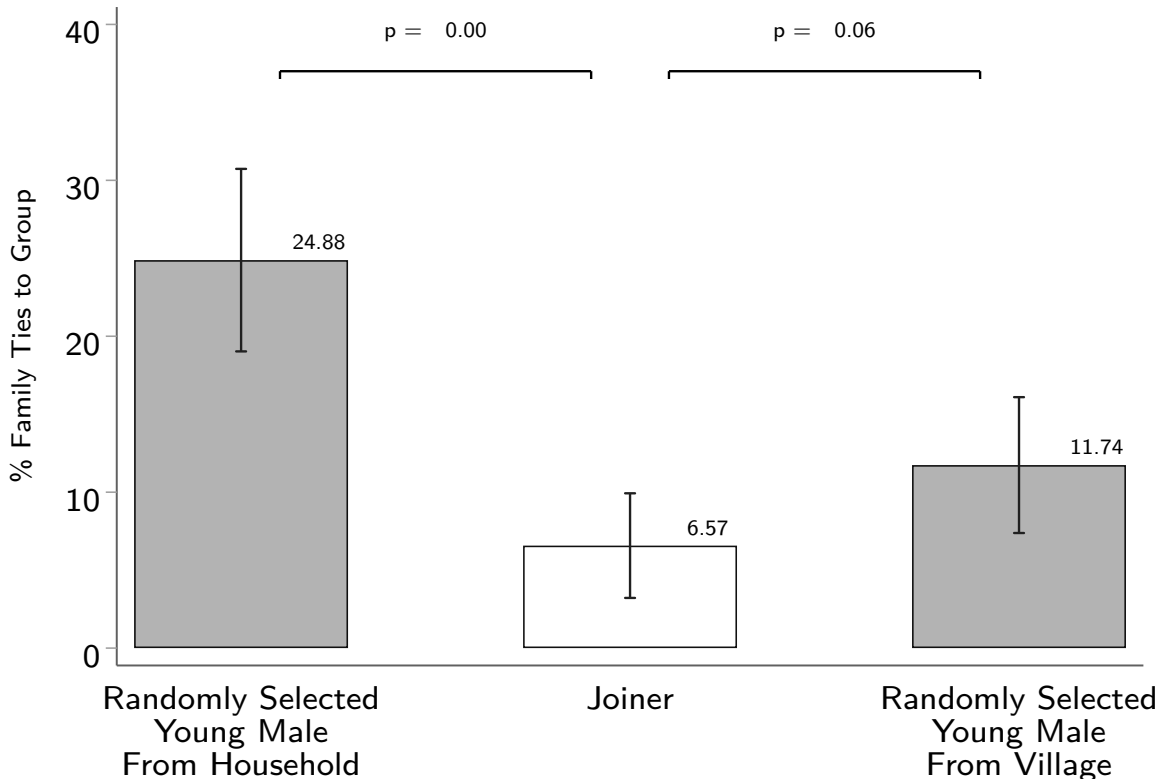
Notes. This figure presents the mean amount transferred by the respondent playing as the first mover in the trust game to the opponent playing as the second mover (Panel A) and the mean amount that the respondent, when playing as the second mover in the trust game, expects the first mover to transfer (Panel B). In both panels, the young randomly selected male from the joiner's household and the young randomly selected male from the randomly selected village of the joiner's household play the trust game. The trust game is played twice: in one round, the opponent is presented as a de-identified "random young male" from the village, and in the other round, it is introduced as [name] where [name] is the name of the actual joiner corresponding to this joiner household and randomly selected village household, which has previously been loaded, identified through the S1 call. In both rounds, the respondent plays first as a first mover (the choices are reported in Panel A), and can transfer any amount from 0 to 10 to the other player, acting as second mover, knowing that any amount transferred is multiplied by three, and that the second mover then can choose whether and how much to transfer back to the first player. The respondent is then asked to play as player 2, and is offered to guess how much they think the opponent will transfer to them. The figure presents the choices given, separately, when the respondent playing is the randomly selected young male of the randomly selected household of the joiner's village (labeled Villager), and when the respondent is the randomly selected young male from the corresponding joiner's household.

Figure A16: Correlation Between Moral Foundations Dimensions



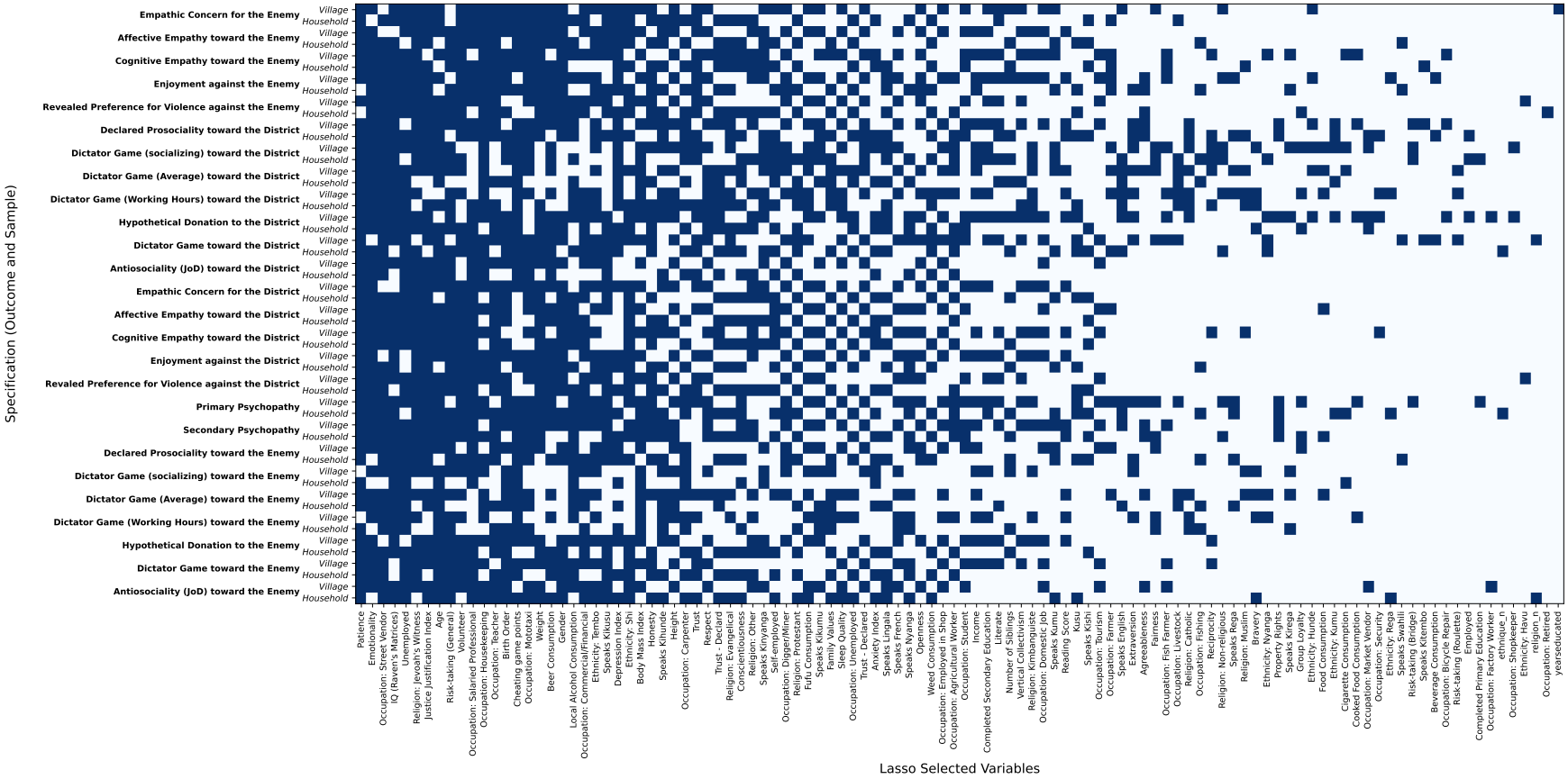
Notes. This figure presents the correlations between the measures of Oliver Scott Curry and Whitehouse (2019)'s moral foundations and Geert Hofstede and Minkov (2010)'s horizontal and vertical collectivism in the whole sample.

Figure A17: Social Ties with Group Members



Notes. This figure presents the raw data for the question "Is this sibling already in an armed group in the District"?

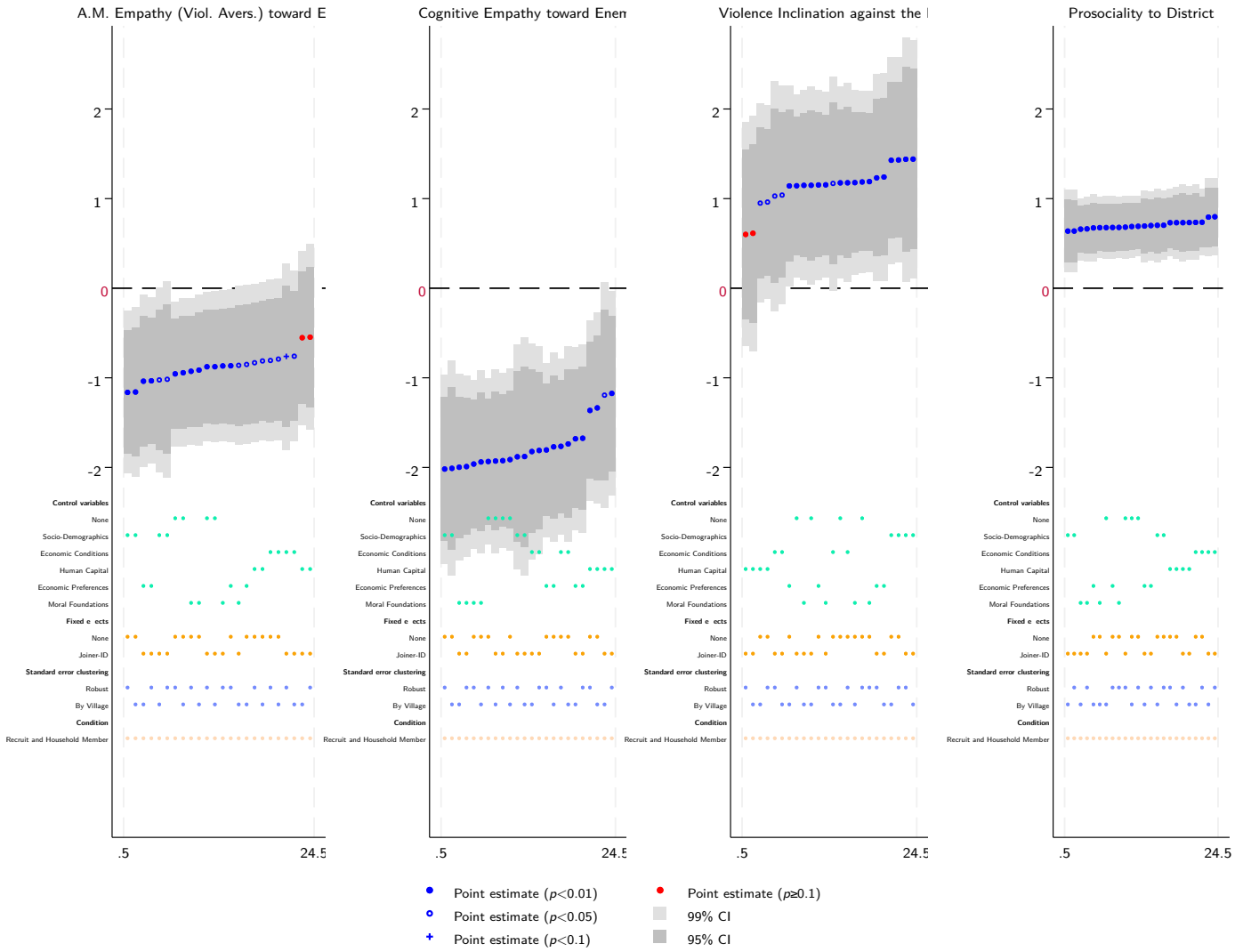
Figure A18: Variables Selected through LASSO Procedure



Notes: To address potential model selection and regularization issues in the estimates of Equation 1 presented in Figure 5, we implement the LASSO (Least Absolute Shrinkage and Selection Operator) double-selection procedure with cross-validation (**?**). This figure presents the variables selected through this procedure for each specification.

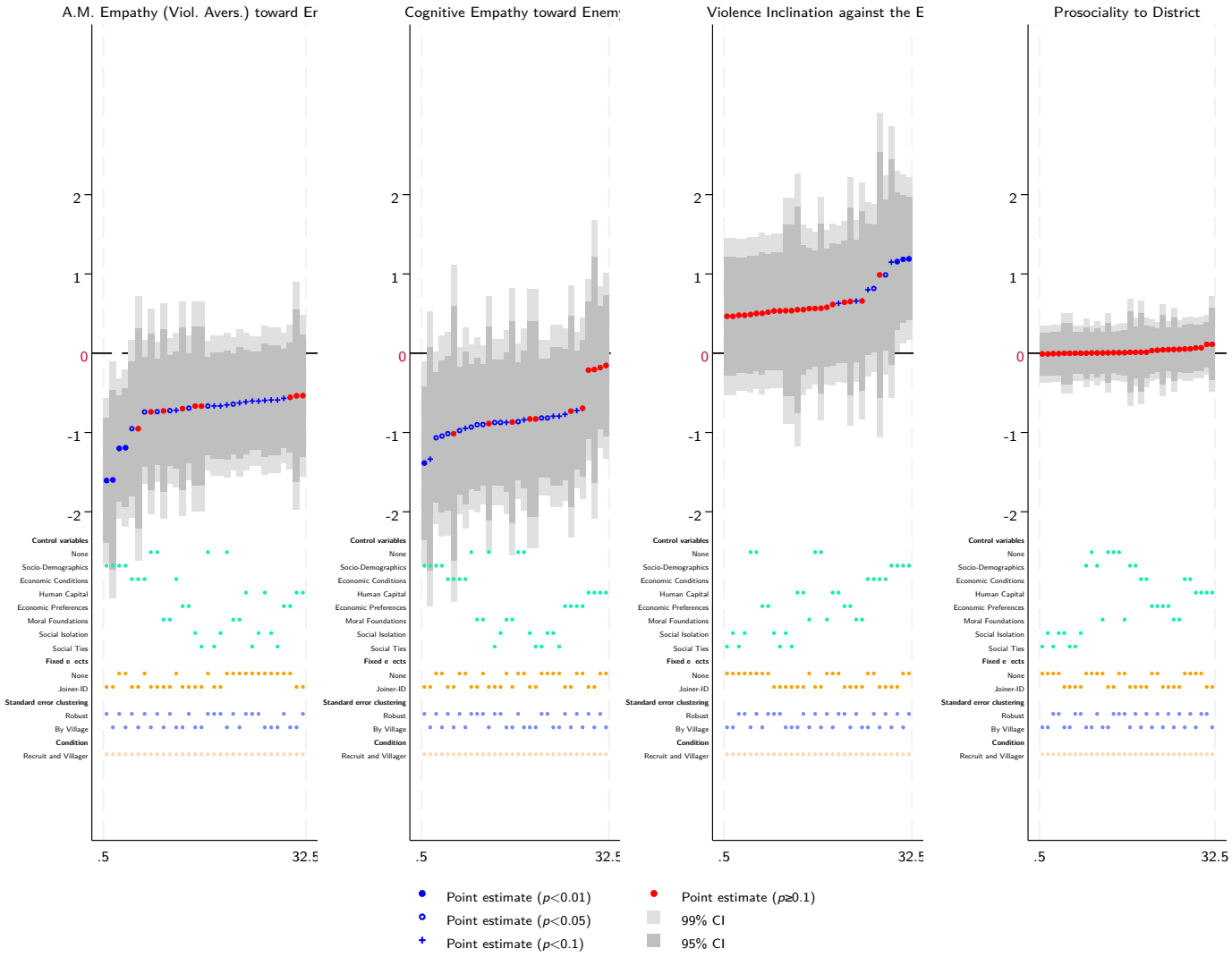
Figure A19: Specification Robustness

Panel A. Within Households



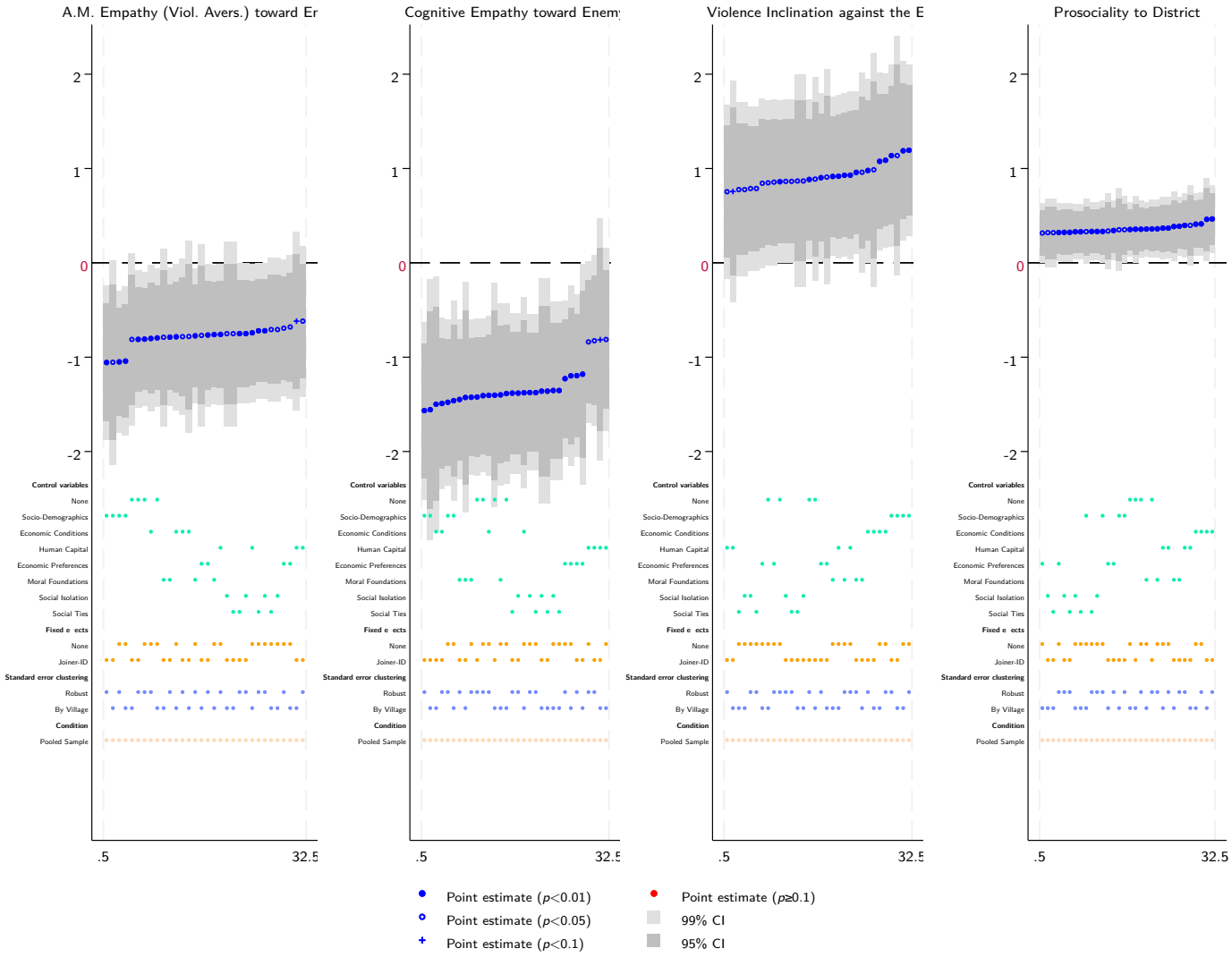
Notes. This figure presents the estimates from all possible specification choices within household.

Panel B. Within Villages



Notes. This figure presents the estimates from all possible specification choices within village.

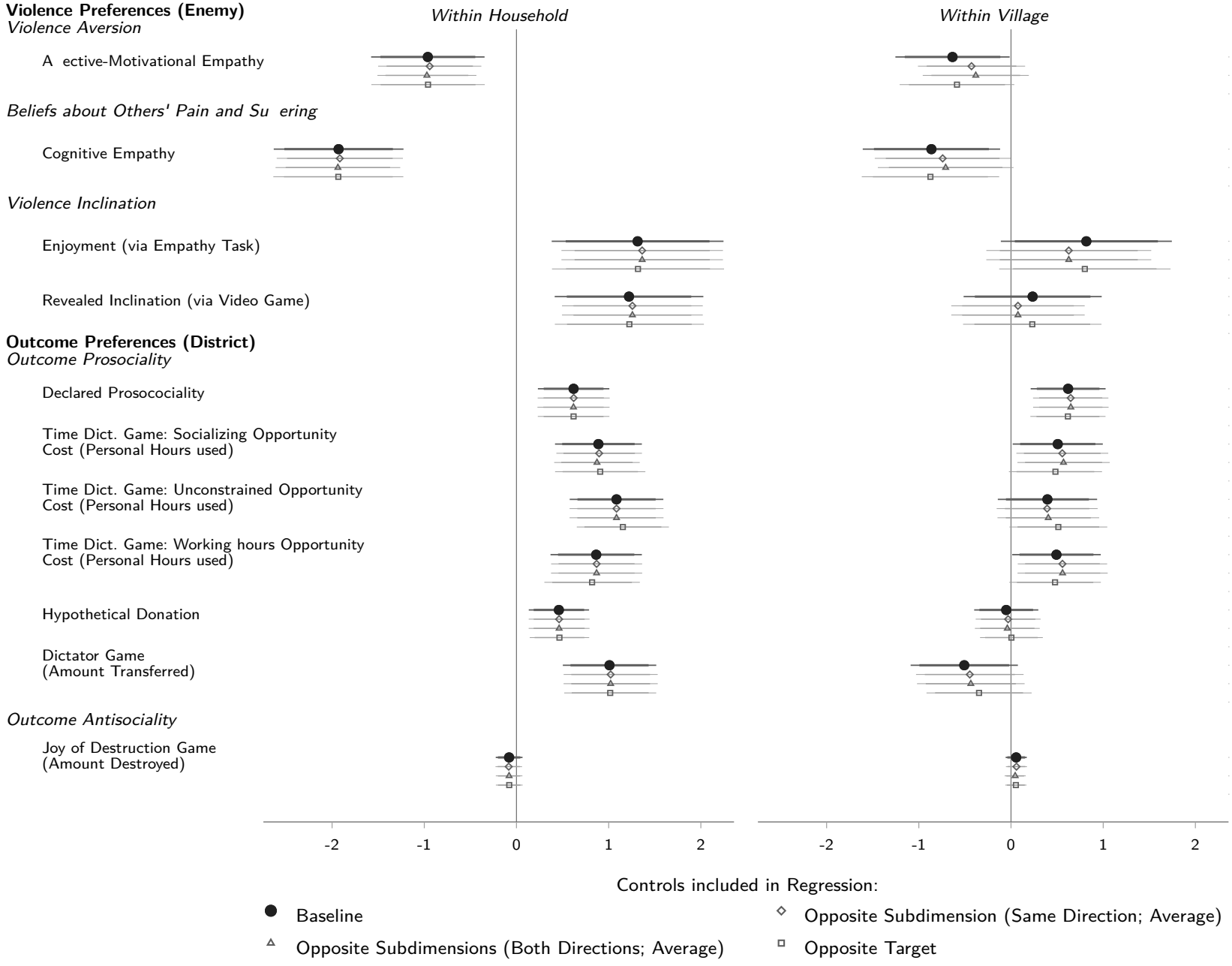
Panel C. Pooled Joiners, Within Households and Within Villages



Notes. This figure presents the estimates from all possible specification choices within village.

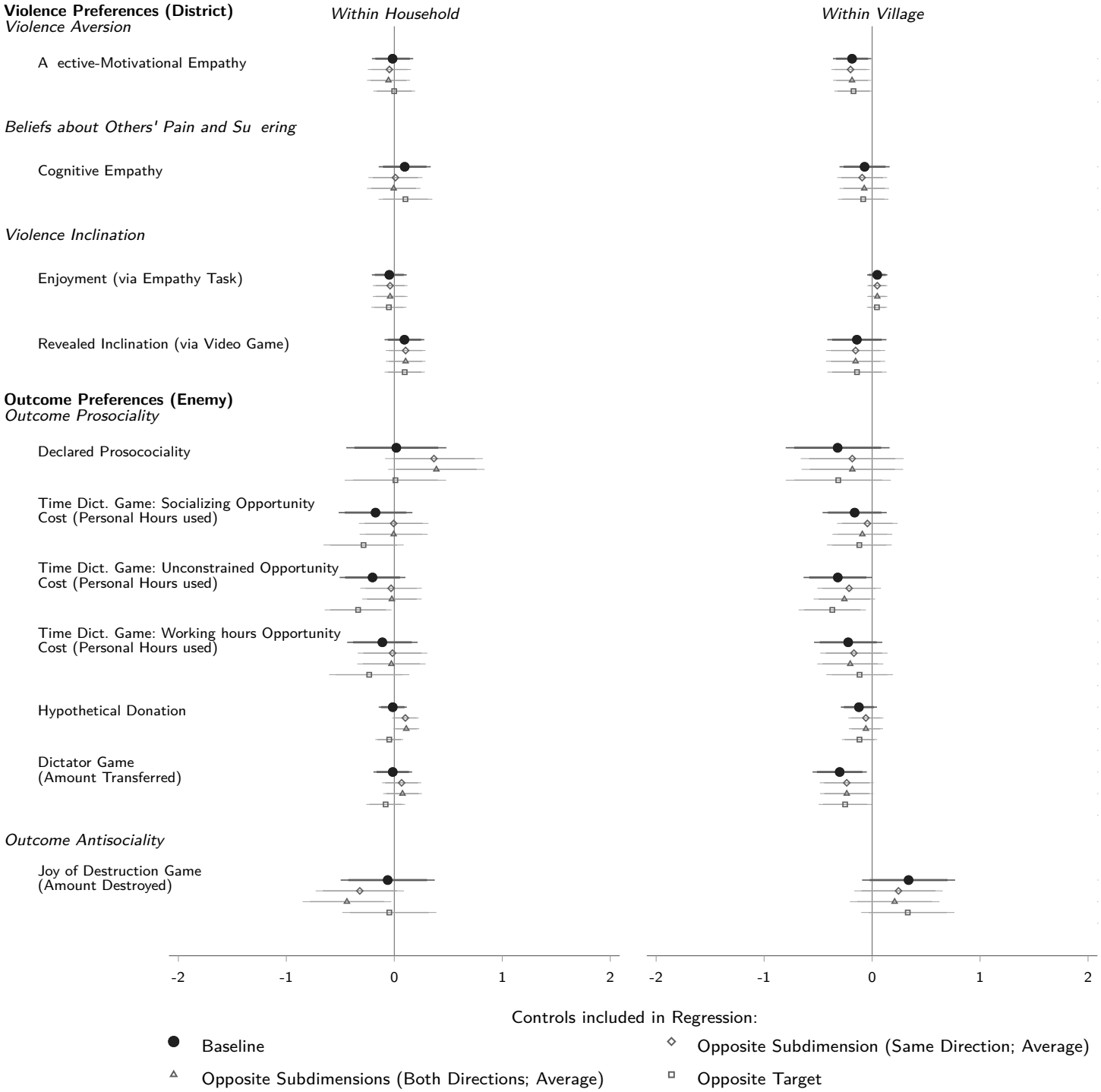
Figure A20: Joiners are the "Zealots"—Controlling for Opposite Dimensions of Social Preferences

Panel A. Joiners are those with Greater Violence Preferences Toward the Enemy



Notes are in Panel B.

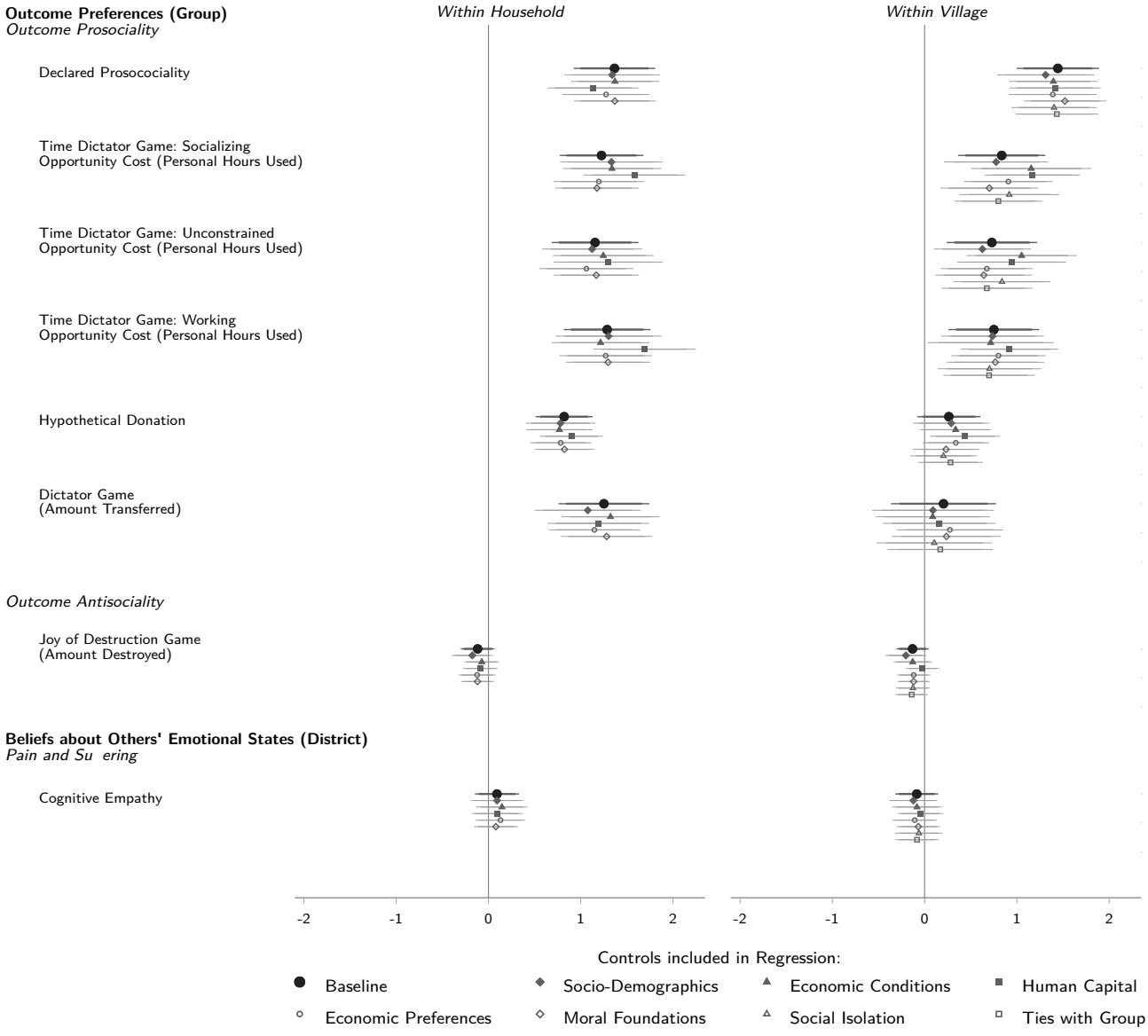
Panel B. Joiners are those with the Largest Social Preferences toward Members of the District



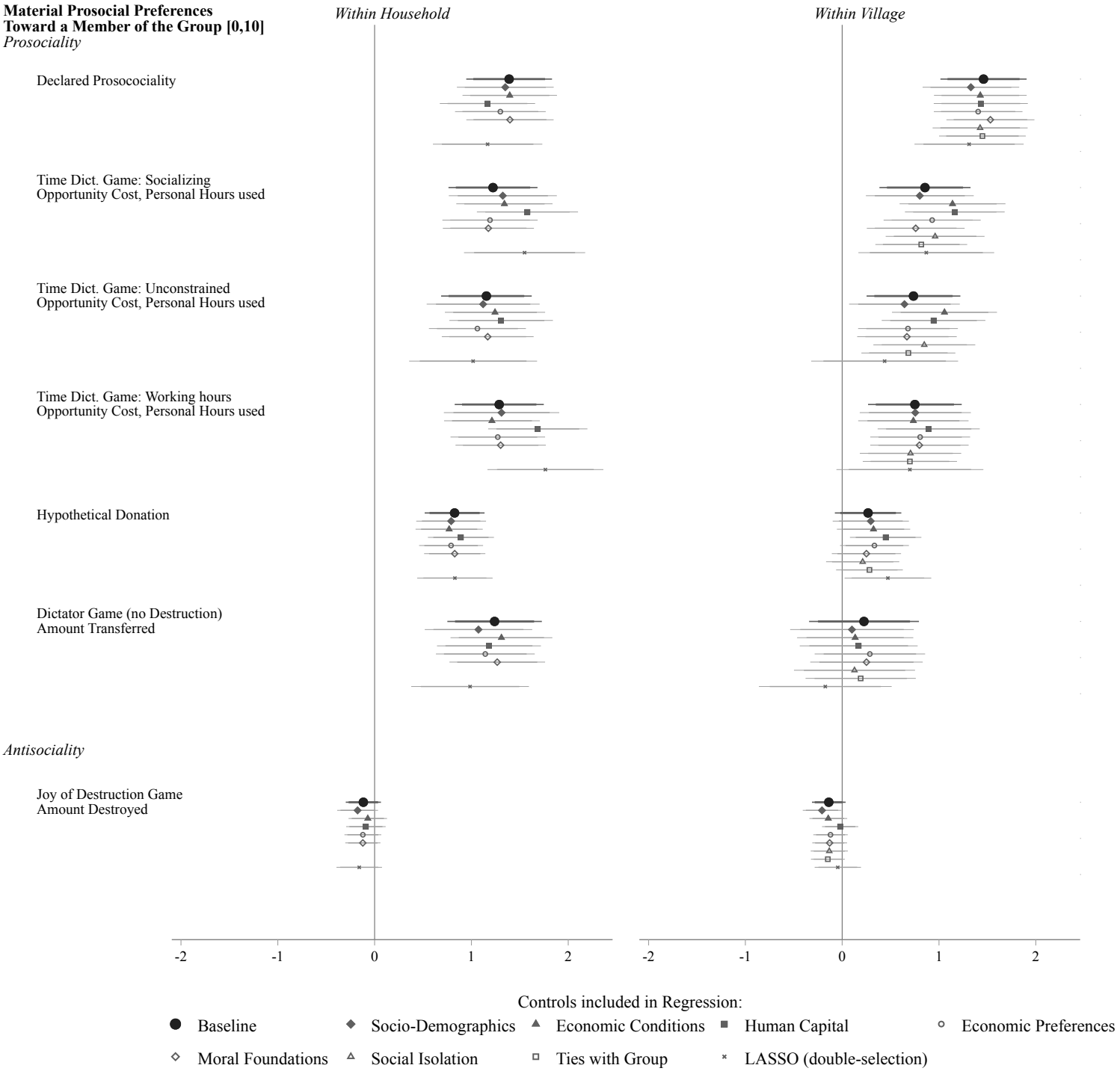
Notes. This figure presents the estimates from the following specification. Let i index the individual. We estimate the following equation in Ordinary Least Squares in the following sample: new joiners, the randomly selected youth in their own household, and the randomly selected youth in the village, $y_i = \alpha + \beta J_i + \epsilon_i$, where y_i is a (unbundled) component of social preferences of individual i presented always on a scale from 0 to 10, and J_i is an indicator taking value one if individual i belongs to the set of new joiners and zero otherwise. Thus, the coefficient β measures the difference between joiners and the average of young adult of their household and the average young adult of the same gender of the average household of the village. In Panel A, the dependent variable are the measures of social preferences toward a member of the District/toward an enemy member; in Panel B, the dependent variables are the measures of violence preferences against enemy members/against members of the District. The coefficients and confidence intervals are presented without additional controls (Baseline) and including controls indicated in the legend. In both panels, we report regression estimates without controls and with the following controls added separately: (i) the opposite social preference component (averaged across all sub-components) directed toward the same target as in the dependent variable; in this case, the coefficients and confidence intervals are qualitatively indistinguishable regardless of what sub-components of the opposite social preference components we include in the average, as well as whether we narrow the controls to *pro*social outcome preferences or also include *anti*-social outcome preferences. And: (ii) the same social preference sub-component as in the dependent variable, but directed toward the opposite identity.

Figure A21: Joiners are the "Zealots:" Falsification Test District's Armed Group Members as Targets

Panel A. Violence Preferences



Panel B. Prosocial Preferences



Notes. This figure replicates Figure 5 using members of the District's group as targets, instead of District members or of enemy members.

Figure A22: The Moral Enlisting Moral Dilemma: Screenshots

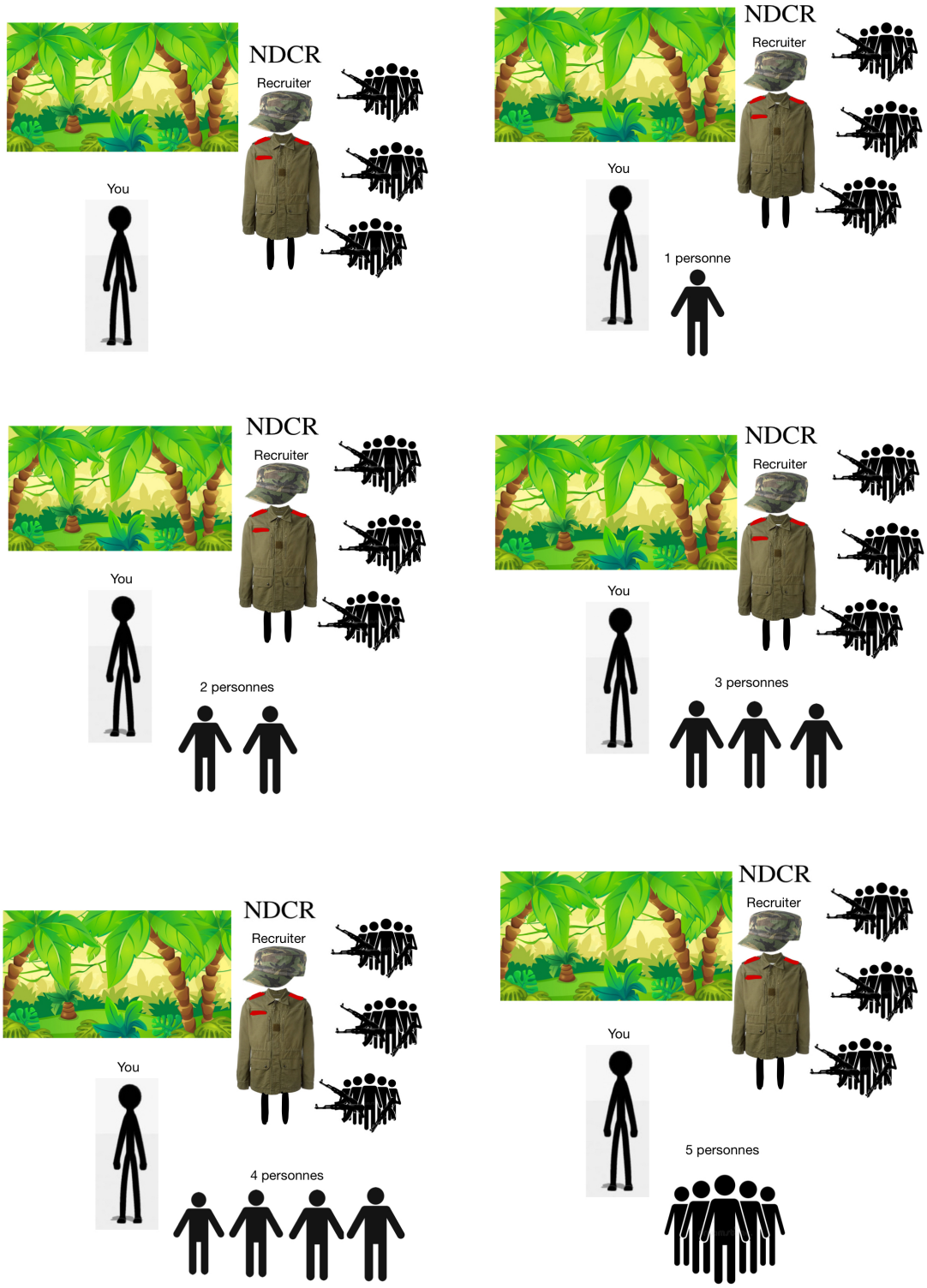
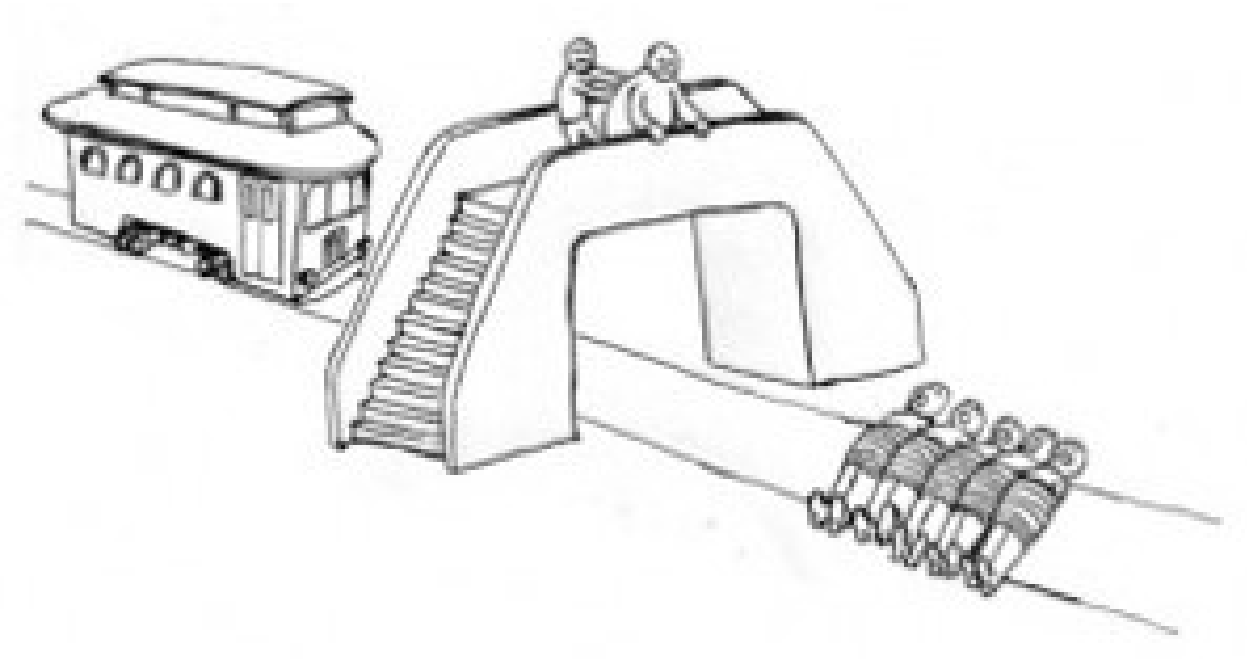


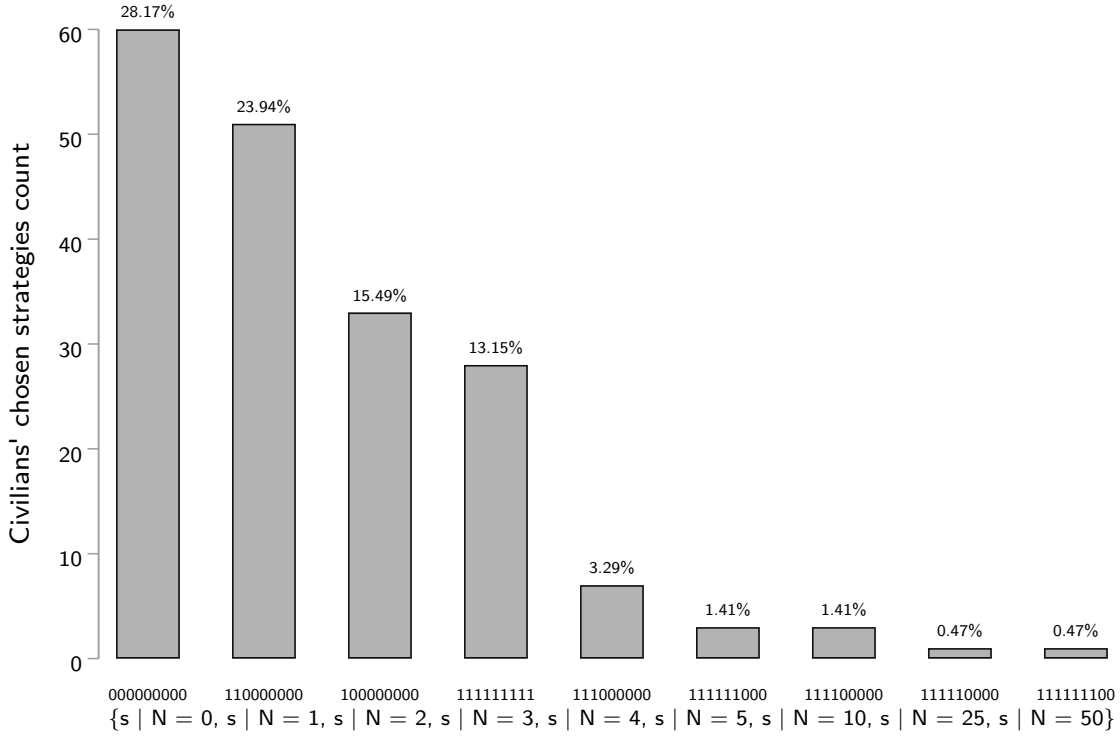
Figure A23: The Morality of Killing enemy members



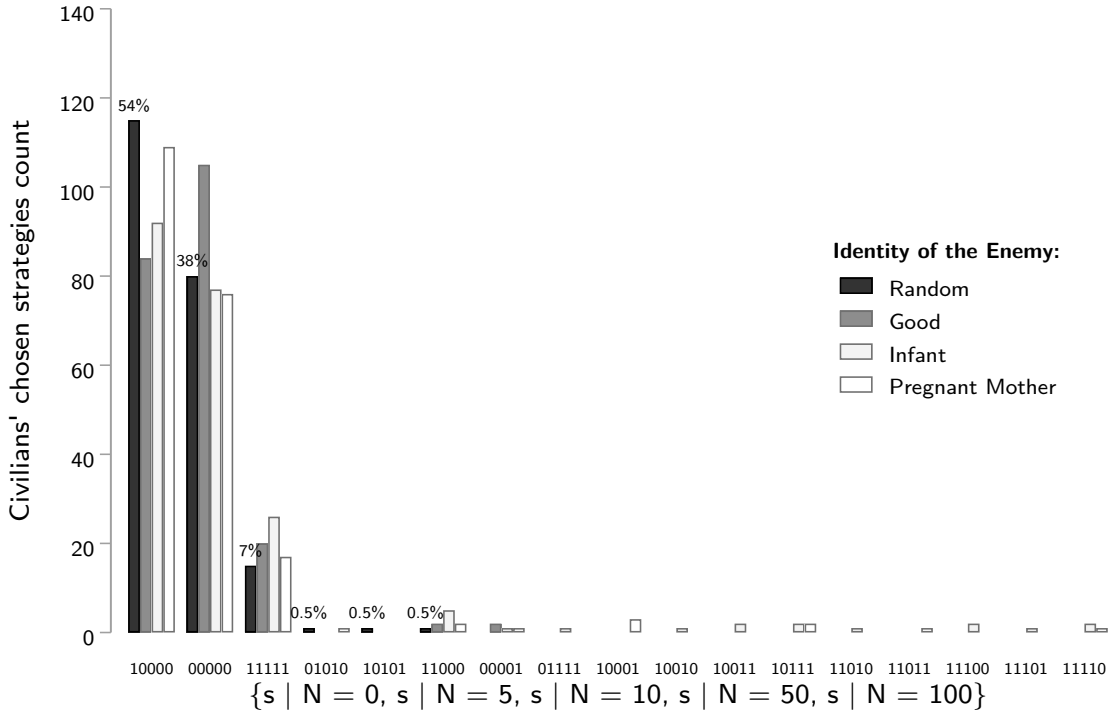
Notes. This figure shows the image displayed for the trolley dilemmas. The choice is elicited through the strategy method, in the following increments for the number of the District's civilians sacrificed: [0; 5; 10; 25; 50; 100] and similarly for the scenarios in which the sacrificed individuals are other enemy members; after the last choice, the respondent is asked to make their strategy explicit. The specific script for this dilemma is: *"You are on this bridge [see the screen]. Under you, there is a tram, going toward a group of people from the District tied to the rails. Next to you, there is an enemy member. Choice A: if you do nothing, the tram stops, because it is not going fast, and none will die; Choice B: if you push the member of the enemy off the bridge, the person dies, but this also creates a short-circuit that results in the death of the members tied to the rails. Regardless of what you would do, what <u>should</u> you do? Regardless of your choice, it is important to remind that your choice would carry no consequence (e.g., legal) and that none, in the scenario or in the real world, will be able to identify that the choice was made by you."*

Figure A24: Moral Dilemmas—Raw Data: Strategies Played

Panel A. Civilians' Hold Enlisting as Morally Right—Strategies

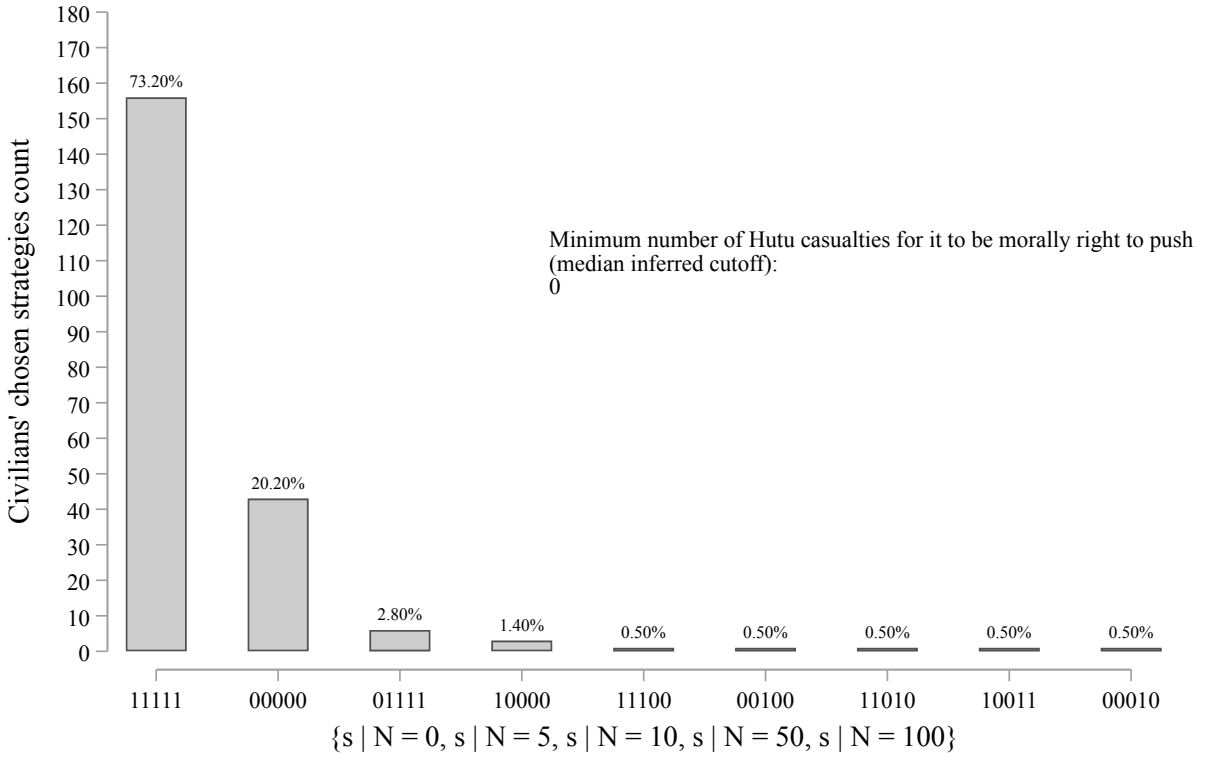


Panel B. Civilians' Hold Killing Enemies as Morally Right—Strategies



Notes. Panel A presents the strategies count for the enlisting dilemma, asking respondents whether they think it is morally right to enlist, when their decision to enlist would imply that $N \in \{0,1,2,3,4,5,10,25,50\}$ other equally qualified members would not be able to enlist. The term $s|N \in \{0,1\}$ denotes the choice of whether to enlist ($s = 1$) or not enlist ($s = 0$) when enlisting implies that N other equally qualified others would not be able to enlist. Panel C presents the strategies count for the killing moral dilemma.

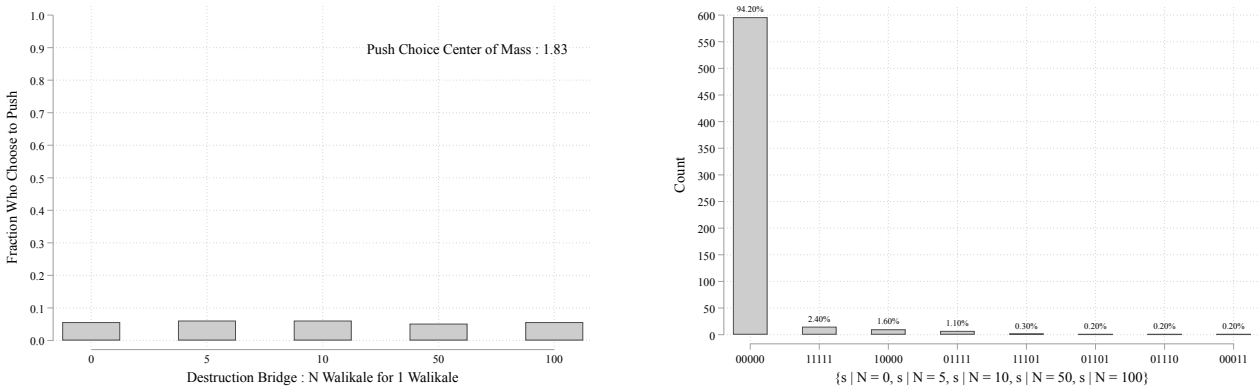
Figure A25: The Prevailing Moral Values: The Killing Dilemma with Enemy Casualties—Strategies Played



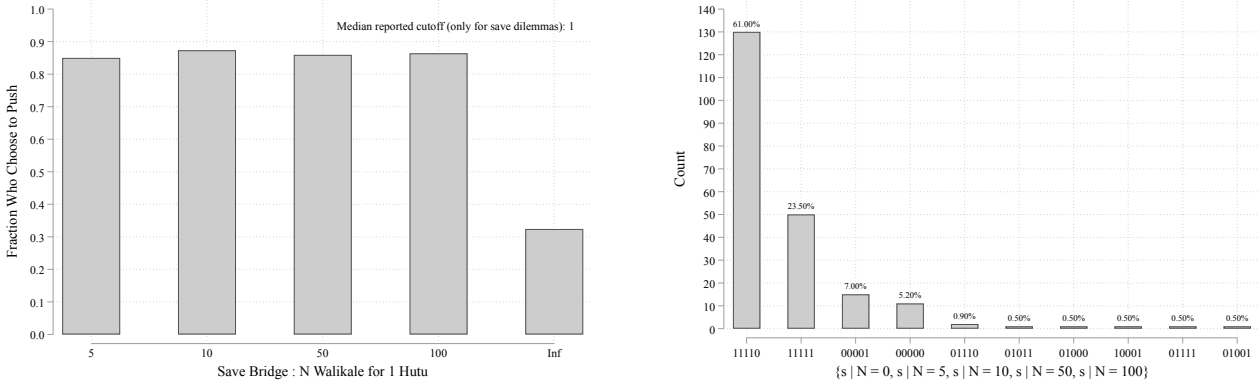
Notes. This figure presents it for the dilemma in which respondents are asked to ascertain whether pushing an enemy member off the bridge is morally right, when it results in the death of $N \in \{0, 5, 10, 50, 100\}$ enemy members. The term $s|N \in \{0,1\}$ denotes the choice of whether to push the member of the enemy off the bridge ($s = 1$) or not enlist ($s = 0$) when pushing implies, in addition to his own death that N enemy members would also die (Panel B).

Figure A26: The Prevailing Moral Values: Robustness

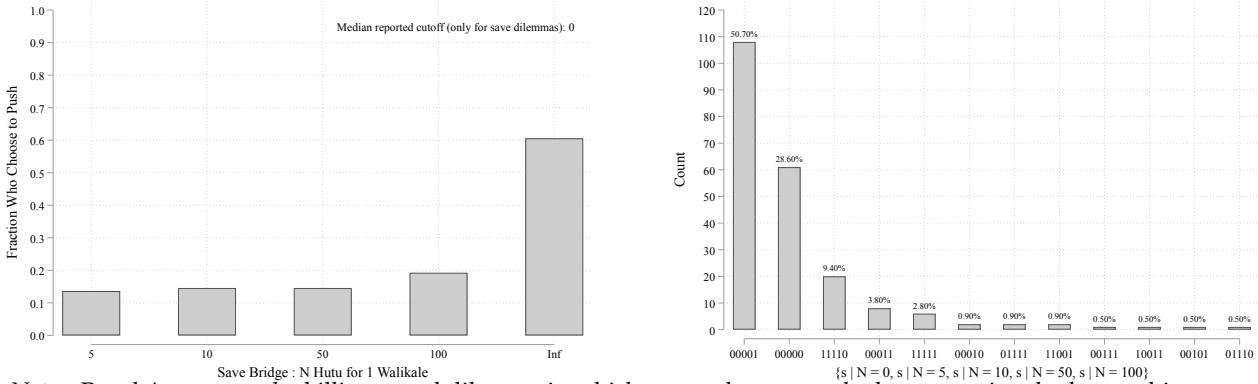
Panel A. Is Killing Own District Members also Morally Right?



Panel B. Saving Bridge Dilemma: Saving N Members of the District by Sacrificing 1 Enemy Member



Panel C. Saving Bridge Dilemma: Saving N Enemy Members by Sacrificing 1 Member of the District



Notes. Panel A presents the killing moral dilemma in which respondents are asked to ascertain whether pushing a civilian of the District off the bridge is morally right, when it results in the death of another $N \in \{0, 5, 10, 50, 100\}$ civilians from the District. Panels B and C present the standard bridge moral dilemma in which not pushing results in the death of the people in the rail, but pushing saves them while killing the person pushed.

Figure A27: Anger Associated to the War

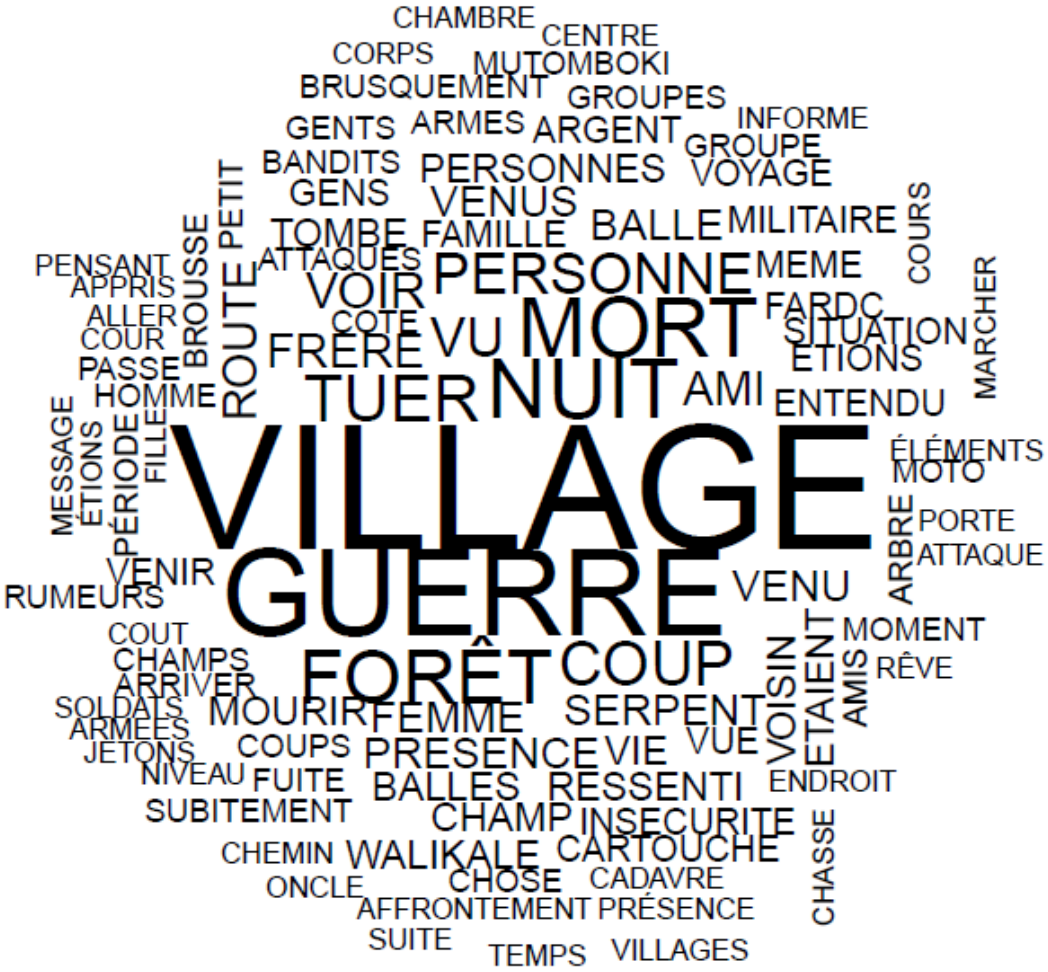
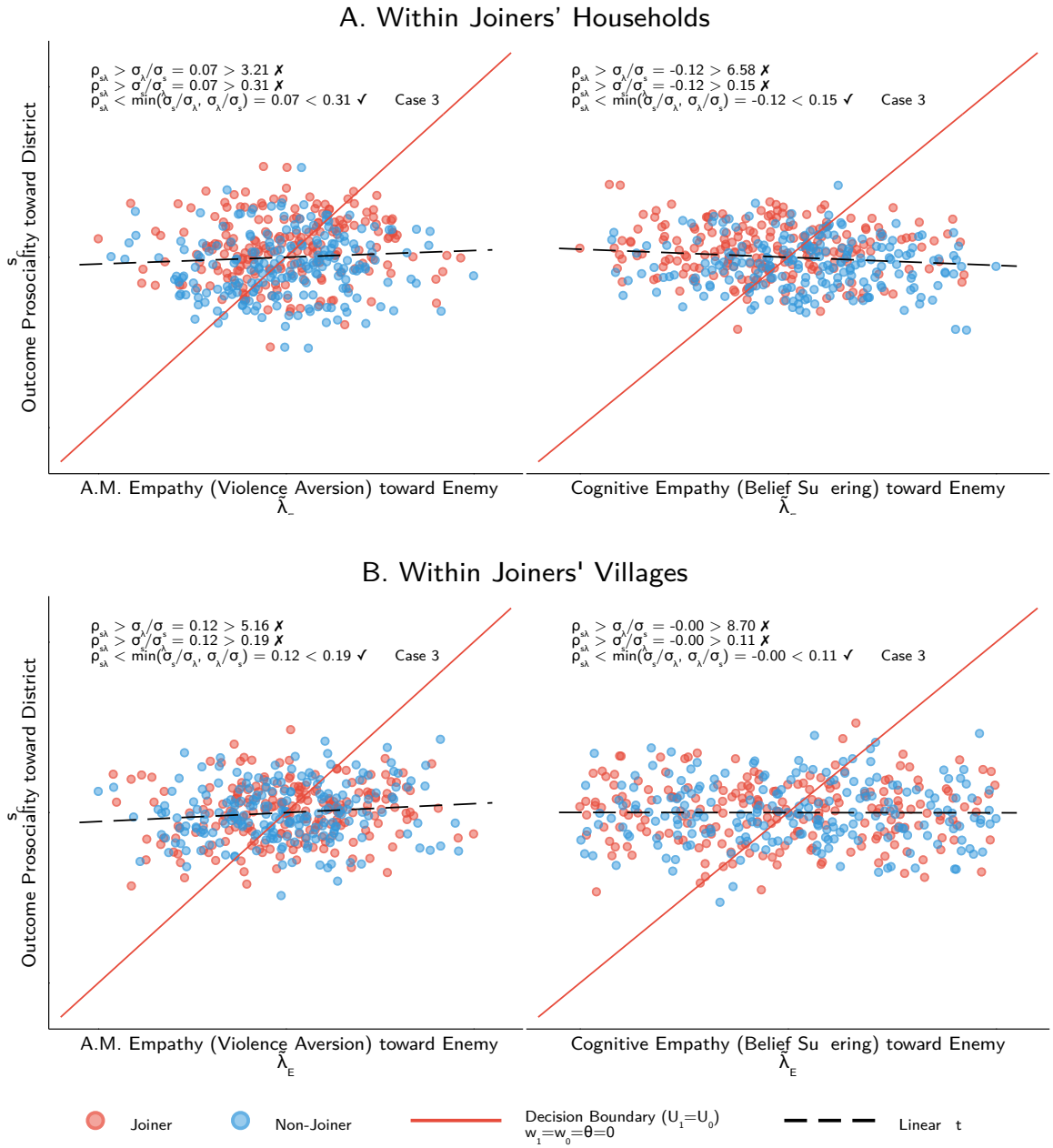
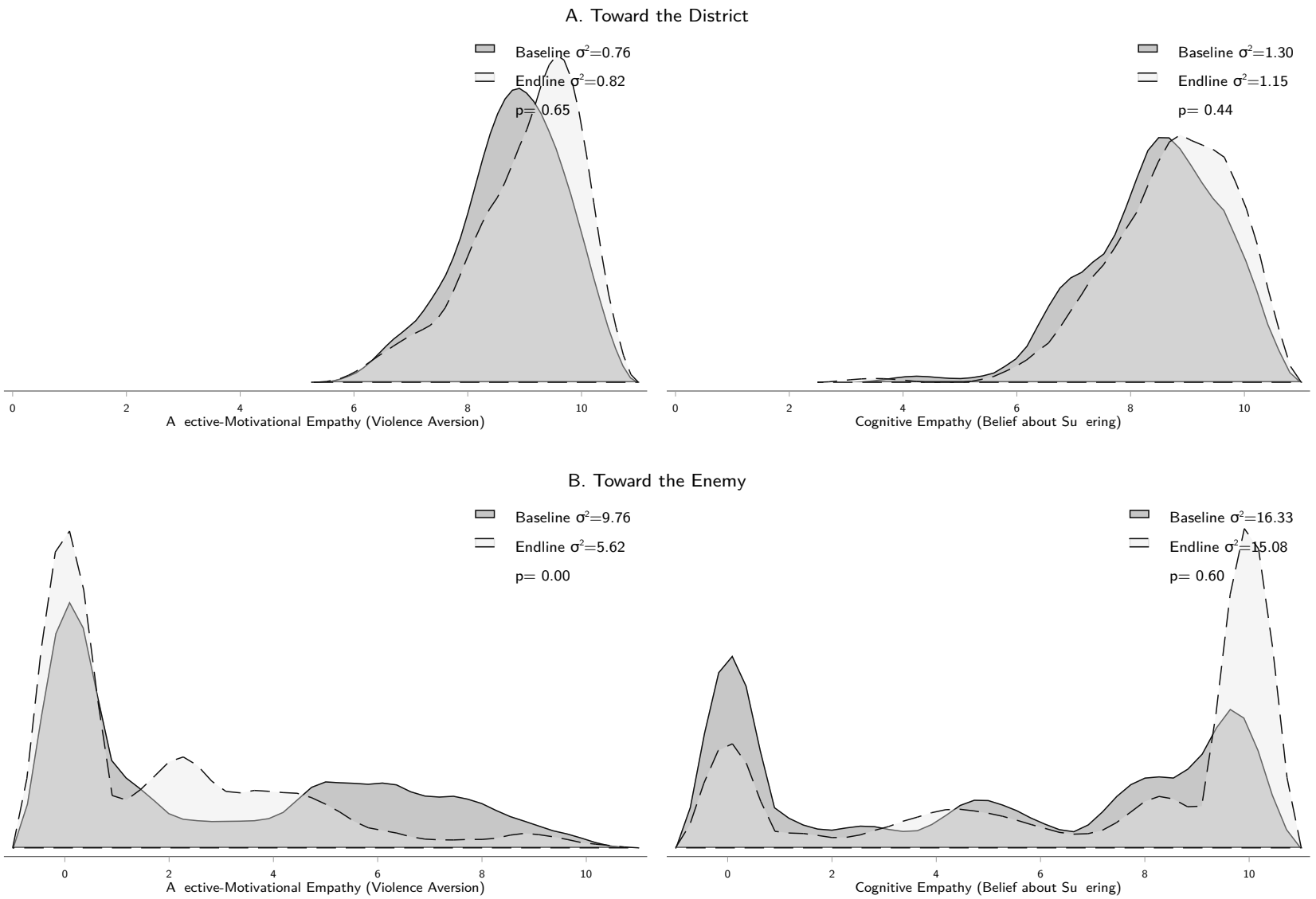


Figure A28: The *Key Relationship*: Outcome Preferences toward the District and Violence Aversion Toward the Enemy—Residualized Controlling for Moral Variables



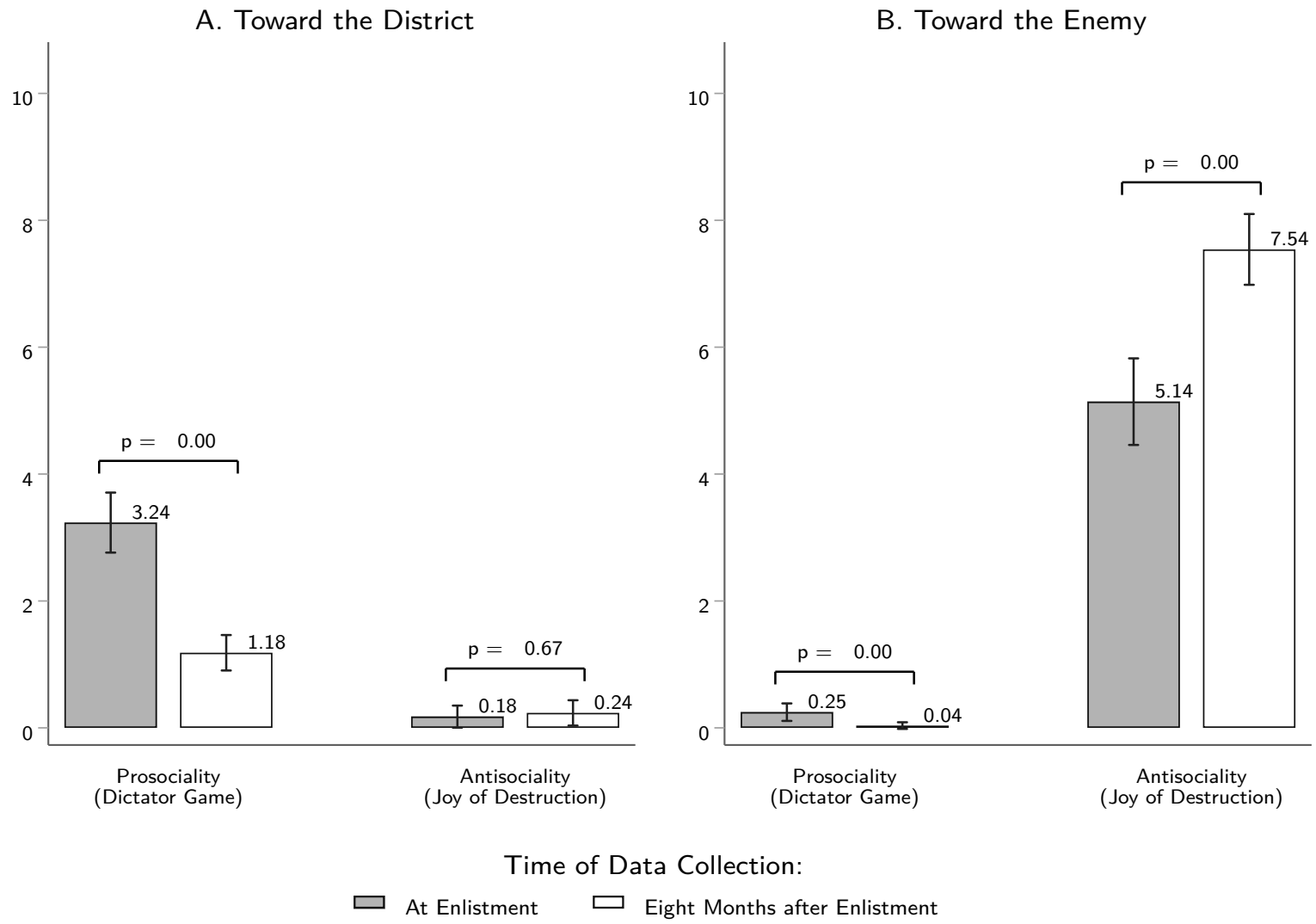
Notes. This figure replicates the original scatter plot of Figure A12 and calculations on the residualized data, showing that inclusion of these three variables leaves the selection of case 3 unchanged.panels) toward an enemy member. The linear fit is estimated through OLS.

Figure A29: Violence Aversion, After Enlisting—Kernel Distributions



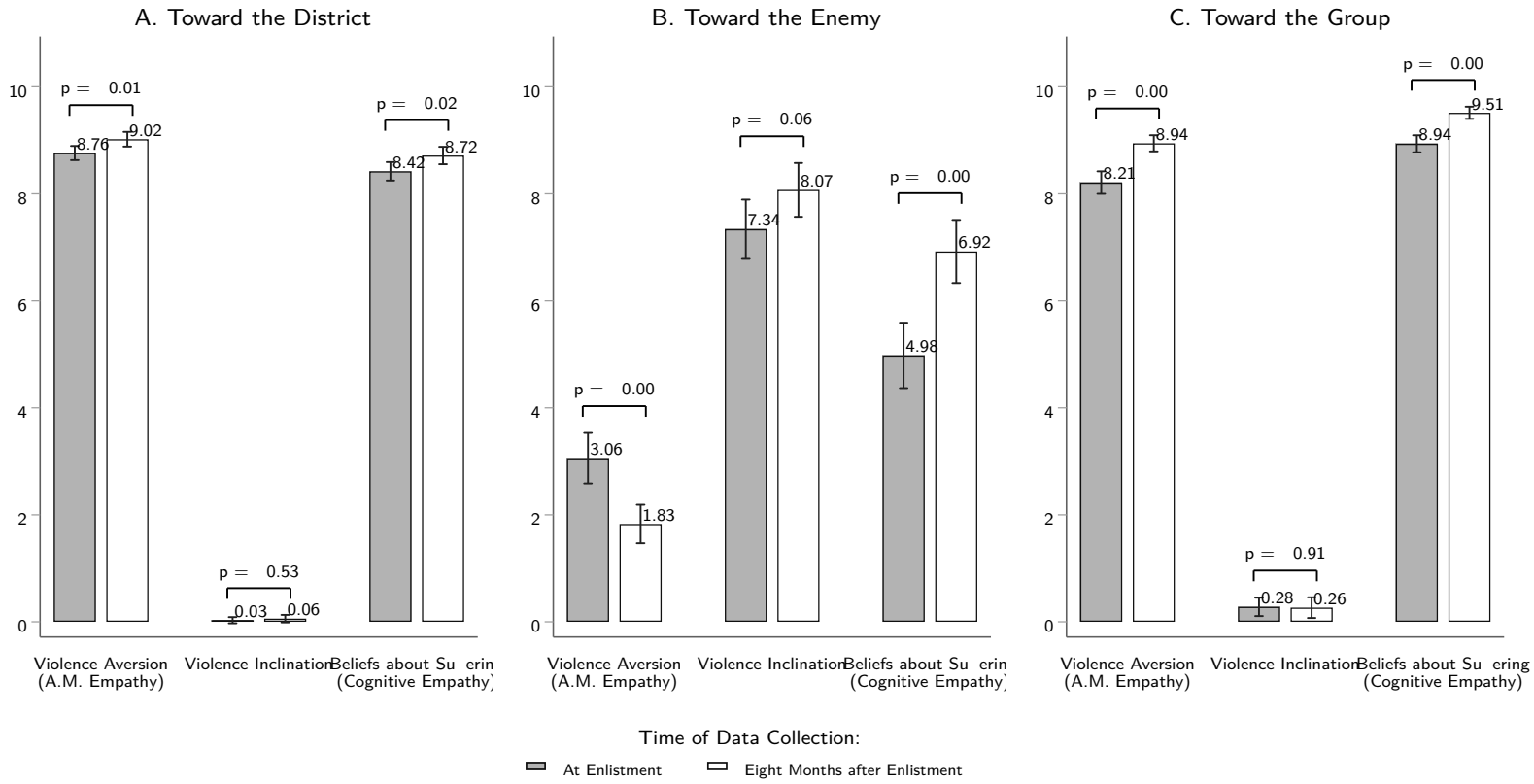
Notes. This figure presents the smoothed kernel distributions of the outcomes presented in Figure 7.

Figure A30: Outcome Preferences, After Enlisting



Notes. All outcomes are rated from 0 to 10. For comparability between baseline and endline measures, the measures of prosocial preferences include only the dictator game, since this was the only measure of prosociality measure at endline due to the shorter window available for an interview. Furtheromre, the sample is restricted to the stayers, that is, it includes only the 169 joiners who did not drop out of the sample at endline (those who left the sample include dead and those who quit). The p-values in those quadrants are the p-values of the two-sided test of equality of means of baseline vs. endline for each corresponding variable in which we include, as controls, treatment indicators of a separate experiment conducted among new joiners in which the treatment was administered after the baseline; thus, the treatment does not affect the baseline measures but could potentially affect the endline measures. Treatments controlled for include participation in perspective taking designed to promote empathy toward civilians, international humanitarian law and norms training.

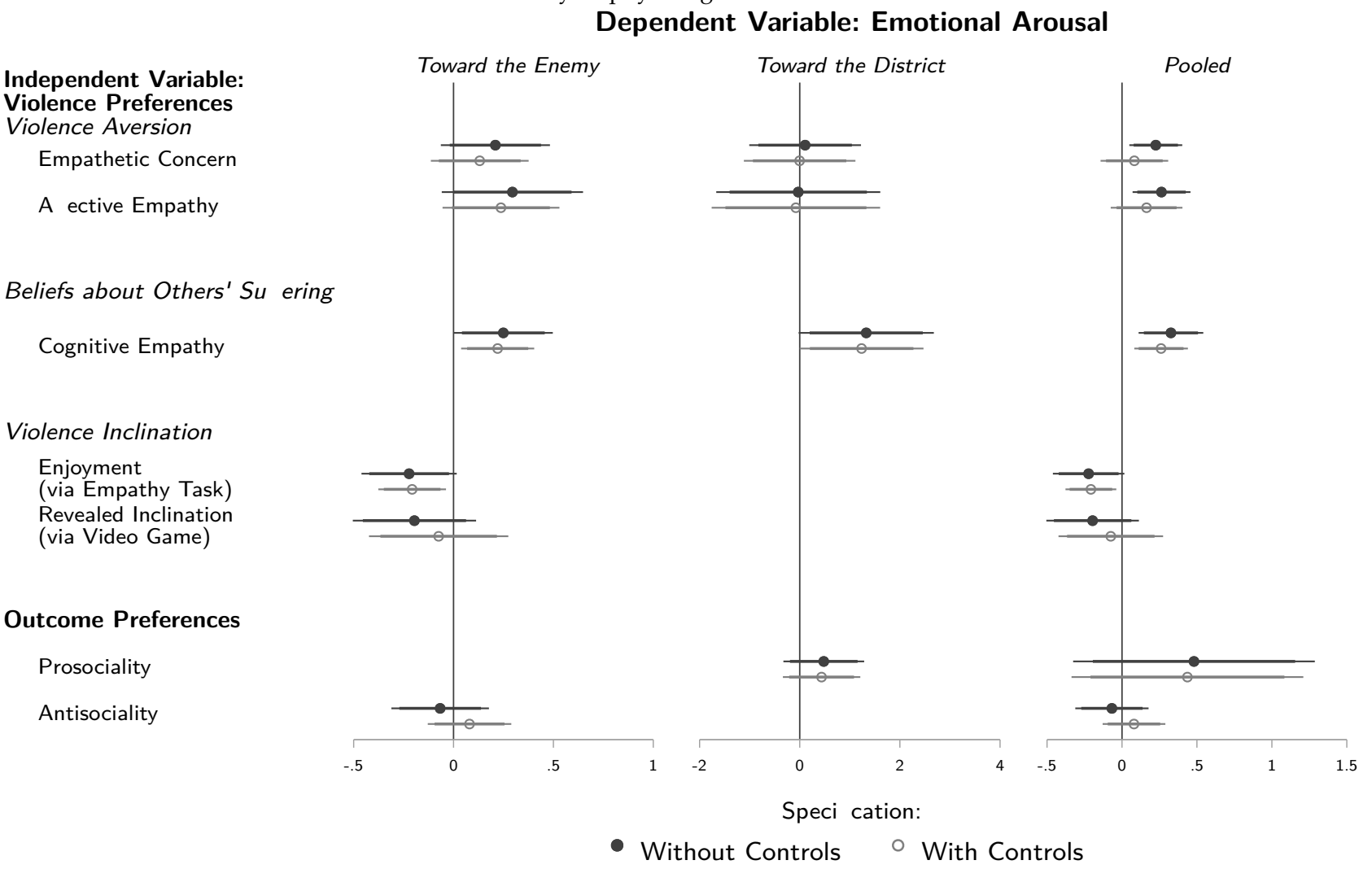
Figure A31: Violence Aversion Toward Group Members, After Enlisting



Notes. This figure replicates the analysis of Figure 7 using group members as target identity.

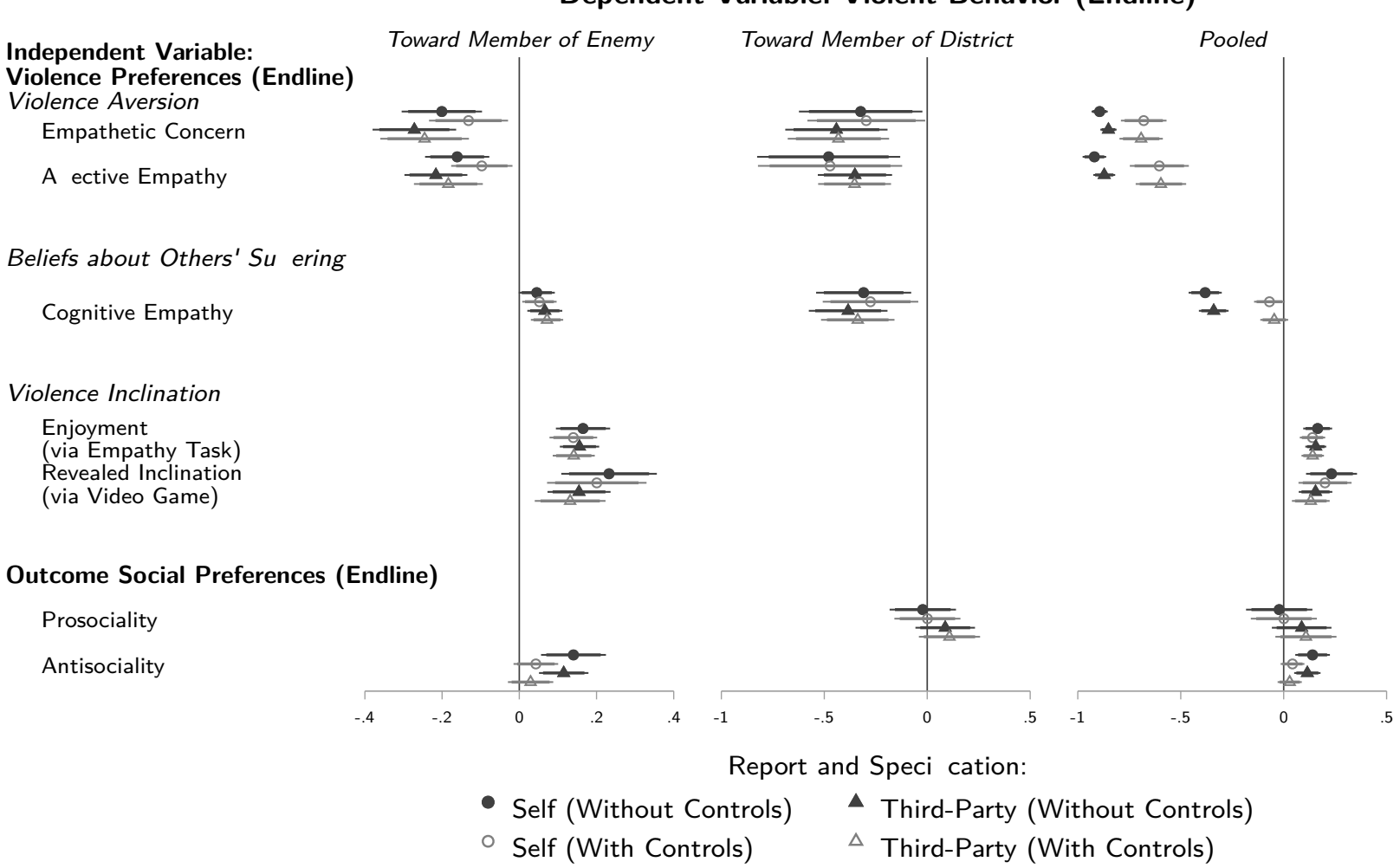
Figure A32: Violence Preferences: Measurement Validation

Panel A. Validation with Psychophysiological Measures of Emotional Arousal



Notes. This figure presents the coefficients and confidence intervals from the following equation estimated through OLS, where i indexes the respondent: $y_i = \alpha + \beta A_i^j + \epsilon_i$, where the dependent variable y_i in Panel A is the measure of *psychophysiological emotional arousal* recorded in the seconds immediately following witnessing the prime of the person suffering in the *suffering empathy task* in the baseline The explanatory variable in both panels A_i^j represents a measure of unbundled social preferences collected in the survey from respondent i towards target identity $j, j \in \{E, D\}$. The specific measure used varies row by row as follows: violence aversion (the average of empathetic concern and affective empathy, measured as affective-motivational empathy, A.M. empathy), violence inclination (enjoyment, revealed inclination), and social preferences (dictator game and joy of destruction game), all rated from 0 to 10, and the target j used varies across columns as follows: left column shows these measures toward a member of the enemy, middle column towards a member of the District, and right column regardless of identity hence pooling the targets. For comparability with the analysis of the endline measures presented in Panel B, where only a sub-set of social preferences was gathered, the measures of social preferences include only the dictator/joy of destruction game, but the results are qualitatively identical for other measures of social preferences. Standard errors are clustered at the level of the respondent. The measure of psychophysiological emotional arousal is obtained through data recorded in bracelets wore by the respondents, and analysis of electrodermal activity in that time window. We use the standard measure of emotional arousal using skin conductance response/electro-dermal activity: the sum of the area of Skin Conductance Response (SCR), in units of MicroSiemens Seconds ($\mu S \times s$). Due to logistical and budgetary constraints, we obtained usable data on electrodermal activity for a non-random sample of 53 randomly selected youth (out of the 213), 47 randomly selected youth in the joiners' households (out of the 213), and 112 new joiners (out of the 213). For consistency with the available data of violent behavior in the group, and due to the small sample of civilians with EDA measurements (which were designed for the joiners due to the prohibitive costs of EDA measuring equipment), the sample in all columns is the joiners. Thus, the sample in the first two columns is constituted of (n=112 joiners) $\times$ type of situation/hurt (n=2: fall from tree, knife cutting ear), totaling 224. The sample in the third column is constituted of (n=112 joiners) $\times$ type of situation/hurt (n=2: fall from tree, knife cutting ear) $\times$ target identity (n=2: Enemy, District) totaling 448. The analysis of dimensions of unbundled social preferences that have almost identical responses (eg., violence inclination toward members of the District) are removed for presentability (due to their large standard errors masking the other coefficients and uninformative values). The panel includes as controls social preferences when the explanatory variable being analyzed belongs to the set of variables in violence preferences, and conversely when the explanatory variable being analyzed belongs to the set of variables in outcome preferences.

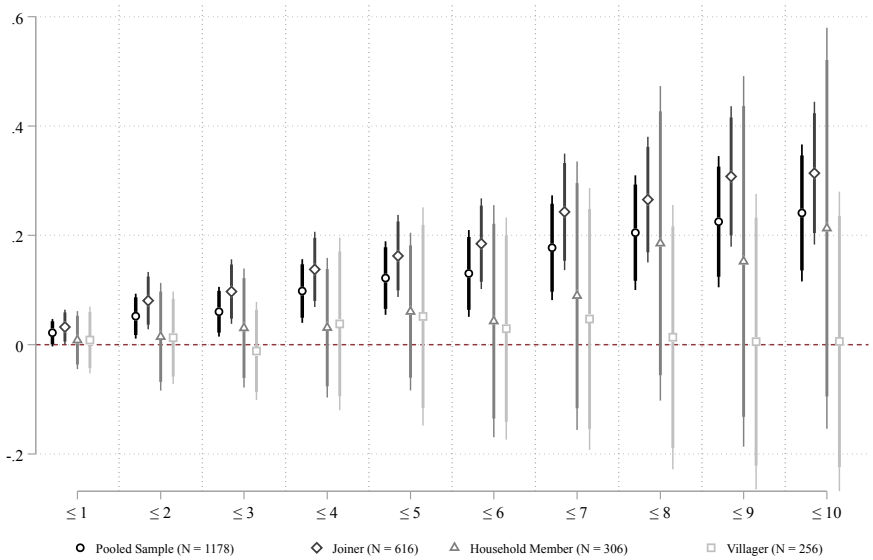
Panel B. Validation with Third-Party Assessments of Actual Behavior in the Group After Enlisting



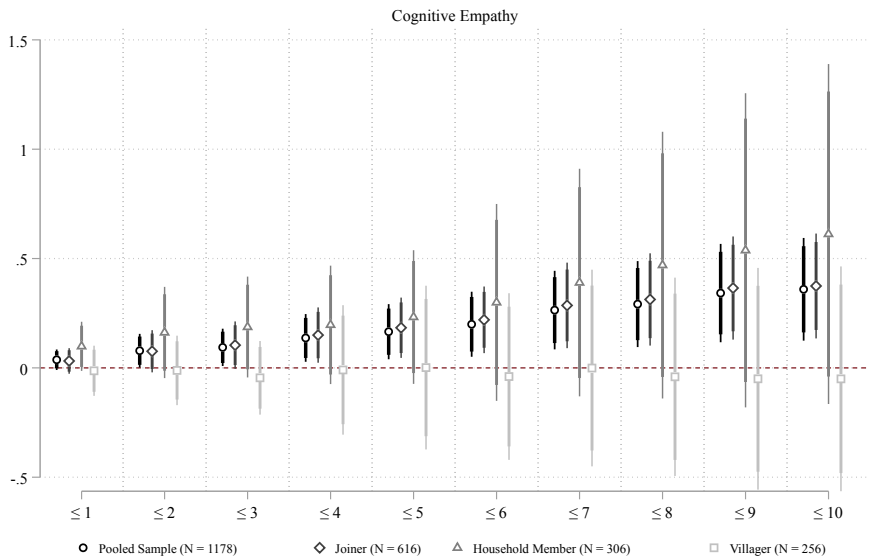
Notes. As in Panel A, this panel presents the coefficients and confidence intervals from the following equation estimated through OLS, where i indexes the respondent: $y_i = \alpha + \beta A_i^j + \epsilon_i$, where the dependent variable y_i is the measure of combatant behavior obtained through behavioral assessments eight months after enlisting at endline from the supervisor, a peer, and the combatant. The explanatory variable and panel structure is the same as in Panel A but using the measures collected at endline so that they can be compared to the measure of behavior also collected at endline. The supervisor was the commander directly in charge of the joiner, and the peer was randomly selected from a roster that we created for each joiner listing all the peers of the same level that know the joiner well-enough. To measure violent behavior across targets in a context-relevant fashion, we asked the evaluator to report how high is the tendency of the joiner being evaluated to, on a scale from 0 to 10 (from less to more violent): a. abuse the District's civilians; b. be violent against enemy members. It was natural to ask the following: "under what circumstances might this person be violent or even shoot an enemy combatant when they could" for violence toward enemy members; and "under what circumstances might this person abuse District civilians when having the opportunity to do so." The panel also includes as controls social preferences when the explanatory variable being analyzed belongs to the set of variables in violence preferences, and conversely when the explanatory variable being analyzed belongs to the set of variables in outcome preferences.

Figure A33: Empathy Measurements and Psychophysiological Emotional Arousal: All Empathy Types

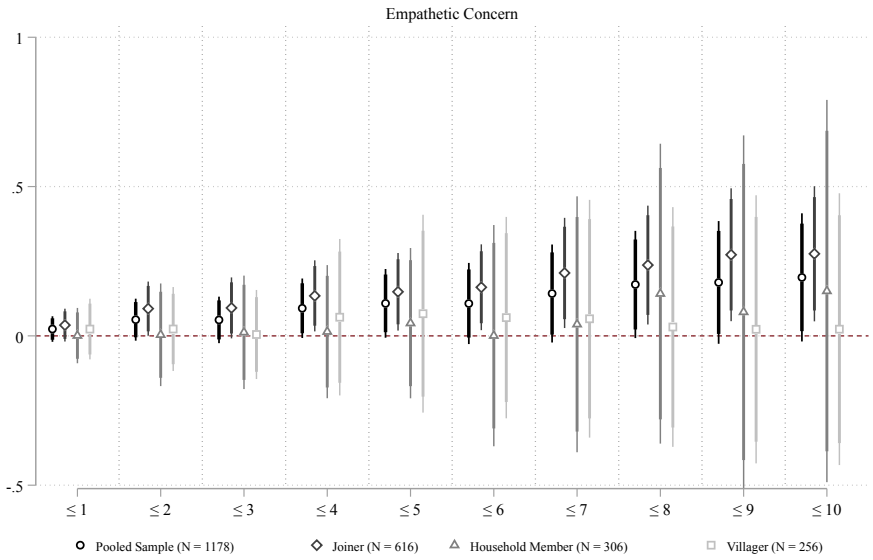
Panel A. All Empathies Pooled



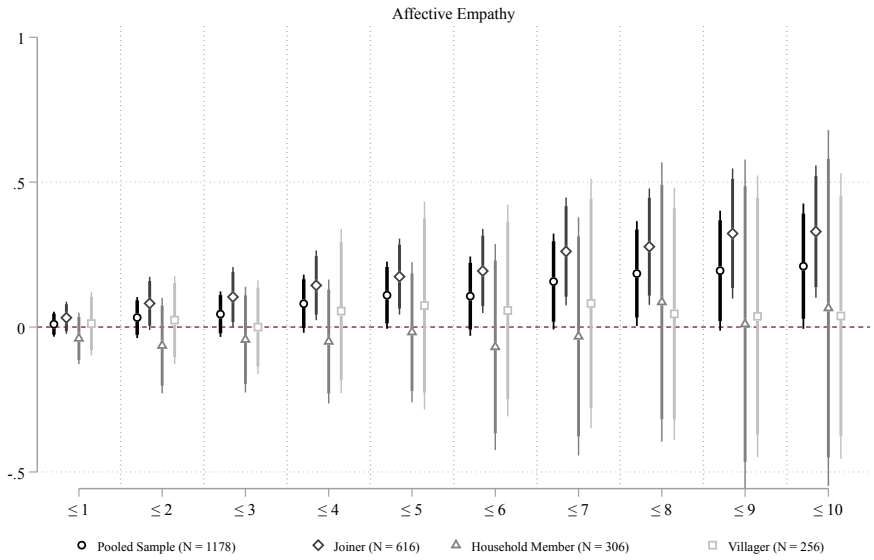
Panel B. Cognitive Empathy



Panel C. Empathic Concern

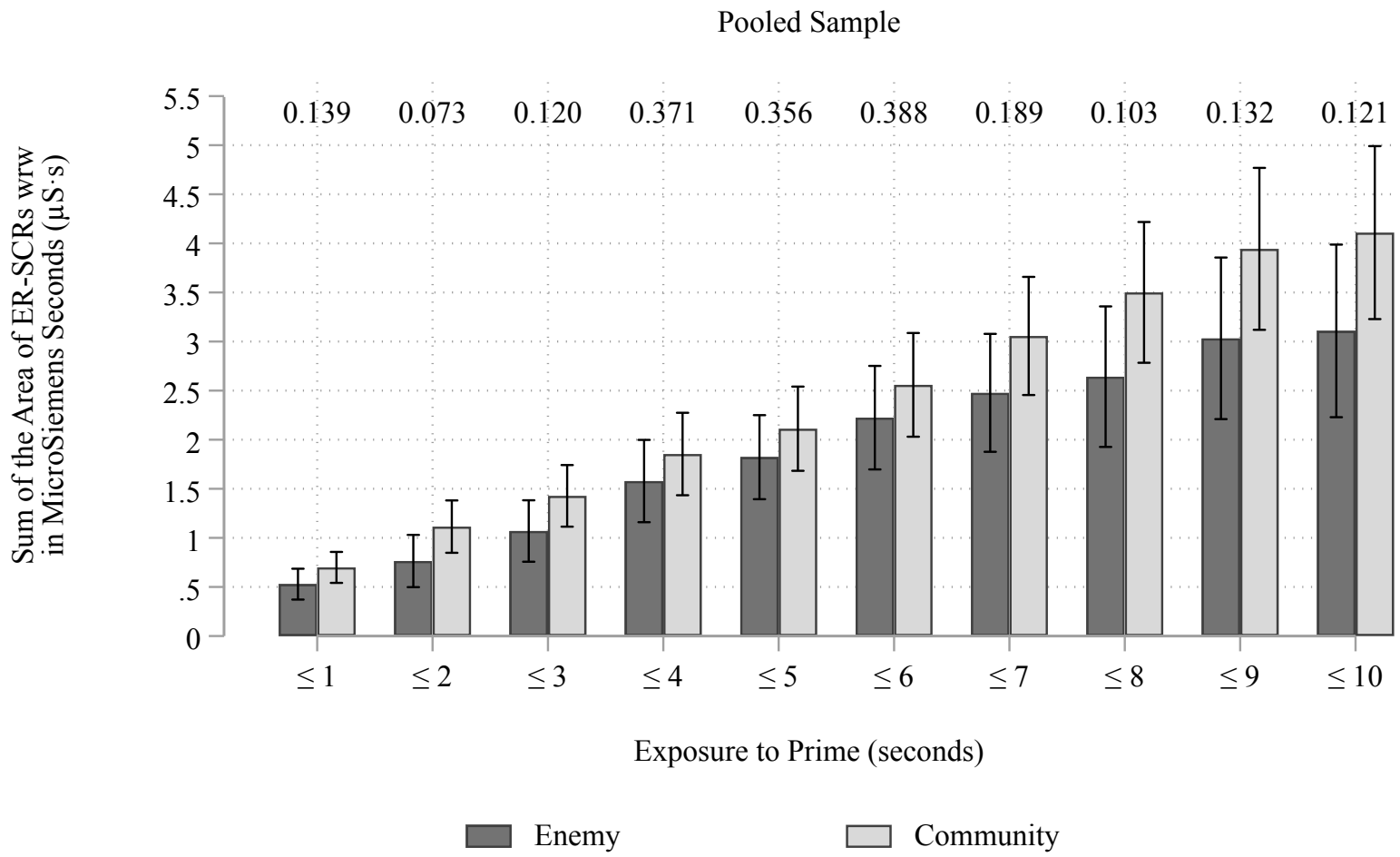


Panel D. Affective Empathy



Notes. This figure presents the coefficients of a regression of the measure of emotional arousal immediately after witnessing the prime of the person suffering in the *suffering empathy task*, obtained through analysis of electrodermal activity, on the reported empathic response for affective empathy, empathetic concern, and cognitive empathy in the *suffering empathy task* upon observing the same prime. We estimate a regression of the three reported measures of empathy from the *suffering empathy task* (Panel A: pooled, Panel B: cognitive empathy; Panel C: Empathic concern; Panel D: affective empathy) on the measure of emotional arousal in psychophysiological data obtained within seconds upon observing the corresponding prime. The dependent variable is the average rating (from 0 to 10) reported in affective empathy, empathic concern, cognitive empathy. The independent variable is the standard measure of emotional arousal using skin conductance response/electro-dermal activity: it is the sum of the area of Skin Conductance Response (SCR), in units of MicroSiemens Seconds ($\mu S \times s$). The sample (n=1,117) is the number of times a respondent interacts with a prime, generated by 125 new joiners, 59 villagers, and 44 joiner household members (n=228 respondents), three identities (District member, enemy, group's combatant), and two primes (an image of ear cutting, and an image of falling from a tree). Thus, in all subsequent analysis of emotional arousal, standard errors are clustered at the level of the respondent. Due to logistical and budgetary constraints, we obtained usable data on electrodermal activity for a non-random sample of 44 randomly selected youth (out of the 213), 59 randomly selected youth in the joiners' households (out of the 213), and 125 new joiners (out of the 213). The sample in this panel is the number of times a *suffering empathy task* prime is played. The sample sizes in the table indicate the number of respondents, times the number of primes observed (6, composed of three identities and two types of situations—fallen from a tree, and ear being cut).

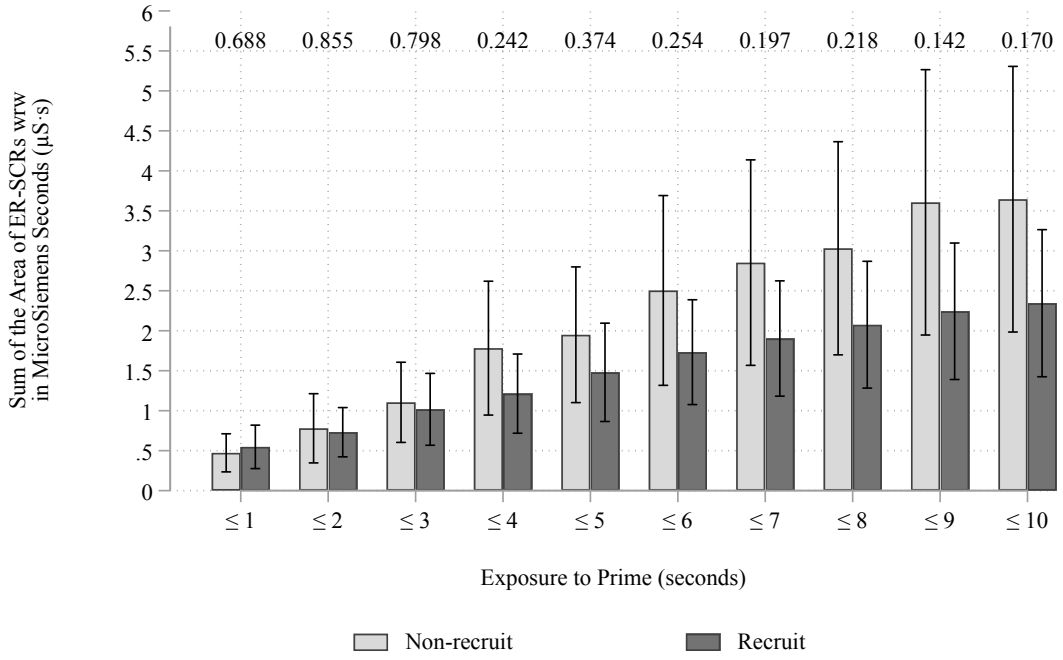
Figure A34: Lower Empathy toward enemy members Replicates Using Emotional Arousal



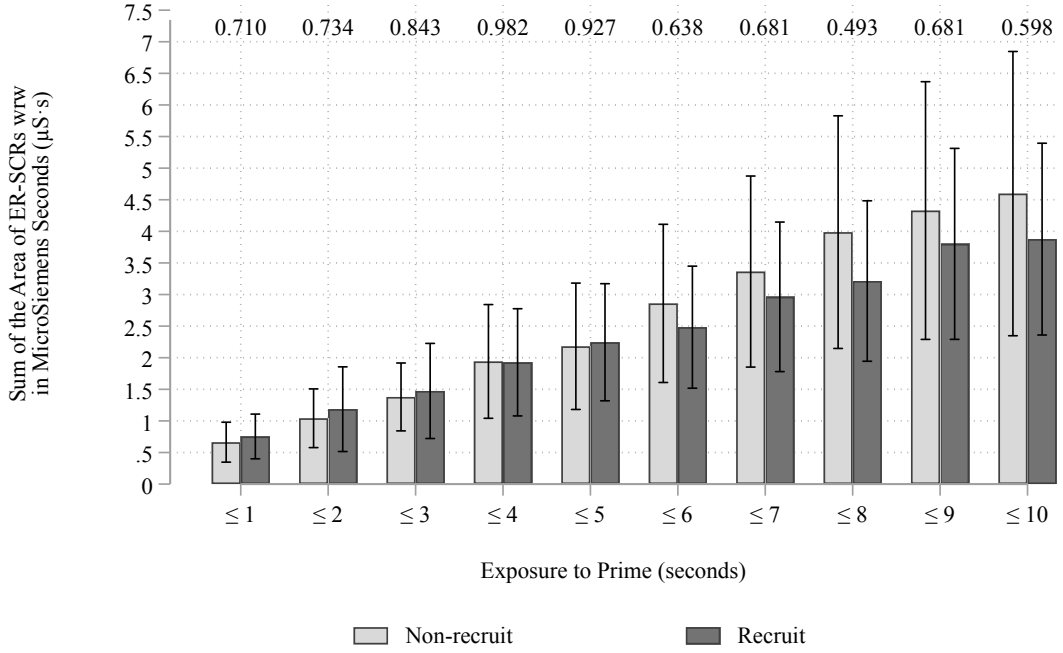
Notes.

Figure A35: The Joiners' Distinctive Violence Preferences, in Emotional Arousal

Panel A. Witnessing the Pain of an enemy member Arouses Joiners Less than Others



Panel B. Witnessing the Pain of a District Member Arouses Joiners as Others



Notes. This figure shows the emotional arousal over the pain of an enemy member (upper panel) and over the pain of a District member (lower panel), separately for non-joiners and for joiners, respectively.

Table A1: Outcome Preferences and Violence Preferences Measure Different Things

Panel A. The Relationship Between Outcome Preferences and Violence Preferences

	Toward X			
	A.M Empathy (Violence Aversion)	Cognitive Empathy (Suffering Belief)	Violence Inclination	
	(1)	(2)	(3) Enjoyment	(4) Revealed I.
Outcome Prosociality toward X	1.086***	0.561***	-1.242***	-1.456***
	(0.056)	(0.061)	(0.078)	(0.080)
Mean Dependent Variable	6.34	7.15	2.90	4.13
R^2	0.47	0.17	0.38	0.44
Observations	424	424	424	426

Panel B. Social Preferences to District vs. Social Preferences to Enemy: An (Unbundled) Social Preferences Budget Constraint?

	Toward the Enemy				
	Outcome Prosociality	A.M Empathy (Violence Aversion)	Cognitive Empathy (Suffering Belief)	Violence Inclination	
	(1)	(2)	(3)	(4) Enjoyment	(5) Revealed I.
Prosociality toward the District	0.201***				
	(0.073)				
A.M. Empathy toward the District		0.448*		0.520	-0.016
		(0.252)		(0.584)	(0.493)
Cognitive Empathy toward the District			0.091	-0.524	0.432
			(0.214)	(0.432)	(0.378)
Mean Dependent Variable	0.91	3.78	5.87	5.80	7.98
R^2	0.04	0.02	0.00	0.01	0.02
Observations	213	212	212	212	212

Notes. In Panel A, we regress measures of violence preferences toward a given target on the index of social preferences toward the same target. Panel B examines the correlations of these notions across targets.

Table A2: Outcome Preferences and Violence Aversion—All Targets Pooled, Understanding the Correlations

Panel A. Violence Aversion toward X vs. Outcome Prosocial Preferences toward X

	Toward X					
	A.M Empathy (Violence Aversion)			Cognitive Empathy (Suffering Belief)		
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Prosociality toward X	0.489***	1.265***	0.498***	0.297***	0.652***	0.361**
	(0.083)	(0.063)	(0.129)	(0.098)	(0.069)	(0.154)
Target FE	✓		✓	✓		✓
Individual FE		✓	✓		✓	✓
Mean dep. var	6.34	6.34	6.34	7.15	7.15	7.15
R^2	0.56	0.75	0.79	0.19	0.59	0.60
Observations	424	424	424	424	424	424

Panel B. Violence Inclination toward X vs. Outcome Prosocial Preferences toward X

	Toward X					
	Enjoyment (Violence Inclination)			Revealed Inclination (Violence Inclination)		
	(1)	(2)	(3)	(4)	(5)	(6)
Outcome Prosociality toward X	-0.588***	-1.467***	-0.744***	-0.273**	-1.848***	-0.472***
	(0.120)	(0.087)	(0.189)	(0.107)	(0.087)	(0.168)
Target FE	✓		✓	✓		✓
Individual FE		✓	✓		✓	✓
Mean dep. var	2.90	2.90	2.90	4.13	4.13	4.13
R^2	0.44	0.70	0.72	0.62	0.74	0.81
Observations	424	424	424	426	426	426

Notes. This table presents the estimates from a regression of the dependent variable listed on the column headers on the explanatory variable listed on the row header. The sample is at the respondent-social preference target level. Thus, each respondent is a group of two observations, each defined by the corresponding social preference toward the District, and toward the enemy, where $X \in \{Enemy, District\}$. Standard errors are clustered at the level of the individual respondent.

Table A3: The *Key Relationship*

	Toward the Enemy					
	Cognitive Empathy (Suffering Beliefs)		A.M. Empathy Violence Aversion		Violence Inclination	
	(1)	(2)	(3)	(4)	(5) Enjoyment	(6) Revealed I.
A. Joiners and Household Members						
Prosociality toward the District	-0.503***	-0.367**	-0.184	0.073	0.191	0.066
	(0.183)	(0.147)	(0.139)	(0.112)	(0.228)	(0.212)
A.M. Empathy toward the Enemy		0.740***				
		(0.062)				
Cognitive Empathy toward the Enemy				0.511***		
				(0.042)		
Joiner ID FEs	✓	✓	✓	✓	✓	✓
Mean Dependent Variable	5.97	5.94	3.61	3.68	5.91	7.73
R^2	0.48	0.67	0.50	0.69	0.51	0.52
Observations	420	420	420	420	420	426
B. Joiners and Random Villagers						
Prosociality toward the District	-0.044	-0.209	0.202	0.228*	-0.519**	-0.002
	(0.213)	(0.146)	(0.194)	(0.133)	(0.258)	(0.195)
A.M. Empathy toward the Enemy		0.816***				
		(0.055)				
Cognitive Empathy toward the Enemy				0.581***		
				(0.041)		
Joiner ID FEs	✓	✓	✓	✓	✓	✓
Mean Dependent Variable	5.46	5.94	3.47	3.68	5.91	7.73
R^2	0.46	0.72	0.45	0.71	0.48	0.54
Observations	418	418	418	418	418	426

Notes. This table shows the *Key Relationship* separately for the sample of joiners and household members (Panel B.1.) and for joiners and randomly selected villagers (Panel B.2). Since the comparison groups are tied to each joiner, they are self selected, hence the analysis always includes fixed effects for joiner ID, which is shared between each joiner, its randomly sampled household member, and its randomly sampled villager. The analysis always uses fixed effects for joiner ID, shared between each joiner, its randomly sampled household member, and its randomly sampled villager, to exclude the selection of comparison groups. The analysis reveals substantial heterogeneity in the *Key Relationship*. Within the households of the joiners, individuals who are more prosocial to the District tend to have less affective empathy and especially less cognitive empathy toward the enemy.

Table A4: Joiners are the "Zealots" of Society

	Household (Mean)	Joiner - HH (Difference)	Villager (Mean)	Joiner - Vill (Difference)	Pooled (Mean)	Joiner - Pooled (Difference)
	(1)	(2)	(3)	(4)	(5)	(6)
A. Violence Preferences						
A.1. Toward the District						
Violence Aversion						
Affective-Motivational Empathy	8.73	-0.03	8.90	-0.19**	8.82	-0.10
Violence Inclination						
Enjoyment (via Empathy Task)	0.09	-0.05	0.00	0.05	0.05	-0.00
Revealed Inclination (via Video Game)	0.05	0.10	0.28	-0.14	0.16	-0.02
A.2. Toward the Enemy						
Violence Aversion						
Affective-Motivational Empathy	4.10	-0.96***	3.78	-0.67**	3.94	-0.81***
Violence Inclination						
Enjoyment (via Empathy Task)	5.31	1.29**	5.80	0.84*	5.55	1.07***
Revealed Inclination (via Video Game)	7.00	1.18***	7.98	0.23	7.49	0.73**
B. Outcome Preferences						
B.1. Toward the District						
Prosociality						
Declared Prosociality	6.54	0.62***	6.54	0.61***	6.54	0.62***
Time Dictator Game: Socializing Opp. Cost, Personal Hours Used	2.16	0.89***	2.54	0.52	2.34	0.71***
Time Dictator Game: Unconstrained Opp. Cost, Personal Hours Used	2.19	1.09***	2.88	0.37	2.53	0.73***
Time Dictator Game: Working Hours Opp. Cost, Personal Hours Used	1.83	0.87***	2.20	0.48	2.02	0.67***
Hypothetical Donation	2.67	0.45***	3.18	-0.06	2.92	0.20
Dictator Game, Amount Transferred	2.08	1.01***	3.60	-0.53*	2.84	0.25
Antisociality						
Joy of Destruction Game, Amount Destroyed	0.21	-0.08	0.07	0.05	0.14	-0.01
B.2. Toward the Enemy						
Prosociality						
Declared Prosociality	1.47	0.01	1.81	-0.32	1.64	-0.15
Time Dictator Game: Socializing Opp. Cost, Personal Hours Used	0.69	-0.18	0.68	-0.16	0.68	-0.17
Time Dictator Game: Unconstrained Opp. Cost, Personal Hours Used	0.70	-0.20	0.81	-0.33**	0.76	-0.26**
Time Dictator Game: Working Hours Opp. Cost, Personal Hours Used	0.56	-0.10	0.68	-0.22	0.62	-0.16
Hypothetical Donation	0.29	-0.01	0.39	-0.13	0.34	-0.07
Dictator Game, Amount Transferred	0.24	-0.02	0.53	-0.30**	0.38	-0.16*
Antisociality						
Joy of Destruction Gam, Amount Destroyed	2.59	-0.05	2.19	0.34	2.39	0.14
C. Beliefs about Others' Mental States						
C.1. Toward the District						
Pain and Suffering						
Cognitive Empathy	8.26	0.09	8.42	-0.08	8.34	0.01
C.2. Toward the Enemy						
Pain and Suffering						
Cognitive Empathy	6.94	-1.93***	5.87	-0.86**	6.40	-1.41***
D. Psychopathy						
Leveonson, Kiehl, and Fitzpatrick (1995)						
Primary Psychopathy	5.45	-0.18*	5.38	-0.12	5.41	-0.16*
Secondary Psychopathy	4.31	0.22	4.74	-0.19	4.53	0.01
Target Observations	213	426	213	426	426	639

Notes. This table presents the means of the measures of social preferences and violence preferences along the estimates from Equation 1.

Table A5: The Driving Force of Cognitive Empathy

	Dependent Variable: Enlists					
	(1)	(2)	(3)	(4)	(5)	(6)
Affective Empathy toward the Enemy	-0.020***			0.000		0.008
	(0.008)			(0.009)		(0.015)
Empathic Concern toward the Enemy		-0.024***			-0.005	-0.011
		(0.008)			(0.010)	(0.016)
Cognitive Empathy toward the Enemy			-0.031***	-0.032***	-0.029***	-0.031***
			(0.007)	(0.009)	(0.009)	(0.009)
Joiner ID FEs	✓	✓	✓	✓	✓	✓
Mean Dependent Variable	0.33	0.33	0.33	0.33	0.33	0.33
R^2	0.03	0.03	0.05	0.05	0.05	0.05
Observations	635	635	635	635	635	635

Notes. This table presents the estimates from estimating the following equation in OLS: we estimate the following baseline regression in each panel, where the dependent variables differ across panels and where we sequentially add controls as indicated in each panel:

$$y_i = \alpha + \beta_1 AE_i^e + \beta_2 EC_i^e + \beta_3 CE_i^e + \epsilon_i \quad \text{(A1)}$$

where y_i is an indicator variable taking value 1 if individual i is a joiner and zero otherwise, and AE_i^e, EC_i^e, CE_i^e are the measures of affective empathy, empathic concern, and cognitive empathy to the enemy, respectively, all in the scale [0-10]. All regressions include Joiner ID Fixed effects (defined as the triple of joiner, young male in their household, and young male in their village) and includes the whole sample of joiners, young males in joiners' households, and young males in joiners' villages.

Table A6: Socio-Demographic Characteristics

	Household (Mean)	Joiner - HH (Difference)	Villager (Mean)	Joiner - Vill (Difference)	Pooled (Mean)	Joiner - Pooled (Difference)
	(1)	(2)	(3)	(4)	(5)	(6)
Age, Gender, and Literacy						
Age	28.89	-3.58***	27.04	-1.86***	27.97	-2.72***
Female (%)	0.00	1.44*	0.00	1.42	0.00	1.43**
Literate (%)	96.71	-0.84	98.12	-2.40	97.42	-1.63
Reading score [1-4]	3.11	-0.21**	2.74	0.16*	2.92	-0.03
Ethnicity						
Majority ethnic group (%)	67.61	8.34***	69.01	6.31**	68.31	7.37***
Ethnic group 2 (%)	15.02	-2.68	14.08	-0.95	14.55	-1.92
Ethnic group 3 (%)	10.33	-4.23**	7.98	-1.98	9.15	-3.09**
Religion						
Protestant (%)	56.34	9.24**	66.67	-1.14	61.50	3.88
Catholic (%)	17.84	-4.81	12.68	0.34	15.26	-2.13
Other (%)	25.82	-4.43	20.66	0.80	23.24	-1.75
Language						
Speaks Swahili (%)	97.65	0.04	99.06	-1.48	98.36	-0.73
Speaks French (%)	63.38	-4.11	38.03	20.94***	50.70	8.45*
Speaks lang. of maj. ethn. group (%)	53.99	16.27***	52.58	17.03***	53.29	16.64***
Family						
Number of siblings	6.26	-0.35***	5.55	0.37*	5.90	0.02
Birth rank	3.33	0.40**	2.93	0.78***	3.13	0.60***
Observations	213	426	213	426	426	639

Notes. This table presents the estimates of Equation 1, in which the dependent variables are the variables indicated on the left.

Table A7: Economic Conditions

	Household (Mean)	Joiner - HH (Difference)	Villager (Mean)	Joiner - Vill (Difference)	Pooled (Mean)	Joiner - Pooled (Difference)
	(1)	(2)	(3)	(4)	(5)	(6)
Occupation past 5 years						
Farmer (%)	34.27	3.69	38.03	-0.39	36.15	1.84
Student (%)	25.82	-3.38	24.41	-1.99	25.12	-2.61
Digger/Miner (%)	5.63	4.13	5.63	4.25*	5.63	4.25*
Herding (%)	6.57	1.40	5.16	2.79	5.87	2.09
Salaried professional (%)	7.51	-6.59***	3.29	-2.78**	5.40	-4.91***
Employment status past month						
Work in exchange for income (%)	23.47	-8.49**	15.49	-0.57	19.48	-4.48
Work for self (%)	54.93	8.18	66.20	-3.35	60.56	2.31
Volunteer (%)	7.04	4.64	6.10	5.54**	6.57	5.13**
Unemployed (%)	7.04	5.54	6.57	6.23**	6.81	5.95**
Inactive (%)	12.21	-3.29	9.86	-0.92	11.03	-2.08
Labor supply						
Hours worked past month	86.36	-1.04	68.00	16.48**	77.11	7.52
Consumption						
Food consumption (USD/month)	108.30	0.22	95.63	13.39**	101.97	6.50*
Soft drink consumption (USD/month)	18.30	0.16	19.54	-1.14	18.92	-0.59
Ate pre-made food in past 14 days	13.27	-0.17	14.08	-0.82	13.68	-0.37
Ate cooked food in past 14 days	100.00	0.00	98.12	1.87**	99.06	0.95
Times ate dish with carbs (foufou) in past 14 days	9.21	0.02	7.96	1.28***	8.59	0.63**
Beliefs about future income						
PDV of income in group (CRRA discounted)	919.83	111.67**	963.46	68.07*	944.28	88.17**
PDV of income outside group (CRRA discounted)	1017.20	9.98	1037.06	-2.51	1026.66	5.32
Observations	213	426	213	426	426	639

Notes. This table presents the estimates of Equation 1, in which the dependent variables are the variables indicated on the left.

Table A8: Human Capital

	Household (Mean)	Joiner - HH (Difference)	Villager (Mean)	Joiner - Vill (Difference)	Pooled (Mean)	Joiner - Pooled (Difference)
	(1)	(2)	(3)	(4)	(5)	(6)
Education						
Years of education	10.52	-0.31	10.20	-0.01	10.36	-0.17
Primary completed	91.08	0.56	92.96	-1.34	92.02	-0.43
Secondary completed	56.34	-10.90**	46.01	-0.83	51.17	-6.05
Cognitive ability						
Correct Raven Matrices (%)	30.44	3.33	29.66	4.17**	30.05	3.86**
Personality (HEXACO-60)						
Honesty [0,1]	0.45	0.00	0.44	0.02	0.44	0.01
Emotionality [0,1]	0.60	-0.03**	0.60	-0.03***	0.60	-0.03**
Extraversion [0,1]	0.65	-0.00	0.65	-0.00	0.65	-0.00
Agreeableness [0,1]	0.71	-0.02	0.69	-0.00	0.70	-0.01
Conscientiousness [0,1]	0.67	-0.05***	0.62	-0.00	0.64	-0.03**
Openness [0,1]	0.61	-0.01	0.61	-0.01	0.61	-0.01
Physical health						
Height (cm)	161.68	0.25	154.47	7.23***	158.06	3.79***
Weight (kg)	58.58	0.90	57.77	1.62*	58.17	1.20
BMI (kg/m^2)	22.57	-0.17	24.77	-2.31***	23.67	-1.29***
Mental health						
Anxiety (HSCL-25, index [0,1])	0.18	0.08***	0.25	0.01	0.21	0.04***
Depression (HSCL-25, index [0,1])	0.24	0.06***	0.29	0.01	0.27	0.04***
Sleep quality [0,1]	0.41	-0.05***	0.40	-0.04**	0.40	-0.04***
Smoking						
Cigarettes (sticks/day)	0.27	0.46**	0.50	0.23	0.38	0.35*
Weed (sticks/day)	0.02	0.02	0.03	0.01	0.03	0.02
Alcohol consumption						
Beer (liters/day)	0.76	0.55***	0.76	0.55***	0.76	0.56***
Local alcohol (liters/day)	0.18	0.48***	0.38	0.27**	0.28	0.38***
Blood Alcohol Concentration (p.p., ethylotest)	0.56	0.16	0.52	0.18	0.54	0.18
Observations	213	426	213	426	426	639

Notes. This table presents the estimates of Equation 1, in which the dependent variables are the variables indicated on the left. Smoking cigarettes and weed, and drinking beer and local alchol were asked as "how many [unit] do you consume in a typical day?"

Table A9: Economic Preferences

	Household (Mean)	Joiner - HH (Difference)	Villager (Mean)	Joiner - Vill (Difference)	Pooled (Mean)	Joiner - Pooled (Difference)
	(1)	(2)	(3)	(4)	(5)	(6)
Fundamental economic preferences						
Risk taker (%)	40.38	-14.27***	34.74	-8.52**	37.56	-11.74***
Risk preference (switching point)	87.64	-31.85***	88.59	-32.80***	88.12	-33.60***
Risk taker - deadly bridge (%)	7.51	1.43	9.86	-0.52	8.69	0.25
Risk preference (deadly bridge; switching point)	86.79	231.98	271.43	19.05	181.65	117.53
Risk taker - Russian roulette (%)	3.76	0.39	2.82	1.41	3.29	0.94
Risk preference (Russian roulette; switching point)	43.40	32.16	103.57	-40.75	74.31	-7.34
Patient (%)	25.82	-13.30***	19.72	-6.92**	22.77	-10.08***
Time preference (switching point)	242.97	14.62***	249.67	7.64**	246.33	11.09***
Social Preferences						
Amount transfered in trust game (District)	0.46	0.04**	0.52	-0.01	0.49	0.02
Declared trust (good itentions) (District)	0.61	-0.01	0.56	0.05**	0.59	0.02
Declared trust (good itentions) (General)	0.55	-0.05***	0.51	-0.01	0.53	-0.03
Cheating game						
Number of points	36.67	1.39**	36.96	1.08**	36.81	1.24**
Justice sensitivity scale (Schmitt et al., 2005)						
Index [0,1]	0.56	0.04***	0.59	0.01	0.58	0.03**
Observations	213	426	213	426	426	639

Notes. This table presents the estimates of Equation 1, in which the dependent variables are the variables indicated on the left.

Table A10: Moral Foundations

	Household (Mean)	Joiner - HH (Difference)	Villager (Mean)	Joiner - Vill (Difference)	Pooled (Mean)	Joiner - Pooled (Difference)
	(1)	(2)	(3)	(4)	(5)	(6)
Individualism and Collectivism						
All, Mean [0,1]	0.72	-0.01	0.71	-0.00	0.71	-0.01
Horizontal, Mean [0,1]	0.71	-0.01	0.72	-0.02	0.72	-0.01
Vertical, Mean [0,1]	0.73	-0.02*	0.69	0.02	0.71	-0.00
Morality as Cooperation						
Family Values, Mean [0,1]	0.86	-0.02*	0.86	-0.02*	0.86	-0.02**
Group Loyality, Mean [0,1]	0.85	-0.02	0.84	-0.01	0.85	-0.01
Reciprocity, Mean [0,1]	0.85	-0.01	0.84	-0.00	0.85	-0.00
Bravery, Mean [0,1]	0.84	-0.01	0.84	-0.01	0.84	-0.01
Respect, Mean [0,1]	0.74	-0.01	0.76	-0.03***	0.75	-0.02*
Fairness, Mean [0,1]	0.46	-0.00	0.46	-0.00	0.46	-0.00
Property Rights, Mean [0,1]	0.38	-0.02	0.37	-0.01	0.37	-0.02
Observations	213	426	213	426	426	639

Notes. This table presents the estimates of Equation 1, in which the dependent variables are the variables indicated on the left.

Table A11: Past Victimization by Enemy Members: The Facts

	Mean	SD	Min	P25	Median	P75	Max
All Attacks	5.62	8.53	0.00	0.00	2.00	8.00	47.00
Enemy Attacks							
Murder	1.92	2.69	0.00	0.00	1.00	3.00	12.00
Sexual violence	0.67	1.50	0.00	0.00	0.00	1.00	10.00
Torture	1.70	3.13	0.00	0.00	0.00	2.00	12.00
Looting	0.56	0.50	0.00	0.00	1.00	1.00	1.00
Non-enemy Attacks							
Murder	0.66	1.70	0.00	0.00	0.00	0.00	12.00
Sexual violence	0.15	0.82	0.00	0.00	0.00	0.00	8.00
Torture	0.54	1.76	0.00	0.00	0.00	0.00	12.00
Looting	0.30	0.46	0.00	0.00	0.00	1.00	1.00

Notes. This table shows the distribution of attacks recorded by the civilian respondents. The data reflect the answer to the question "have your family members been victim of an armed attack at any time since 1990?" "If so, for each time, let's discuss the details." A mean of 5.62 means that, for the average respondent, there have been 5.62 attacks against at least one member of their family.

Table A12: Morality of Killing—Why is Killing Enemy Members Held as Morally Right?

	Infant		Pregnant Mother	
	Feels Good	Feels Bad	Feels Good	Feels Bad
Precentage	40.2%	59.8%	41.18%	58.82%
Mean Degree of Immorality [0-5]	2.54	4.18	2.81	4.23
Percentage Feels Guilty	18.33%	81.67%	17.24%	82.76%

Feels Good		Feels Bad
I would much rather massacre all Hutus in District		Hutus are immoral
It is good to turture them this way		They massacred our families, they deserve it
They came to our land to kill, we defended us		This follows Hutus stealing, raping, killing us
They had the same impacts in our population		This would be wrong, but given how they acted…
They did these massacres to our population		Nothing, eye for eye tooth for tooth (Bible)
Hutus did a lot of massacres in Congo		Because Hutus steal and rape our siblings
They deserve the massacres they did to us		Innocents are dying. Rwanda must take them
They made us suffer a lot		The Hutus saw an involuntary situation
They also harmed us in previous years		This is the fault of Tutsi-Hutu problem
They are bad people, massacred us		I feel very good, the Hutus aggressed us
They should leave our country		They are right, Hutus FDLR did the same
Who attacks with the sword will die with the sword (Bible)		Pushing them back so that we can be calm
Mai mais are just defending territorial integrity		Feel guilty, but it is our self-defense (x2)
When Hutus take us, they treat us the same way		This is a result of their own misconduct (x3)
Even better if it is executed by our mai mais		They are right; Hutus came as bandits
They are invaders		It's due to their own behavior
Hutus are Rwandan, they are the enemy		They also wronged our population
They are criminal people		Good or nothing (x5)
Mai mais do right, Hutus don't want to go home	*	It's a bad behavior (x2)
The Hutus did the same acts	*	Bad, bit bad, pity, nostalgy, very bad (x19)
Hutus FDLR do same aggression to our people	*	They really suffered
We reciprocate the actions	*	Bible says not to kill, it's not good
Their Rwandan Hutu brothers do the same acts	*	It's bad to massacre Hutus
I feel good because they are rebels	*	Bad in heart, they are humans like other (x2)
The FDLR are the root of massacres here in our country	*	I feel pity for them (x2)
The one who kills by shooting will die by shooting	*	Bad because it is immoral (x2)
They also massacred us much more than this	*	Very bad our brothers did this to her
Our mai mai brothers helped us push them back	*	This killing doesn't help us. We need cohesion
They are the ones who started massacring us	*	What mai mais did to her is bad
They shouldn't stay here, they should go back home	*	We are all Congolese siblings
They [the Hutu] are murderers	*	They did a big sin
They started massacring the Congolese	*	It hurts the heart really
The mai mais diminish our enemies	*	It's bad to massacre people
Very good (x5)	*	If the mai mai attacks them, it's wrong
Good (x7)	*	Bad (x10)

Notes. This table presents the answers to the module in which respondents are asked to express their opinion and potentially justify the violence perpetrated by the Congolese Mayi Mayi (in this case the group) on the enemy. To trigger them expressing their opinion and justify the violence when needed, respondents are shown the image of: a. an enemy member infant with wounds; b. a Rwandan mother with wounds, and are told that the wounds were caused by the Mayi Mayis. The respondents are then asked for both a and b to indicate whether they feel good or bad about it, to rate the degree of morality of the image, and to express whether they feel guilty. For b, they are also asked to articulate, in an open ended question, how they feel in their heart about it. The star indicates that the text to the right expresses moral condemnation of the violence. Sentences without the star are expressions of justification of the violence. The following sentence, extracted from a civilian respondent in a joiner's household, provides an elaborate explanation, pointing at feelings of betrayal and frustration: "*I feel good because these people had come as refugees. We had helped them with fields so they could cultivate while they were still in Congo, preparing for their return to Rwanda. We had even given them food supplies and cooking pots so they could settle in the country. We even gave them our daughters in marriage just so they would feel comfortable. But later, these people turned against us, starting to act as if they were the owners of all our possessions. They had now taken up their weapons, which they had crossed the border in 1994 after the genocide against the Tutsi. This is why, to this day, these people are destabilizing the eastern part of the country, the Democratic Republic of Congo. They rape, kill, and steal. So when I see such an image, I feel good because they are unconscionable, those people.*"

Table A13: Justifications of Group's Violence Against the Enemy Civilians

A. Moral Justification		
	Count	Percentage
Provides justification	168	55.1%
Morally condemns	137	44.9%
Total responses	305	100.0%
B. Type of Moral Justification (If Provided)		
	Count	Percentage
Retaliatory justification only	93	55.4%
Liberatory justification only	17	10.1%
Both retaliatory and liberatory justification	5	3.0%
Approval without specific justification	53	31.5%
Total justifications	168	100.0%

Notes. This table presents the results of classification analysis of open-ended text provided by the respondents in response to the following module. To elicit a a justification for holding moral convictions that support violence against enemy members, respondents were presented images of an enemy member civilian expressing suffering and carrying with scars (first infants, then a pregnant mother); surveyors explained that these scars were caused by lethal violence by one Congolese armed group. A prime was introduced before asking the respondents anything, containing moral condemnation of such violence in order to trigger respondents' moral justification. Respondents were then asked whether they felt good, the degree of immorality, whether they felt guilty and, in the case of the pregnant mother, they were offered an opportunity to articulate their feelings. The sample reflects that only half of the respondents, randomly selected, was offered the opportunity to articulate their justification, for conciseness. The sample of joiners, the randomly selected young male in their household, and the randomly selected young male in the randomly selected household of their village is included.

Table A14: Cognitive Processes that Rationalize Violence Against Enemy Members: Moral Disengagement

	Mean of Random Civilians, of which			
		Bridge Dilemma		
	All	Kills	Does Not Kill	Difference
A. Moral Justification (Index)	**0.71**	**0.76**	**0.63**	**0.130*** **
It's OK to be violent on Hutus to protect the community.	0.60	0.64	0.53	0.112**
It's OK to beat a Hutu who hurt Walikaleans.	0.80	0.85	0.71	0.139***
It's OK to be violent to a Hutu to restore population honor.	0.71	0.75	0.64	0.113**
It's OK to lie to a Hutu to get you out of trouble.	0.73	0.78	0.63	0.155***
B. Euphemistic Language (Index)	**0.65**	**0.72**	**0.54**	**0.182*** **
Kicking is the only language Hutus understand.	0.64	0.71	0.53	0.184***
Being violent with a Hutu, it's just giving them a lesson.	0.70	0.77	0.59	0.180***
Taking Hutu's beer without paying them is just borrowing.	0.62	0.69	0.50	0.194***
It's not bad to hit Hutus a bit from time to time.	0.65	0.72	0.55	0.170***
C. Advantageous Comparison (Index)	**0.62**	**0.67**	**0.53**	**0.136*** **
Damaging Hutu's things is OK given they are abusing us.	0.63	0.68	0.56	0.126**
Stealing a bit of cash from a Hutu is ok given they steal too.	0.67	0.74	0.56	0.183***
Torturing a Hutu is OK, because killing it would be worse.	0.56	0.59	0.51	0.079
Given illegal things they do, stealing from a Hutu is OK.	0.60	0.66	0.50	0.157***
D. Diffusion of Responsibility (Index)	**0.56**	**0.61**	**0.49**	**0.115*** **
A combatant cannot be blamed if his platoon kills H.c.	0.59	0.64	0.51	0.125**
A commander abusing H.c. cannot be blamed when others also do it.	0.57	0.63	0.49	0.135***
If a group hurts a H.c., it's unfair to blame an individual for it.	0.53	0.55	0.49	0.057
It's unfair to blame a combatant if his platoon hurt Hutus.	0.56	0.61	0.47	0.142***
E. Displacement of Responsibility (Index)	**0.61**	**0.67**	**0.52**	**0.145*** **
If combatants have bad conditions, can't be blamed for abusing H.c.	0.61	0.68	0.50	0.177***
If combatants are not trained, they can't be blamed for misconduct.	0.54	0.58	0.48	0.097*
Combatants can't be blamed for hurting H.c. when peers do it too.	0.62	0.67	0.55	0.119**
Combatants can't be blamed for torturing H.c. if peers pressured.	0.67	0.74	0.55	0.189***
F. Distorting Consequences (Index)	**0.64**	**0.70**	**0.55**	**0.153*** **
It's ok to lie a Hutu because he doesn't really feel pain.	0.64	0.71	0.53	0.182***
Hutu doesn't mind to be teased or confused.	0.60	0.64	0.54	0.100*
Teasing a Hutu does not really hurt them.	0.66	0.71	0.59	0.122***
Hassling a Hutu does not hurt anyone.	0.66	0.74	0.53	0.208***
G. Dehumanization (Index)	**0.64**	**0.70**	**0.52**	**0.180*** **
Hutus deserve to be treated as animals.	0.60	0.67	0.48	0.193***
It is normal to treat someone who behaved like a rat.	0.71	0.76	0.63	0.127***
A Hutu does not deserve to be treated as a human.	0.56	0.63	0.45	0.184***
Hutu must be treated violently, they lack feelings, can't feel pain.	0.67	0.75	0.54	0.217***
H. Attribution of Blame (Index)	**0.68**	**0.73**	**0.60**	**0.120*** **
If combatants hurt Hutu, it's commander's fault.	0.60	0.63	0.56	0.070
If Hutu leaves property unsupervised, it's their fault if stolen.	0.66	0.71	0.60	0.110**
Hutu who have been abused usually did things to deserve it.	0.74	0.81	0.64	0.166***
Combatants can't be blamed, if commanders pressure.	0.71	0.76	0.62	0.134***
Moral Disengagement Index	**0.64**	**0.69**	**0.55**	**0.145*** **

Notes. This table presents the answers to the Bandura (2011) scale of moral disengagement, as recorded among the randomly selected young males in the randomly selected households of the villages of the joiners. The table shows the mean fraction of respondents who respond "yes" to the corresponding question, the mean for respondents who had selected that the morally right action is to kill in the moral killing bridge dilemma, for those who had selected not kill in said dilemma, as well as the difference between these two.

Table A15: The Retaliatory and Motivated Reasoning Roots of Suppressed Aversion to Violence Against Enemy Members

	Empathic Concern toward Enemy			Affective Empathy toward Enemy			Cognitive Empathy toward Enemy		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
B.1 Joiners and Household Members									
Material Prosocial Pref. toward District	-0.191	-0.000	0.042	-0.274*	-0.093	-0.022	-0.554***	-0.454**	-0.435**
	(0.136)	(0.126)	(0.126)	(0.153)	(0.155)	(0.148)	(0.176)	(0.181)	(0.186)
Moral Disengagement (Bandura)		-8.063***	-7.503***		-7.610***	-6.821***		-4.251***	-3.830***
		(0.835)	(0.918)		(0.933)	(1.046)		(1.125)	(1.241)
Negative Reciprocity toward Enemy			-0.157**			-0.236**			-0.102
			(0.073)			(0.093)			(0.108)
Family Victimized by Enemy			0.286			0.665			-0.080
			(0.378)			(0.453)			(0.536)
Joiner ID FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean Dep. Var	3.38	3.38	3.38	3.97	3.97	3.97	5.94	5.94	5.94
R^2	0.49	0.63	0.64	0.49	0.60	0.62	0.47	0.50	0.50
Observations	420	420	420	420	420	420	420	420	420
B.2 Joiners and Random Villagers									
Material Prosocial Pref. toward District	0.188	0.174	0.188	0.298	0.284	0.301	-0.025	-0.033	-0.024
	(0.190)	(0.172)	(0.175)	(0.210)	(0.193)	(0.196)	(0.210)	(0.210)	(0.213)
Moral Disengagement (Bandura)		-5.711***	-5.251***		-5.493***	-4.895***		-3.194***	-2.872**
		(0.971)	(0.988)		(1.112)	(1.153)		(1.090)	(1.136)
Negative Reciprocity toward Enemy			-0.170**			-0.224**			-0.120
			(0.079)			(0.086)			(0.097)
Family Victimized by Enemy			0.326			0.610			0.283
			(0.440)			(0.494)			(0.564)
Joiner ID FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean Dep. Var	3.38	3.38	3.38	3.97	3.97	3.97	5.94	5.94	5.94
R^2	0.42	0.50	0.51	0.43	0.49	0.51	0.45	0.47	0.47
Observations	418	418	418	418	418	418	418	418	418

Notes. This table analyzes the role of motivated reasoning ("moral disengagement" (Bandura, 2011)) and of negative reciprocity against past victimization on the affective-motivational components of empathy toward enemy membersand on cognitive empathy toward enemy members. To do so, we estimate the following baseline regression in each panel, where the dependent variables differ across panels and where we sequentially add controls as indicated in each panel: $y_i = \alpha + \beta_1 NR_i + \beta_2 V_i + \beta_3 MD_i + \epsilon_i$ where NR_i, V_i, MD_i are the measures of negative reciprocity toward the Enemy, an indicator taking value 1 if the family of the respondent was previously victimized by the Enemy (as shown in Table A11, the median value is 1, which is true for 53% of respondents/observations), and Bandura (2011)'s Moral Disengagement scale about physically hurting enemy members. The dependent variable y_i are the following measures of empathy in a a scale [0-10]: affective concern toward Enemy (Columns 1-3), empathetic concern toward enemy (Columns (4-6), cognitive concern toward enemy (Columns 7-9);

Appendix C. Framework Extension: Prevailing Morality of Killing and the Correlation Between Social Preferences and Empathy toward the Enemy

Suppose that both social preferences toward the collective and empathy toward the enemy are jointly governed by a common "goodness" parameter, $\tilde{\lambda} \in \mathbb{R}$, as stipulated in the following assumption:

Assumption B *Morality and the Correlation*

$$\tilde{\lambda}^E = \tilde{\lambda}\left(1 - \tilde{s}^C m\right)$$

and

$$\tilde{s}^C = \tilde{\lambda},$$

where $m \in \mathbb{R}$ is a scale parameter measuring the intensity of the prevailing moral values about the moral righteousness of killing enemies.

Applying Assumption B, it follows that:

$$\frac{cov(\tilde{s}^C, \tilde{\lambda}^E)}{\partial m} < 0$$

and

$$cov(\tilde{s}^C, \tilde{\lambda}^E) < 0 iff. m > 1$$

Proof By Assumption B:

$$cov(\tilde{s}^C, \tilde{\lambda}^E) = cov(\tilde{\lambda}, \tilde{\lambda}(1 - m\tilde{s}^C,))$$

After minimal algebra, it follows that:

$$cov(\tilde{s}^C, \tilde{\lambda}^E) = var(\tilde{\lambda})\left(1 - m^2\right)$$

□

Appendix D. Ethics Appendix

To facilitate safe pursuit and execution of the study, the study team obtained authorization documents from the group's human resource and religious order agencies in 2015, as well as by the District's customary authorities, civil society and District administrator, the National army unit controlling the Province, the intelligence agency in the Province (Agence Nationale des Renseignements, in French). The authorization from group to conduct a study of the group was granted verbally in April 2015. The design was also approved by the University of Chicago's IRB, IRB19-1102-AM002. Additional steps were taken to ensure the safety of the surveyors. The surveyors were carefully recruited civilians who were trusted by the leadership of the group and the District. They carried authorization letters and symbols making their activity public: NGO T-shirts and service cards. They traveled by day and we provided them with cutting edge security information. Furthermore, six months prior to the launch of the data collection, we created a local ethics committee, comprised of members of of the data collection and data overseeing organizations, foreign researchers, as well as other Congolese researchers with the mandate to conduct a risk assessment of the design and to propose amendments: a. to reduce the risks to reasonable or otherwise as low as possible; b. mitigate the consequences of the risks; c. assess whether the amended design was below reasonable risk and ethical. The committee modified key components of the design (such as whether the surveyors should be publicly visible and how they ought to be introduced in the communities) and deemed the modified study below reasonable risk. We did not record any incident to individuals or organizations associated.

Finally, a legitimate concern might be that this agreement may feed into an armed militia group. Key features of the context shed light on this concern. First, unlike other militia groups, the group is not listed as a terrorist organization or similar sanctions list. The group is, however, recognized as a reserve army in Congolese law: between April and September 2023, a new Congolese law that recognized various militias, including the group as "Wazalendo" (meaning "patriots" in Lingala) was adopted by the Parliament and the government, recognizing them as a national service of support to the government army and materializing a long history of collaboration (Radio France International, 2023, Actualite.cd, 2023). Second, since colonial times, a huge number of actors have formed a part of, benefited from, and been aware of organizations similar to the group. Their participation and the popular support of their local society, rather than the view of foreign researchers, determines whether they perceive the group to be legitimate.

Appendix E. Measuring Emotional Arousal in Electro-Dermal Activity

A.. Introduction

For one-third of the survey duration, a subsample of 252 respondents wore E4 Empatica Bracelets to record Electrodermal Activity (EDA). This subsample consisted of 133 joiners, 66 household members, and 53 village members. This appendix describes the processing of raw EDA data collected with E4 Empatica wristbands and outlines the decomposition of the raw data. Given the variability in terminologies used in EDA analysis literature, this appendix adopts the terminology established by the authors of the LedaLab software (Benedek and Kaernbach, 2016). We then detail the construction of the dataset for analysis, leading to the final variables used.

Electrodermal Activity (EDA) refers to changes in the skin's electrical properties as a function of sweat gland activity. By applying a low constant voltage, variations in skin conductance (SC) can be measured non-invasively (Boucsein, Fowles, Grimnes, Ben-Shakhar, Roth, Dawson, and Filion, 2012). EDA can reflect either specific responses to stimuli (e.g., stressors) or non-specific responses.

The advent of wearable technology, such as the E4 Empatica Wristband, has enabled the collection of raw biomarker data in real-life settings. These wristbands are not only as reliable as laboratory-grade equipment but also facilitate long-term data collection in the field (Milstein and Gordon, 2020). EDA is typically decomposed into two main components: the tonic component, also known as skin conductance level (SCL), and the phasic component, known as skin conductance responses (SCRs), which are rapid changes driven by sympathetic nervous system activity.

In our analysis, we focus on measuring the Areasum of event-related SCRs (ER-SCRs) to assess the respondents' reactions to survey events, referred to as "primes." Although other parameters such as tonic and phasic changes can be measured, prior research has shown that the Areasum of ER-SCRs provides a clear and reliable measure of a respondent's response to brief, event-based stimuli. For a comprehensive overview, see Boucsein (2012).

B.. Setup and Primes

The Raw EDA is captured by the E4 Empatica Bracelet (Fig A36) which the respondent wears on their left wrist. To ensure that the respondent's response to the primes are not mixed with prior

Figure A36: E4 Empatica Wristwatch

Notes. Source: Reyes, Wozniak, Ham, and Zahabi (2023)

arousal from the survey activities, each respondent who wore the E4 Empatica Watch was asked to rest for 10 minutes before commencing the section of the survey.

Once the section of the survey began, the respondent was exposed to several primes consecutively after which the respondent answered questions. For this study we focus on the EDA response to three primes which are images. These images were shown on the SurveyCTO tablet by the enumerator. The three primes were 1) an image of person who has fallen from a tree, breaking his leg; 2) a person whose ear is being cut with a knife and 3) a photo of a hand or foot. The order of these three images, along with the survey questions, was randomized. Before exposure to each image, the enumerator loudly declared to the respondent the identity of the person in the image. The three identities possible are: a District Member, an enemy member Combattant or a Mayi-Mayi Fighter. Following each image, the respondent was asked

the questions that measure Empathetic Concern, Affective Empathy and Cognitive Empathy.

C.. Data Cleaning and Preparation

Raw Electrodermal Activity (EDA) data from the E4 Empatica Wristband is exported as CSV files containing two columns for each respondent over the session duration. The first column indicates the time since the session started, in seconds, while the second column provides the EDA response measurement in microSiemens.

To be analyzed using LedaLab software, the CSV files must be processed as TXT files with event markers. This is accomplished using Python code. Time stamps recorded by SurveyCTO were merged with the CSV file to identify event markers and time stamps. For each respondent and each prime, a new TXT file was generated by trimming the CSV data to the relevant prime period.

To create these prime-related segments, we retained 30 seconds of data before the prime and 10 seconds after. The moment a respondent was exposed to the prime was marked as an event in a new column. We report event-related skin conductance responses (ER-SCRs) from 1 to 10 seconds after exposure. While LedaLab authors recommend a starting analysis window of 1-4 seconds, researchers are given the discretion to choose an appropriate window. We extended this window to 10 seconds to account for any potential lag in response to the prime.

Once the data was trimmed for each respondent by each prime, each trimmed time period was checked for manual artifacts using the Python open-source algorithm EDA-Artifact-Detection-Script.py (Taylor, Jaques, Chen, Fedor, Sano, and Picard, 2015). Artifacts, such as sudden spikes caused by abrupt hand movements or wristband adjustments, were manually flagged and removed in the code. The final output was a cleaned TXT file in the required format for LedaLab analysis.

Subsequently, all TXT files were processed using LedaLab's batch processing function in MATLAB. This generated MAT files for each TXT file in MATLAB's data format. We then used Python to extract the Areasum measure for each response window from the MAT files and exported this data to Stata for final analysis. Based on LedaLab's recommendations, only SCRs with a minimum amplitude threshold of 0.01 S were retained. We calculated the Areasum for ER-SCRs with an onset occurring up to 1 second before the prime.

At the end of data preparation, we identified data corruption issues for 28 respondents (16 joiners, 8 household members, and 4 village members) due to either missing data or movement

artifacts. Observations for these respondents were excluded from the analysis. Additionally, some data was dropped for specific primes when no corresponding timestamps were found. Detailed reports on the data cleaning process can be found in the replication folder for each prime under prime/report/summary_data_cleaning.txt.

D.. Analysis

a. EDA Decomposition: A walkthrough example

This example illustrates the decomposition of the raw EDA signal for the respondent with record ID R.tcho.2.mp.0 during the prime event "person fallen from a tree" (see Figure A37). The red line indicates the moment the individual was exposed to the prime. The trimmed data include 30 seconds before the prime to provide LedaLab with a baseline for analysis. For this example, we concentrate on the 10-second window following the prime exposure.

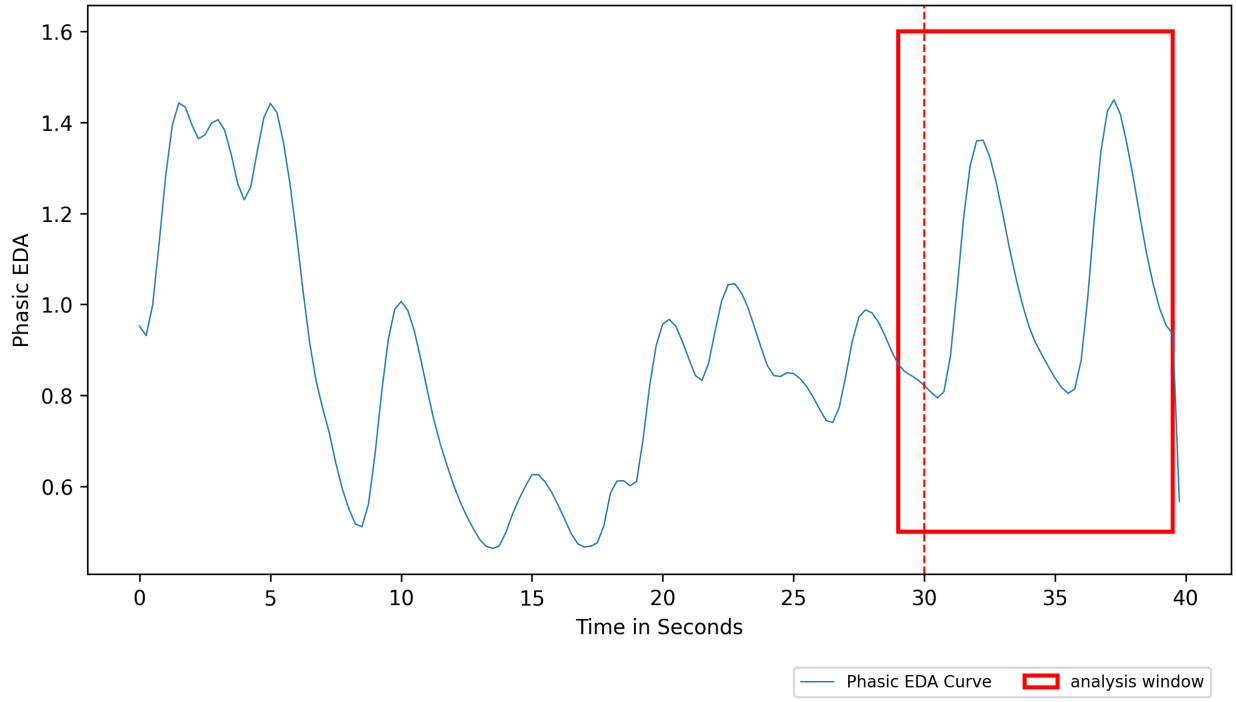


Figure A37: EDA Raw Data for R.tcho.2.mp.0 for the "person fallen from a tree" prime

For the extraction of measures we use the E4 Empatica Wristband recommended software LEdlab. Ledalab uses Continuous Decompostion Analysis (CDA) to decompose the Raw EDA into its onic or Skin Conductance Level (SCL) and the Phasic or Skin Conductance Response (SCR) components (Benedek and Kaernbach, 2010).

Figure A38 below shows the Tonic component. Tonic data signifies the general arousal for each individual. For the purpose of this analysis tonic data is not analysed as it is not a useful measure for short prime exposure but better used when a person is exposed to an event for several seconds or even minutes.

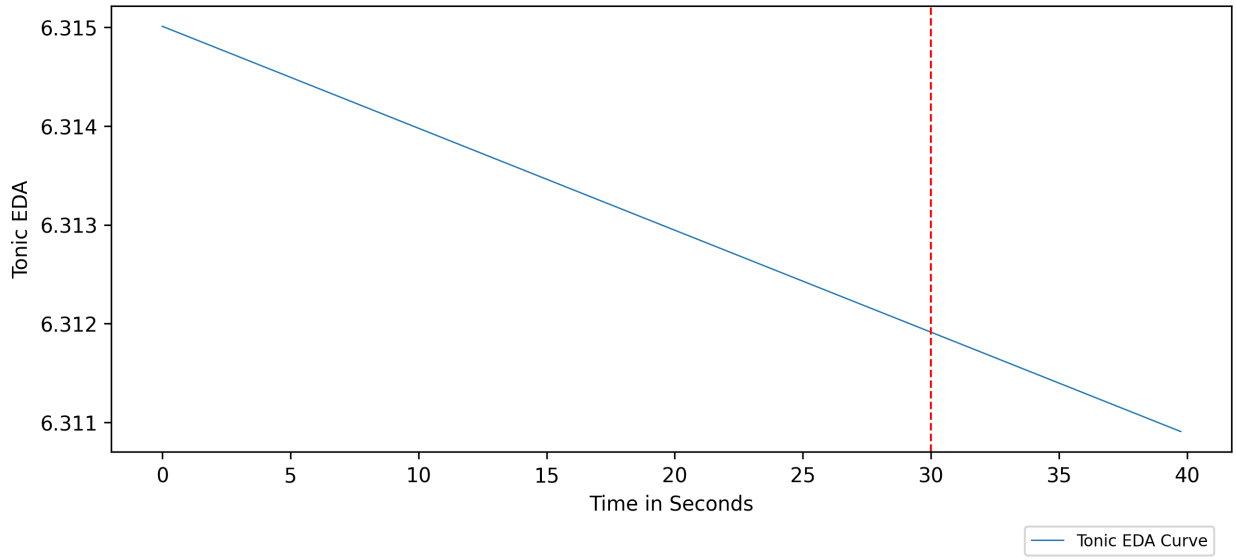


Figure A38: Tonic Component for Response Window of 10 seconds

Figure A39 shows the Phasic component which represents fast moving signals. Separating the Tonic component from the phasic component allows for specific event or prime analysis. This ensures that response observed is event related and not due to general background arousal. The response window is the time period that is studied to be Event Related. Figure A40 shows the locations which is the onset of the ER-SCR's for the respondent was.

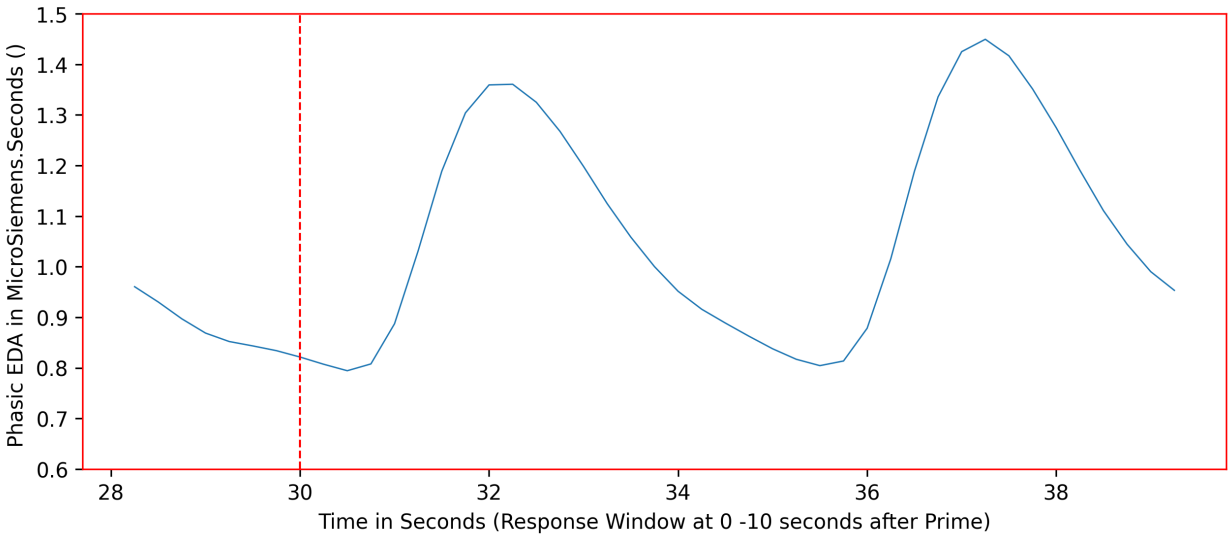


Figure A39: Phasic Component for Response Window of 10 seconds

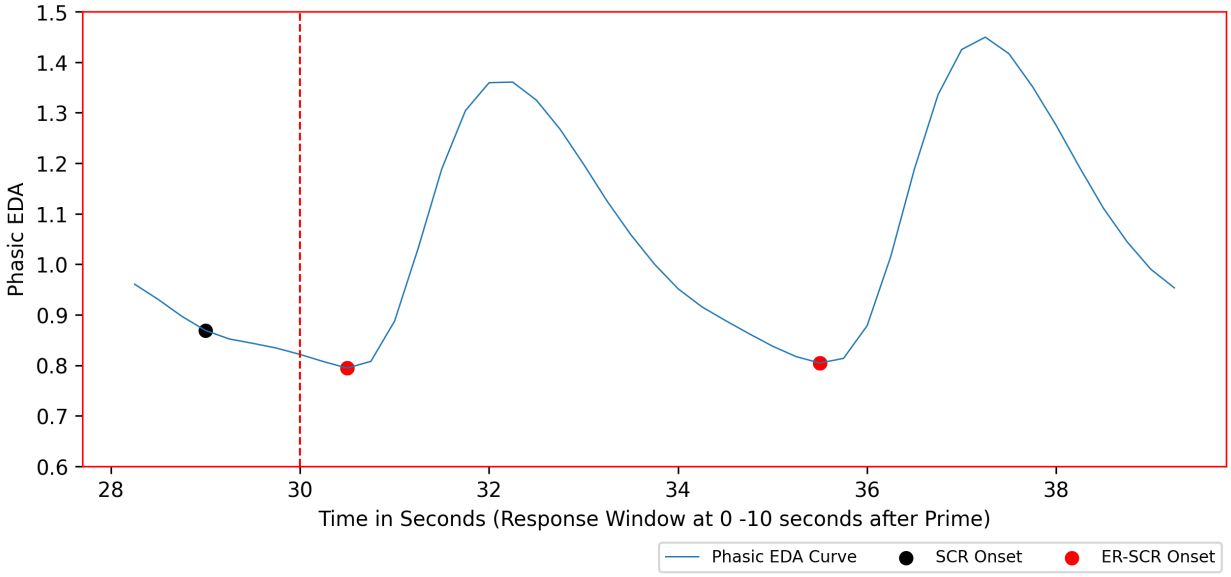


Figure A40: Onset location of Skin Conductance Responses: SCRs for the Prime.

Finally, Figure A41 indicates the measure for Areasumw which is used in our analysis.

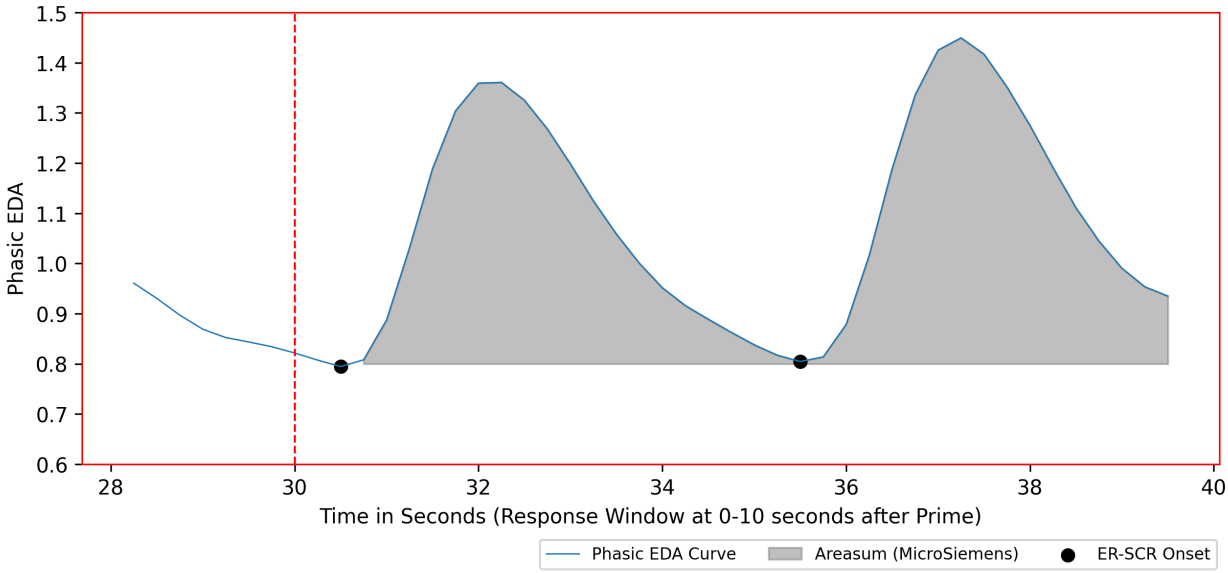


Figure A41: Phasic Activity with Areasum

b. Figure Outputs

The bar figures, Figures A34 and A35, compares the Areasum of the ER-SCR to the primes when told that the identity of the person in the photo is an enemy member or the District member for each respondent type. The coefplot figures in Figure A33 measures the coefficient of the Areasum of the ER-SCRs for the prime on the response to the three empathy questions. An additional result is added for all Empathy responses pooled.

References

Actualite.cd (2023): "Loi instituant la Réserve Armée de la Défense en RDC: pour Claudel Lubaya, ce texte va à l'encontre de la résolution de l'Assemblée nationale sur les groupes armés et légalise les milices au sein des FARDC," .

Benedek, M., and C. Kaernbach (2010): "A continuous measure of phasic electrodermal activity," *Journal of Neuroscience Methods*, 190(1), 80–91.

——— (2016): "LedaLab Documentation," http://www.ledalab.de/documentation.htm, Accessed: 2024-11-03.

Boucsein, W. (2012): *Electrodermal Activity*, The Springer series in behavioral psychophysiology and medicine. Springer US.

Boucsein, W., D. C. Fowles, S. Grimnes, G. Ben-Shakhar, W. T. Roth, M. E. Dawson, and D. L. Filion (2012): "Publication recommendations for electrodermal measurements," *Psychophysiology*, 49(8), 1017–1034, Epub 2012 Jun 8.

KST (2021): "Kivu Security Tracker," Discussion paper.

Milstein, N., and I. Gordon (2020): "Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States," *Frontiers in Behavioral Neuroscience*, 14, 148.

Radio France International (2023): "RDC: l'adoption d'un projet de loi sur un corps de réservistes au sein de l'armée se précise," .

Reyes, C., D. Wozniak, A. Ham, and M. Zahabi (2023): "Design and evaluation of an adaptive virtual reality training system," *Virtual Reality*, 27, 1–20.

Taylor, S., N. Jaques, W. Chen, S. Fedor, A. Sano, and R. Picard (2015): "Automatic Identification of Artifacts in Electrodermal Activity Data," in *Engineering in Medicine and Biology Conference*.