Rational Deterrence in Internal Armed Conflicts*

Gaétan Nandong

New York University

Abstract

I study deterrence in internal armed conflicts such as insurgencies, where the identification problem is a defining feature. A repressive agent who can be overzealous must execute a repressive order with limited knowledge of dissidents. Dissenters always reveal valuable information that exacerbates the repressive response. The threat of repression often leads to overreactions among civilians as they assess its marginal impact on the severity of repression. Civilians anticipate the potential for security overreach in enforcement and engage in preemptive resistance despite the perceived risks. Technological advancements support these findings, as I demonstrate that micro-level rather than macro-level inference offers robustness. If insurgent attacks pose significant risks to security forces, agents who would typically exercise restraint are screened out.

Keywords— Repression, Credible threats, Identification problem, Overreaction.

*I would like to thank Sandeep Baliga, Peter Bils, Lisa Blaydes, Peter Buisseret, Steve Callander; Alessandra Casella; Tiberiu Dragu; Timothy Frye, German Gieczewski, Federica Izzo, Kimuli Kasara, Navin Kartic, Gleason Judd, Chris Li, Zhaotian Luo, Cesar Martinelli, Andrea Prat, Carlo Prato, Kris Ramsay, Arturas Rozenas, Mehdi Shadmehr, Alastair Smith, Konstantin Sonin, Jessica Sun, Leonard Wantchekon, and the participants in several seminars and conferences.

> [...] we will not win based on the number of Taliban we kill, but instead on our ability to separate insurgents from the center of gravity—the people. That means we must select and protect the population from coercion and violence—and operate in a manner which will win their support. [...]
>
> I understand that the carefully controlled and disciplined employment of force will entail risks to our troops [...]. But excessive use of force resulting in an alienated population will produce far greater risks. We must understand this reality at every level in our force.
>
> *Retired United States Army General Stanley A. McChrystal, July 2, 2009, Kabul, Afghanistan.*[1]

In conflicts such as rebellions and insurgencies, government forces often resort to repression to deter undesirable behavior (Davenport 2007; Ritter and Conrad 2016). Although repression often cultivates a climate of fear and pessimism (Nandong 2025; Young 2019; Pearlman 2013), traditional concerns about the failure of deterrence are fundamentally rooted in psychological explanations. An expanding body of literature indicates that government repression can elicit relational and emotional responses, leading to backlash even in the presence of perceived risks (Gurr 1970; Fransisco 1995; Rasler 1996; Blaydes 2018; Nugent 2020; Lebas and Young 2024; Rosenfeld 2025). The literature often fails to predict which psychological response state repression will provoke (e.g., fear or anger). This limitation stresses the need to further explore the factors that drive deterrence in political contentions. The paper addresses this gap by focusing on civil conflicts such as insurgencies and civil wars, characterized by the identification problem.

The identification problem in insurgencies refers to the challenges authorities encounter in distinguishing insurgents from the general population (Berman, Shapiro, and Felter 2011; Trinquier 2006; U.S.Army 2007; Kalyvas 2006). While patrolling the streets of Iraq, a soldier from the American troops succinctly described this challenge, stating that "it's hard to separate the good from the bad" (Filkins 2005). Similarly, during the Soviet invasion of Afghanistan, adversaries were referred to as *dukhi*, translating to "ghost" (Baker 2002). This

1. General Stanley McChrystal, *Tactical Directive* (Kabul: International Security Assistance Force, July 6, 2009), http://www.nato.int/isaf/docu/official_texts/Tactical_Directive_090706.pdf.

difficulty is partly due to the nature of irregular warfare, where dissidents often deliberately blend into the civilian population (Kalyvas 2006; Lewis 2020). In the Vietnam War, the Viet Cong operated as an insurgency against the South Vietnamese government and the United States, with fighters who were farmers, shopkeepers, and laborers during the day, seamlessly integrating into the local community. In this context, the challenge of separating dissidents from civilians presents an agency problem of unobserved actions that has thus far been overlooked in the theoretical literature on collective action for political change.

In this article, I present a theoretical framework that deepens our understanding of the various factors undermining deterrence in political contentions. It is widely recognized that the repressive apparatus plays a crucial role in managing opposition challenges (Greitens 2016; Svolik 2012; Stavro and Welch 2024). The implementation of repressive measures falls under the jurisdiction of security agents, with the police, military, and security forces often serving as the primary institutions responsible for identifying and repressing dissenters (Liu and Sullivan 2021; Sullivan 2016; Davenport 2007). Repressive measures typically manifest as threats aimed at deterring aggressors. Security agents patrol communities, detain and interrogate suspected dissidents, and process information to ultimately carry out the threat of repression. The phenomenon of overcompliance with repressive measures frequently characterizes overly zealous government forces, just as a lack of compliance can also be observed. Consequently, the security apparatus often consists of repressive agents who exhibit overzealousness, act out of self-interest, and employ excessive force (Tyson 2018; Egorov and Sonin 2011; Curtice and Brandon 2021).

The paper emphasizes the role of repressive agents and the impact of limited information on the dynamics of repression and dissent by analyzing a standard game of collective action with private and independent types. In the framework, a discrete number of citizens choose either to dissent or to abstain in the hope of achieving political change. In response, a strategic but imperfectly informed repressive agent ("Defender") decides on a course of repression.

Defender prefers to target dissidents and incurs a cost if civilians are subjected to repression. Defender's decision to retaliate is influenced by the signals generated by the dissidents, the repressive measures, and the value of retaliation. The repressive measure ("threat of repression") is common knowledge among all players and is assumed to be credible. Defender's value of retaliation is considered private information. The value of retaliation is Defender's type, with an overzealous type having a high value of retaliation.

The observed level of dissent can send strong signals about the degree of activism, raising Defender's belief about an individual involvement. As a result, dissidents tend to send strong signals and face more retaliation compared to civilians. This argument suggests that, despite the indiscriminate nature of repression, its severity is a function of participation.

Although exposed to repression less frequently, civilians tend to overreact when they anticipate the use of excessive force. They overestimate the marginal impact of a threat of repression executed by an overzealous type. While such an overreaction reduces the marginal risk of joining an insurgency, whether civilians feel compelled to act depends on the perceived likelihood of the campaign's success. An overreaction can spark dissent when there are hopes for the movement's success, especially when the threat of repression is low or moderate.

More importantly, the current mechanism operates through a channel distinct from the channel highlighted by the psychological literature on repression. It is often unclear how repression, which fosters fear and pessimism—resulting in quiescence—can simultaneously ignite indignation, identity formation, and anger, leading to backlash. Rather than being psychological, the mechanism presented in this paper features a Bayesian player that utilizes signals generated at the insurgent group level to enhance its knowledge of dissidents. I demonstrate that relative to abstention, an act of dissent sends strong signals, exacerbating repression on dissidents. However, those who abstain tend to overreact to the threat of repression executed by likely overzealous types. In an extension, I explore an alternative signal structure in which signals are generated at the individual level rather than the group

level. I find that this information structure reinforces the main result.

Taken together, the results suggest that exercising restraint during a counterinsurgency can be a winning strategy. Excessive force and unnecessary violence can weaken deterrence in environments characterized by the identification problem. This finding aligns with the tactical directive issued by retired US Army General McChrystal on the critical impact of courageous restraint among US troops. Senior leadership within the United States military has emphasized the necessity of protecting civilian life and minimizing civilian casualties to comply with international and humanitarian laws. In this regard, the findings presented in this paper underscore an additional benefit that such a tactical directive can provide for a counterinsurgent operation.

The article presents several theoretical, empirical, and political contributions to the scholarship on political contention and conflict. First, I propose a rational explanation for the paradox of repression. The literature on the effects of repression is mixed, with most accounts of deterrence failure being empirical. Second, I address the controversy surrounding the 2009 tactical directive issued by the senior leadership of the U.S. Army.[2] Third, the model engages with the literature on leadership in rebellions and insurgencies. I find that through their dissent actions, aggressive actors sometimes compromise civilians, as these actors reveal valuable information for defenders, initiating a mechanism that encourages civilians to join the insurgency. Fourth, I propose a novel theoretical framework of collective action that incorporates the identification problem. This model can be applied to studies in which a group of players collaborates to achieve a collective goal, while a third player with limited knowledge of individual contributions relies on aggregate statistics of efforts to assess individual performance. Lastly, this article contributes to the literature on security studies by addressing police brutality and the excessive use of violence in the enforcement of social control.

2. https://www.nbcnews.com/id/wbna35409150

The next section reviews the literature. Following that, I present the model and analysis. I examine several extensions in which I modify the information structure and relax the assumption about the impact of the level of dissent. The final section concludes.

Related Literature

Several scholars have found a positive association between political dissent and repression in various conflicts with an identification problem. Rebellion is often viewed as a form of protection when repression is indiscriminate (Kalyvas 2006, Lyall 2009). However, this argument rests on the view that neutrality is more costly than participation. Furthermore, in a seminal book on Iraq under Saddam Hussein, Blaydes 2018 argues that this form of repression reinforces group identities, ultimately increasing political dissent. In her framework, inadequate information compels the regime to rely on ethnicity to improve its limited knowledge of dissidents. Given this mode of inference, expectations of repression increase political dissent by reinforcing group ties (Blaydes 2018, pg 47). This argument is grounded in a long tradition in political science, sociology, and economics that considers ethnicity as the relevant dimension to enhance intelligence in domestic conflict (Greitens 2016, Bates 1983, Cooper 1979, Bayly 1996). In the model I consider, the severity of repression is a function of participation and ethnicity is not the relevant dimension for information updating.

My paper also speaks to the literature on backlash. A backlash effect occurs when an increase in repression is followed by a rise in dissent. The mechanisms through which dissent responds to repression are often associated with emotions such as anger, frustration, perceptions of unfairness, or the activation of identity (Rasler 1996; Lichbach 1987; Opp and Roehl 1990; Gibilisco 2021; Pop-Eleches and Way 2023; Lebas and Young 2024). For example, members of a community may perceive state repression as unjust, which can lead to a heightened willingness to engage in risky actions. This psychological perspective on the

failure of deterrence is often confusing, as repression also tends to provoke fear and pessimism (Pearlman 2013; Young 2019; Nandong 2025).

The impact of repressive measures on the frequency of their application and the likelihood of dissent has been previously examined within a game-theoretical framework (e.g. Shadmehr and Bernhardt 2011, Bueno De Mesquita et al. 2024). Most studies assume, either implicitly or explicitly, that repression is implemented under conditions of perfect information. First, Bueno De Mesquita et al. 2024 explores a framework in which a threat of repression may lack credibility, as its implementation can be challenging for the repressive a. Second, Shadmehr and Bernhardt 2011 investigates a model of collective action where players are uncertain about the value of a successful protest. Their findings indicate a positive correlation between the likelihood of repression and the probability of dissent; specifically, higher levels of punishment (repressive measures) diminish both the likelihood of repression and the incentive to dissent.[3]

While this paper studies high-risk activism, Nandong 2025 examines low-visibility dissent such as leaflets distribution and petition signing.[4] Most people do not get involved in high-risk dissent. The starting point is often the low-risk dissent in which someone distributes leaflets and sign petitions. There are also reasons to believe that involvement in low-visibility dissent can explain engagement in high-risk dissent, a question that can be the subject of future work. In addition, the paper examines the impact of state surveillance, measured by the noise in the dissent signal. The two theoretical frameworks are therefore clearly different.

The theoretical literature also examines frameworks in which the information environment influences the prospects for effective deterrence (Lagunoff 2024; Baliga, Bueno De Mesquita, and Wolitzky 2020; Chassang and Padró I Miquel 2010). Lagunoff 2024 analyzes

3. The article by Shadmehr 2014 find a U-shaped relationship between income inequality (a measure of grievances) and repression. In contrast, my study focuses on the influence of repressive measures on the likelihood of dissent.

4. Scholars have stressed to importance of making the distinction between high and low-risk dissent (McAdam 1986).

a dynamic model in which, at the beginning of each period, a government commits to a compliance rule while imperfectly observing acts of resistance. In Lagunoff's framework, the level of resistance within the population is not a statistic upon which the government relies, nor are citizen incentives driven by coordination motives. Furthermore, Baliga, Bueno De Mesquita, and Wolitzky 2020 examines a framework in which the government faces one attack at a time, whereas multiple attackers can target the government in my framework. Defender refines its information by observing the scale of the attack, implying that there are direct externalities as a member's dissenting action can harm others. Finally, Chassang and Padró I Miquel 2010 analyzes a model of international conflict under strategic risks, where players do not share the same information about their environment. They conclude that standard predictions must be considerably nuanced once strategic risk is considered.

Model

The theoretical framework is an extension of the classic model of collective action described in Figure 1. I will first present the canonical framework. I will then show how the canonical framework can be extended to incorporate the identification problem.

In the canonical framework, there is a community with a discrete number $n \geq 2$ of citizens (she). Each citizen has a choice between joining an insurgency campaign ("dissent") or staying a civilian ("abstain"). Citizens who join the insurgency are dissidents. Participants in a successful campaign receive 1, while civilians receive γ. I assume that $0 \leq \gamma \leq 1$, which indicates that the benefit associated with success may be either selective or non-selective.

Each citizen i has a private cost c_i, independently distributed on $[\underline{c}, \overline{c}]$. The costs are distributed according to the same cumulative distribution function $G(.)$ and a probability density function $g(.)$, continuous on $[\underline{c}, \overline{c}]$. The type space is assumed to be rich enough: $G(0) > 0$ and $G(1) < 1$. Finally, the campaign succeeds with probability $p(s, \tau)$, where s

	Success	Failure
dissent	$1 - c_i$	$-c_i$
Abstain	γ	0

Figure 1: A citizen payoff matrix when Defender has perfect information.
Note: The figure shows the payoff that citizen i receives at the dissent stage, conditional on the outcome of the campaign. When Defender has perfect information, it only targets dissidents.

is the number of dissidents and $\tau \in [\underline{\tau}, \overline{\tau}] \subset [0, 1]$ is the threat of repression assumed to be common knowledge. If the threat consists of sending troops to patrol the street, cordon off an area, or establish checkpoints, this can restrict mobility and ultimately compromise success. The likelihood of success takes the form

$$p(s, \tau) = \frac{\pi(s)}{d(\tau)},$$

where s is the size of attacks, $\pi(s) \in [0, 1]$, and is weakly increasing.[5] Furthermore, the function $d(.)$ is positive and increasing, with $d'(\underline{\tau}) = 0$ and $d'(\overline{\tau}) = \infty$. I also impose the regularity condition that d'/d is strictly increasing; i.e $d(.)$ is log-convex.[6]

I interpret the game depicted in Figure 1 as a reduced form of a three-stage extensive form game. In the first stage, citizens choose whether to dissent; in the second stage, Defender (it) decides whether to retaliate; and in the third stage, the outcomes and payoffs are realized. The classic framework assumes that Defender has perfect information and prefers to retaliate solely against dissidents. Consequently, the second and third stages are straightforward, as Defender retaliates against a dissident with a probability of 1. Under the assumption of perfect information, the analysis and results are standard, as outlined in the following remark.

5. For example, $\pi(s)$ can take the form $\mathbb{1}_{\{s \geq \theta\}}$, where θ is the common knowledge success threshold. This is the most used probability of success in the literature. In this case, π is a step function.

6. The following function satisfies these conditions: $d(\tau) = exp\left[\frac{(\tau - \underline{\tau})^2}{\overline{\tau} - \tau}\right]$ for $\tau \in [\underline{\tau}, \overline{\tau}]$.

Remark 1 *Suppose Defender has perfect information about the dissent action.*

1. *If the distribution $g(.)$ is flat enough (i.e $g(x) < \xi_0$, for some $\xi_0 > 0$) the game depicted in Figure 1 has a unique equilibrium.*[7]

2. *The equilibrium likelihood of dissenting is decreasing in the threat of repression τ.*

Remark 1 describes the conditions for equilibrium uniqueness in the canonical framework. The remark also succinctly states the standard deterrence logic. Deterrence is successful because the credible threat of repression deters dissident activities. The result follows from the negative impact of τ on the likelihood of success. As the threat of repression increases, citizens self-censor, anticipating a detrimental effect on the likelihood of success. The proof of Remark 1 is contained in the Appendix.

In most situations, Defender cannot distinguish between active dissidents and non-combatant civilians. Hence, Figure 1 is very restrictive as it does not account for Defender's inability to observe whether an individual is a dissident. My framework assumes that Defender observes a private signal that depends on the number of dissident attacks in other to improve its knowledge of dissidents. That is, Defender observes the signal

$$s = a_1 + a_2 + \cdots + a_n = \sum_{i=1}^{n} a_i,$$

where "dissent" corresponds to $a_i = 1$, and "abstain" represents $a_i = 0$.

In the second stage, and for each $i \in \{0, 1, 2, \cdots, n\}$, Defender chooses to retaliate against citizen i or not. Defender's payoff is given as follows:

7. An example of a flat distribution is the Uniform distribution.

Retaliation	$\tau \times v - 1 \times \mathbb{1}\{\text{i is not a dissident}\}$
No Retaliation	0.

In addition to establishing checkpoints, the threat of repression (τ) also represents the intent to inflict harm on a suspected dissident. This threat is carried out by the security agent and is assumed to be credible. The credibility of the threat is established when it is perceived as low-cost, thereby enhancing its potential application.

The value of retaliating, $v \in [0, 1]$, is Defender's type: "the zealousness" of Defender. Citizens believe that v is distributed according to a cumulative distribution function $H(.)$ and a density $h(.)$.[8] Additionally, the baseline model assumes that the decision to retaliate depends on the level of dissent only through belief updating. In an extension below, I explore a scenario in which the level of dissent can impose serious injuries to Defender.

In this baseline model, there is a cost of 1 if retaliation is targeted toward a civilian. Hence, Defender cares more about targeting dissidents. Finally, I assume that $\frac{1}{\tau}\left(\frac{n-1}{n}\right) < 1$. That is, the cost of targeting civilians is high relative to the maximum threat and the population size. The timing of the game is as follows.

Timing

1. Citizens observe their private cost. They (simultaneously) decide whether to dissent or abstain.

2. Defender cannot observe the dissent decision, but instead only observes a noisy signal

8. That the value of retaliation is private information is not necessary for the results, but is important to ensure continuity and tractability. Having v common knowledge can preserve the main result, at the cost of tractability as continuity is lost (see Proposition OA.1 in the Online Appendix).

of it. Defender observes $s = \sum_{i=1}^{n} a_i$, and chooses to retaliate or not.

3. The insurgency succeeds with a probability of $p(s, \tau)$, and payoffs are realized. The equilibrium concept is Perfect Bayesian equilibrium.

I look for a symmetric equilibrium (α^*, R, q), where α^* is the likelihood of dissent. R is the probability of retaliation, and $q : \{0, 1, \cdots, n\} \mapsto [0, 1]$ is the belief. Given signal s,

$$q(s) = \Pr[\text{i is a dissident} \mid s].$$

I now highlight the key tensions in the model and provide a brief comment on the model assumptions.

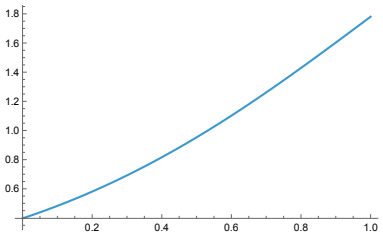
Discussion

Key tensions: Explaining the model.

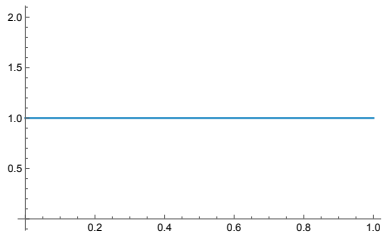
The model illustrates a framework where a repressive agent (Defender) with limited knowledge of dissidents must carry out a repressive order. Defender's retaliation is influenced by the intelligence gathered regarding troublemakers (s), the repressive order (τ), a threat publicly communicated by military leaders to caution potential dissidents, and the value associated with adhering to orders (v).

A number of features emerge from the model setup. First, the signal generated by citizen i, $s = a_i + \sum_{j \neq i} a_j$, depends on the actions of other citizens. Thus, others' actions have an impact on a member's chance of being targeted. Second, the discrete number of citizens implies that a dissident will generate a higher signal compared to a civilian. Hence, relative to a civilian, a dissident may expect a different repressive response.

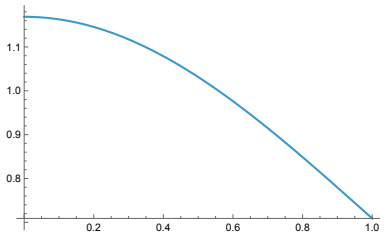
The retaliation decision depends on the threat of repression and the dissent signal. More importantly, it also varies depending on the value of retaliation. If $v = 0$, then Defender



(a) **High prevalence**.
The value of retaliation is distributed according to $\mathcal{N}(1,1)\Big|_{[0,1]}$.



(b) **Moderate prevalence**.
The value of retaliation is distributed according to $\mathcal{U}[0,1]$.



(c) **Low prevalence**.
The value of retaliation is distributed according to $\mathcal{N}(0,1)\Big|_{[0,1]}$.

Figure 2: Characteristics of the repressive apparatus.

Note: The plot outlines the probability density function (PDF) of the CDF $H(.)$, representing the value of retaliation. Since $v \in [0,1]$, the distribution is truncated on $[0,1]$. The distribution $\mathcal{N}(1,1)$ is the normal distribution with mean 1 and variance 1. The distribution $\mathcal{N}(0,1)$ is the standard normal distribution. The distribution $\mathcal{U}[0,1]$ is the uniform distribution on $[0,1]$. Each distribution describes the fraction of overzealous types within the repressive apparatus. Figure 2a has a high prevalence (strictly convex CDF), Figure 2b has a moderate prevalence (weakly convex CDF), while Figure 2c (strictly concave CDF) has a low prevalence. Hence, the shape of the distribution $H(.)$ captures the prevalence of overzealous types.

observes enormous restraint in applying repression. Since a mistake is costly, Defender will only retaliate if the suspect is guilty beyond a reasonable doubt based on the information revealed by the signal. However, a high value of v indicates a strong tendency to use excessive force, including the application of repression even when the evidence does not completely eliminate doubts.

Although the threat of repression (τ) is common knowledge, citizens also know that the fulfillment of the threat depends on Defender's information (s) and its type (v). Citizens only know the distribution $H(.)$ that represents Defender's value of retaliation. The shape

of the distribution can characterize the prevalence of overzealous types within a repressive apparatus. The picture outlined in Figure 2 represents the probability density function (PDF), truncated on $[0, 1]$, of the value of retaliation v. As shown in Figure 2a, a convex CDF $H(.)$ (normal distribution with mean 1 and variance 1 truncated on $[0, 1]$) represents a repressive apparatus with a high prevalence of overzealous types: the density is increasing, so higher values are more likely. For example, Figure 2a represents the repressive apparatus in Iraq under Saddam Hussein. In contrast to Figure 2b where the prevalence of overzealous types is moderate (the uniform distribution on $[0, 1]$), Figure 2c has a low prevalence of overzealous types (the standard normal distribution truncated on $[0, 1]$), since higher values are less likely. Thus, in contrast to Figure 2a and 2b, a type within the repressive apparatus represented by Figure 2c is likely to observe control and discipline in the employment of force as required by the senior leadership of the US Army in Afghanistan. Thus, when Defender belongs to a repressive apparatus characterized by Figure 2a and 2b, citizens anticipate security overreach in enforcement. The aim of the paper is to demonstrate that when citizens (both civilians and dissidents) anticipate that Defender "is likely to" exercise excessive force to enforce the threat (τ), civilians surprisingly overreact to the threat, contributing to a failure of deterrence.

Assumptions

Although this assumption is often unrealistic, standard models of collective action assume that repressive agents make a binary decision between repressing and conceding, treating these options as substitutes. In many instances, however, successful movements face repression and individuals may be injured or even die in the pursuit of a successful resistance campaign. This last feature is in line with my model.

Finally, the payoff matrix in Figure 1 illustrates the concept of "pleasure in agency" (if $\gamma < 1$), which captures the selective psychological benefit derived from engaging in a successful rebellion (Wood, 2003). Furthermore, all types $c_i < 0$ receive an expressive benefit

from participation. These aggressive types join the movement to achieve a cathartic release of their grievances (Correa, Nandong, and Shadmehr 2025). In contrast, all types $c_i > 0$ assess the costs, benefits, and likelihood of the movement's success (Tilly 1978).

Analysis

The analysis begins by examining how beliefs about dissidents influence retaliation. Given the observed level of dissent $s \in \{0, 1, 2, \cdots, n\}$, what is Defender's posterior probability that citizen i is a dissident? Assume that the probability of dissent is α. Defender does not directly observe the decision to dissent, but will ultimately learn α in equilibrium.

When $s = 0$, it is straightforward that $q(0) = 0$. If there is no attack, a rational Bayesian infers that there is no dissident. Similarly, if $s = n$, then $q(n) = 1$.

Now suppose $s = 1$. By Bayesian updating, Defender believes that

$$q(1) = \frac{\Pr[s = 1 \mid i \text{ is a dissident}] \times \Pr[i \text{ is a dissident}]}{\sum_{k=1}^{n} \Pr[s = 1 \mid k \text{ is a dissident}] \times \Pr[k \text{ is a dissident}]} = \frac{\binom{n-1}{0}\alpha(1-\alpha)^{n-1}}{\binom{n}{1}\alpha(1-\alpha)^{n-1}} = \frac{1}{n}.$$

Since a given citizen chooses to dissent with probability α, any other citizen does not with probability $1 - \alpha$. Moreover, the probability of observing a signal $s = 1$ when i is already a dissident is given by $\binom{n-1}{0}(1-\alpha)^{n-1}$. I now state the next Lemma to describe Defender's posterior belief after any signal s. The proof is contained in the Appendix.

Lemma 1 *Consider a dissent signal $s \in \{0, 1, 2, \cdots, n\}$. Defender's belief in equilibrium is given by $q(s) = \frac{s}{n}$.*

Thus, in any symmetric equilibrium, Defender's posterior belief is a well-behaved function of the magnitude of the attack. This suggests that an increase in the number of attacks

elevates Defender's belief that an individual within the community is a dissident. Consequently, in an attempt to resolve its Bayesian inference problem, Defender may erroneously assign the dissident label to a civilian.

The model elucidates a novel logic of indiscriminate repression. Existing literature posits that, in the presence of imprecise information, government forces tend to target individuals based on their ethnic background, or spoken language (see Kalyvas (2006), Chap. 6 for a comprehensive review). In contrast, the functional form of Defender's posterior belief presents an alternative mechanism grounded in Bayesian updating.[9]

Regarding retaliation, the decision depends on the posterior belief, the threat of repression, and the value of retaliation. If Defender retaliates against citizen i, it receives

$$q(s) \times \tau \times v + (1 - q(s))(\tau \times v - 1).$$

With probability $q(s)$, an arrested suspect is a dissident, while with probability $1 - q(s)$ the arrested suspect is a civilian. Defender pays a cost if it retaliates against a civilian. Hence, after observing a signal s, it is optimal to retaliate against citizen i if and only if the expected payoff is positive:

$$v \geq \frac{(1 - q(s))}{\tau} = \frac{1}{\tau}\left(1 - \frac{s}{n}\right),$$

for $s \geq 1$. Defender retaliates if its zealousness exceeds the ratio between the probability of a mistake and the threat of repression. Interestingly, overzealous types do not need clear and convincing evidence to retaliate. For example, a type v close to 1, retaliates even if a suspect generates a low signal. However, less zealous types need substantial evidence before executing the threat. For example, when v is small, the dissent signal must be high for Defender to retaliate. I can then state Defender's best-response.

9. Refer also to Rozenas (2020) and Brandsch (2020), which emphasize rationality as a factor in the prevalence of indiscriminate repression within social movements.

Proposition 1 *In equilibrium, any type of Defender believes a citizen is a dissident with probability* $q(s) = \frac{s}{n}$. *Moreover, a type* v *retaliates if and only if* $v \geq \frac{1}{\tau}\left(1 - \frac{s}{n}\right)$.

Defender retaliates if its type exceeds the probability of making a mistake. If v is distributed according to a distribution $H(.)$, *the severity* of repression is measured by

$$\Pr\left[v \geq \frac{1}{\tau}\left(1 - \frac{s}{n}\right)\right] = 1 - H\left(\frac{1}{\tau}\left(1 - \frac{s}{n}\right)\right) \equiv R(s, \tau),$$

for $s \geq 1$. It is cleat that $R(0, \tau) = 0$, since $q(0) = 0$ and $\tau v - 1 < 0$.

I denote $R(s, \tau)$ the severity of repression. It is the proportion of types who retaliate conditional on the dissent signal s and the threat of repression τ.

The next Lemma states that the number of attacks increases the severity of repression a citizen could face.

Lemma 2 *Dissenting reveals more information and increases the severity of repression. That is,* $R(s, \tau) < R(s+1, \tau)$ *for* $s \leq n - 1$.

A high signal raises Defender's posterior belief, leading to a more severe repression. Additionally, it is important to note that the severity of repression increases with its threat, which supports the assumption of credible threats.

Moving backward, I analyze the decision to dissent. Given that others participate with probability α, a dissident has an expected payoff given by

$$\frac{1}{d(\tau)}\left[\sum_{s=0}^{n-1} \Pr[S^i = s] \times \pi(s+1)\right] - c_i \underbrace{\left[\sum_{s=0}^{n-1} \Pr[S^i = s] \times R(s+1, \tau)\right]}_{expected\ severity\ for\ dissidents}, \tag{1}$$

where S^i is the random variable equal to the number of those who choose to dissent, excluding i; and $R(s, \tau)$ is the severity of repression, conditional on a dissent signal s and the threat of repression τ.

The likelihood of success depends on others' dissent decision. Given that types are independent, this decision is independent across individuals. Furthermore, the expected severity of repression depends on the aggregate number of dissidents, as Defender imperfectly observes whether the citizen is a dissident, and improves its knowledge using the dissent signal.

For a civilian, the expected payoff reads as follows

$$\frac{\gamma}{d(\tau)}\left[\sum_{s=0}^{n-1}\Pr[S^i=s]\times\pi(s)\right]-c_i\underbrace{\left[\sum_{s=0}^{n-1}\Pr[S^i=s]\times R(s,\tau)\right]}_{expected\ severity\ for\ civilians} \tag{2}$$

It is important to note that a dissident expects a more severe repression, compared to a civilian

$$\sum_{s=0}^{n-1}\Pr[S^i=s]\times R(s+1,\tau)>\sum_{s=0}^{n-1}\Pr[S^i=s]\times R(s,\tau). \tag{3}$$

This is because dissidents send strong signals. By Lemma 2 higher signals improves the belief about dissidents. Thus, by dissenting, dissidents exacerbate the severity of repression they face. Strong signals reveal more evidence, inducing less zealous types who would not retaliate otherwise to retaliate. On the other hand, as civilians reveal less information, they should expect a less severe repression. An important implication is that although repression is often indiscriminate, the severity of repression is a function of participation. This observation is stated in the next proposition. The proposition also outlines the crucial impact of the threat of repression on civilians compared to dissidents. The proof is in the Appendix (Lemma A.1 and its proof).

Proposition 2 *Due to dissidents revealing more evidence, they expect repression to be more severe:*

$$\sum_{s=0}^{n-1} \Pr[S^i = s] \times R(s+1, \tau) > \sum_{s=0}^{n-1} \Pr[S^i = s] \times R(s, \tau). \tag{4}$$

Assume $G(0) > \underline{\kappa}$. That is, the community is hostile and there are aggressive types in the model. If Defender is likely to be an overzealous type, then

$$\frac{\partial}{\partial \tau}\left[\sum_{s=0}^{n-1} \Pr[S^i = s] \times R(s, \tau)\right] > \frac{\partial}{\partial \tau}\left[\sum_{s=0}^{n-1} \Pr[S^i = s] \times R(s+1, \tau)\right]. \tag{5}$$

That is, civilians overreact in estimating the marginal impact of the threat of repression.

Although the threat of repression is common knowledge, its perceived impact on the severity of repression depends on whether a citizen is a dissident or not. The left-hand side of expression (5) is the impact of the threat of repression on the severity a civilian expects, while the right-hand side reflects the effect on the severity a dissident expects. Thus, compared to dissidents, civilians tend to overreact to the threat of repression if Defender is likely to be an overzealous type ($H(.)$ weakly convex).

I provide an intuition behind the inequality (5). Perceptions of the impact of the threat of repression on its severity are minimal for dissidents for at least three common reasons in conflict environments. First, dissidents are already involved in the conflict and have surmounted the barriers to participation, making it less likely for them to overreact. Second, repression is already more severe for dissidents compared to civilians, as they reveal more information when there is an identification problem. Third, dissidents frequently interpret the associated risks differently, given their ongoing confrontations with government troops.

Regarding civilians, they often experience less severe repression as they reveal less information. However, a slight increase in the threat of repression has a bigger marginal impact on the severity of repression for civilians, particularly if they anticipate that the threat will likely be executed by an overzealous type. This result is crucial because an overreaction

implies that the marginal risk of repression exhibits decreasing differences:

$$\frac{\partial}{\partial \tau}\left[\sum_{s=0}^{n-1} \Pr[S^i = s] \times (R(s+1,\tau) - R(s,\tau))\right] < 0 \quad \text{if} \quad G(0) > \underline{\kappa}.$$

The condition on aggressive types ($G(0) > \underline{\kappa}$), is only sufficient for the result. This condition arises because the signal generated by a citizen depends on the actions of others. It ensures that at least one citizen, except citizen i, chooses to dissent. In this regard, if the signal generated by others is 0 (or $\sum_{j\neq i} a_j = 0$), the signal Defender observes only depends on the action of citizen i. In this case, the overreaction result is less likely to arise since the individual's act of dissent raises the severity from $R(0,\tau) = 0$ to $R(1,\tau)$. Thus, for a civilian to overreact, she should expect to sometimes be targeted. Consequently, one needs a sufficient condition ensuring that an individual other than citizen i will always initiate an attack (i.e $\sum_{j\neq i} a_j \neq 0$). The result then highlights a novel contribution of aggressive types (those with a cost below 0), during an insurgency. Hence, the proposition has implications for the role of leadership in high-risk dissent.

Proposition 2 is essential for the main result as it emphasizes the role of inadequate information and the characteristics of security agents in shaping how dissidents and civilians perceive the severity of repression in rebellions and insurgencies. To be more precise, note that an individual with a private cost c_i chooses to dissent if the cost is smaller than

$$\frac{\frac{1}{d(\tau)}\left[\sum_{s=0}^{n-1}(\pi(s+1) - \gamma\pi(s))\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}\right]}{\sum_{s=0}^{n-1}\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s} \times (R(s+1,\tau) - R(s,\tau))} \equiv c(\alpha).$$

The threshold $c(\alpha)$ represents an individual's best response function to other individual decision to dissent, captured by the probability α. A symmetric equilibrium in which individuals participate with probability α^* must satisfy the indifference condition

$$\alpha = G\left(\frac{\frac{1}{d(\tau)}\left[\sum_{s=0}^{n-1}(\pi(s+1)-\gamma\pi(s))\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}\right]}{\sum_{s=0}^{n-1}\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}\times(R(s+1,\tau)-R(s,\tau))}\right),$$

where the equation states that the equilibrium probability of dissenting (α) should be equal to the probability that a citizen has a private cost lower than the marginal type $c(\alpha)$. The next proposition shows equilibrium uniqueness.

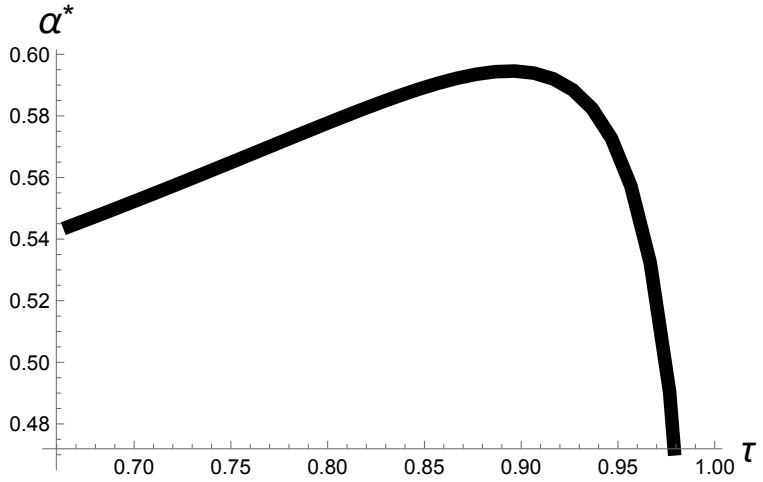


Figure 4: Deterrence with Inadequate Information and overzealous security agents.
The distribution $H(.)$ follows the Uniform distribution on $[0,1]$. Hence, the repressive apparatus has a (moderate) prevalence of overzealous types. The parameters are $n=3$ and $\underline{\tau}=0.67$, $\overline{\tau}=0.98$; $c_i \sim \mathcal{U}[-0.5,2]$. The likelihood of dissent (α^*) increases and then decreases. Therefore, there exists τ^*.

Proposition 3 *There exists $\xi > 0$ such that if $g(z) < \xi$ for all $z \in [\underline{c}, \overline{c}]$, then the game has a unique perfect Bayesian equilibrium (α^*, R, q).*

In this model, equilibrium uniqueness is determined by the fact that individuals coordinate on the unique equilibrium in the dissent stage. Furthermore, the next proposition outlines the failure of deterrence in a framework with the identification problem and a prevalence of overzealous types within a repressive apparatus.

Proposition 4 *Suppose $g(z) < \xi$ and that aggressive types live in the community. If the repressive apparatus has a prevalence (high or moderate) of overzealous types, there exists $\tau^* \in (\underline{\tau}, \overline{\tau})$ such that the probability an individual chooses to dissent is increasing for $\tau \leq \tau^*$. In order words, the threat of repression has a positive impact on the likelihood of dissent.*

The threat of repression τ has two countervailing effects on the decision to dissent. First, the impact on the likelihood of success is negative. But more importantly, the threat of repression negatively affects the net expected severity of repression if Defender is likely to be overzealous ($H(.)$ weakly convex). That is, civilians overreact and the marginal risk of repression exhibits decreasing differences. Consequently, as shown in Figure 4, when the threat of repression is below some threshold, the "overreaction" effect dominates. Civilians employ preemptive resistance as they anticipate the use of excessive force to enforce social control. However, for harsh threats of repression, the likelihood of success becomes small, and the overall effect on the decision to dissent is negative. The proof of Proposition 4 is contained in the Appendix. Proposition A.1 in the Appendix shows the uniqueness of τ^* when $H(.)$ is the uniform distribution on $[0, 1]$.

More formally, for a general weakly convex distribution $H(.)$, the effect of the threat of repression on the likelihood of dissent $d\alpha^*/d\tau$, takes the form of

$$\frac{d\alpha^*}{d\tau} = \frac{\frac{P(\alpha^*)}{d(\tau)M(\alpha^*,\tau)} g\left(\frac{P(\alpha^*)}{d(\tau)M(\alpha^*,\tau)}\right)}{\mathbf{H}'(x)} \times \left[-\frac{d'(\tau)}{d(\tau)} - \frac{\frac{\partial M(\alpha^*,\tau)}{\partial \tau}}{M(\alpha^*,\tau)}\right],$$

where $\mathbf{H}'(x)$, defined in the Appendix, is positive given the condition for equilibrium uniqueness;

$$P(\alpha) = \sum_{s=0}^{n-1} (\pi(s+1) - \gamma\pi(s)) \binom{n-1}{s} \alpha^s (1-\alpha)^{n-1-s}$$

is the net likelihood of success.

Furthermore,

$$M(\alpha,\tau)=\sum_{s=0}^{n-1}\left[H\left(\frac{1}{\tau}\left(1-\frac{s}{n}\right)\right)-H\left(\frac{1}{\tau}\left(1-\frac{s+1}{n}\right)\right)\right]\binom{n-1}{s}\alpha^{s}(1-\alpha)^{n-1-s}$$

is the marginal risk of repression (difference between severity for dissidents and civilians), which is positive since dissidents reveal more information (Equation (4)). In addition, since civilians overreact to the threat of repression, $\frac{\partial M(\alpha^*,\tau)}{\partial \tau}<0$ (Equation (5)). As a result, when τ is small, its effect on the likelihood of success is small as well ($d'(\tau)/d(\tau)$ is small), which indicates that $\frac{d\alpha^*}{d\tau}>0$.

The key lesson is that the often observed failure of deterrence in political contention may not be explained by in-group affinity, emotional reactions to repression, or grievances that spur new dissent. There is a rational mechanism that operates based on the information environment and the characteristics of the repressive apparatus. Grievances play a completely different role in the model. As shown in Proposition 2, the result is sustained if there are aggressive types in the model (players with a cost below zero). The condition $G(0)>\underline{\kappa}$ ensures that aggrieved citizens—those who care less about benefit and costs—maintain a resistance, while incriminating civilians. When such a condition exists, civilians may perceive a positive likelihood of being targeted. Consequently, repression does not generate grievances. In the next section, I show that by altering the information structure, the findings are strengthened, as this condition concerning aggressive types is rendered irrelevant.

In a seminal book on Iraq under Saddam Hussein, Blaydes 2018 studies repression with the identification problem. In her framework, inadequate information compels the regime to substitute ethnicity for lack of information. Given this mode of inference, expectations of repression increase political dissent by reinforcing group ties: people tend to believe that they have a 'linked-fate' and are willing to dissent despite the high risk of repression (Blaydes 2018, pg 47). Identities may have an explanatory power in the decision to dissent in the

shadow of repression. However, Proposition 4 highlights an alternative explanation.

The main point is that government forces, particularly those recognized for employing excessive force, can weaken deterrence when operating with limited information. This finding is consistent with the tactical directive issued by retired US Army General McChrystal, which emphasizes the importance of courageous restraint among US troops. Senior leadership within the United States military has reiterated the necessity of safeguarding civilian lives and minimizing civilian casualties, in accordance with international and humanitarian laws. Reducing civilian casualties is not only a legal obligation but also a moral imperative.

The finding presented in this paper highlights an additional value that such a tactical directive can provide to a counterinsurgency campaign. Observing restraint can limit the amount of support for the enemy. If the repressive apparatus has a low prevalence of overzealous types (e.g $H(.)$ is strictly concave), civilians hardly overreact to a threat of repression. Indeed, additional conditions are required for civilians to overreact when Defender is less zealous; Proposition 2 (Equation 5) may not hold when Defender is less zealous.

Extensions

I explore two extensions. First, I examine the role of technological advancement by considering a signal structure that provides fine-grained information. The number of insurgent attacks often inflicts significant harm on security forces, undermining their ability to retaliate. A second exercise studies the influence of this channel on the main result.

Alternative Mode of Inference.

Thus far, the analysis has relied on the assumption that Defender learns about individual participation at the macro level. This assumption is now relaxed in a setup where signals are individual-specific. With the reliance on advanced technology, security forces often monitor

suspected individuals, and as a result, can observe a noisy, and more precise signal about the individual's behavior. In light of this fact, I assume that Defender has fine-grained information about individual players and observe a signal

$$x_i = a_i + \epsilon$$

where $\epsilon \in \{-1, 0, 1\}$ is an integer, has mean 0, and variance $\sigma \in (0, 1)$, and is distributed according to

$$f(x_i \mid a_i) = \begin{cases} \frac{\sigma}{2} & \text{if } x_i = a_i + 1 \\ \frac{\sigma}{2} & \text{if } x_i = a_i - 1 \\ 1 - \sigma & \text{if } x_i = a_i, \end{cases}$$

which implies that $x \in \{-1, 0, 1, 2\}$. In this setup, I assume that the distribution of the value of retaliation H is distributed on $[0, \overline{v}]$ where $\overline{v} \geq 1$.

I show the robustness of the results, which is driven by the assumption that citizen's risk of repression is independent of the group behavior. This finding is unexpected, as one might assume that the primary result could be altered with more precise information. First, Defender is more informed about individual behavior in comparison to the previous setup. Second, because the risk of repression is now more closely associated with individual behavior and less with group behavior, we can no longer rely on aggressive types to encourage participation.

I further assume that σ is such that $\frac{f(x|1)}{f(x|0)}$ is strictly increasing in x. This is the Monotone Likelihood Ratio assumption. It holds if and only if $\sigma \in (0, \frac{2}{3})$.[10] Intuitively, the two variables a and x are affiliated implies that higher value of x are likely to be associated with higher values of a.

10. One only needs to check if $\frac{f(0|1)}{f(0|0)} < \frac{f(1|1)}{f(1|0)}$. That is, the condition must ensure that $\frac{\sigma}{2(1-\sigma)} < \frac{2(1-\sigma)}{\sigma}$.

If we fix the likelihood of dissenting, α, Defender's posterior belief is given by

$$q(x) = \begin{cases} 0 & \text{if } x = -1 \\ \frac{\alpha\sigma}{\alpha\sigma + 2(1-\alpha)(1-\sigma)} & \text{if } x = 0 \\ \frac{2\alpha(1-\sigma)}{2\alpha(1-\sigma)+\sigma(1-\alpha)} & \text{if } x = 1 \\ 1 & \text{if } x = 2. \end{cases}$$

Repression in this model remains indiscriminate. A civilian may be targeted primarily due to the noise in the signal they generate. Consequently, regardless of the number of individuals involved in dissent, a civilian is targeted if $\epsilon \in \{0, 1\}$. Thus, in contrast to the previous setting, an individual does not need to be concerned about the aggregate level of dissent when assessing the risks of repression.

It turns out that when Defender observes a signal $x = -1$, it never retaliates. Thus, $R(-1, \tau) = 0$. This is because a signal $x = -1$ can only be generated by a civilian. Defender retaliates after observing a signal of $x \in \{0, 1\}$ with probability

$R(x, \tau) = \Pr\left[v \geq \frac{1}{\tau}\left(1 - q(x)\right)\right] = 1 - H\left(\frac{1}{\tau}\left(1 - q(x)\right)\right)$. Moreover, when $x = 2$, Defender always retaliates; $R(2, \tau) = 1$. A direct implication is that as x increases, the severity of repression $R(x, \tau)$ increases as well. Hence, dissidents always reveal more information, increasing the severity of repression. Furthermore, the expected severity of repression for a civilian is given by

$$\sum_{x=-1}^{x=1} R(x, \tau) f(x \mid a = 0) \quad = \quad 0 + (1 - \sigma) R(0, \tau) + \left(\frac{\sigma}{2}\right) R(1, \tau),$$

while a dissident expects a severity

$$\sum_{x=0}^{x=2} R(x,\tau) f(x \mid a=1) \quad = \quad \left(\frac{\sigma}{2}\right) R(0,\tau) + (1-\sigma) R(1,\tau) + \frac{\sigma}{2}.$$

The net expected severity of repression is the difference between the risk of retaliation for a dissident and a civilian.

$$\sum_{x} R(x,\tau) \left(f(x \mid a=1) - f(x \mid a=0)\right) \quad = \quad \frac{\sigma}{2} + \left(1 - \frac{3}{2}\sigma\right) \left[H\left(\frac{1}{\tau}\left(1-q(0)\right)\right) - H\left(\frac{1}{\tau}\left(1-q(1)\right)\right)\right].$$

I now state the next lemma that highlights the trade-offs citizens face in this extended framework.

Lemma 3 *When Defender observes individual-specific signals, dissenting makes repression more severe:*

$$\sum_{x} R(x,\tau) \left(f(x \mid a=1) - f(x \mid a=0)\right) > 0.$$

Suppose there is a prevalence of overly zealous types in the repressive apparatus ($H(.)$ is weakly convex). Civilians always overreact to the threat of repression:

$$\frac{\partial}{\partial \tau} \left[\sum_{x} R(x,\tau) \left(f(x \mid a=1) - f(x \mid a=0)\right)\right] < 0.$$

Civilians confronted with a repressive apparatus characterized by a significant presence of overzealous types tend to overreact to the perceived threat of repression. Lemma 3 specifies that the condition regarding aggressive types is unnecessary when the noise in the generated signal is independent of other actions. Moreover, there exists a threshold τ_* such that this overreaction results in a failure of deterrence if $\tau \leq \tau_*$. This is outlined in the next

proposition.

Proposition 5 *Suppose $H(.)$ is (weakly) convex.*

1. *There exists $\xi_1 > 0$ such that if $g(z) < \xi_1$ for $z \in [\underline{c}, \overline{c}]$, the game has a unique equilibrium.*

2. *There exists $\tau_* \in (\underline{\tau}, \overline{\tau})$ such that if $\tau \leq \tau^*$, the threat of repression increases the likelihood of dissent.*

Harmful Insurgent Attacks

I also explore an extension where insurgent attacks are harmful to security forces. I assume that the extent of harm is proportional to the level of dissent. Insurgent attacks are harmful if Defender engages with insurgents by retaliating. A Defender who chooses not to retaliate cannot be harmed. I assume that Defender's payoff is given by

$$\text{Retaliation} \qquad \tau \times v - \lambda \tfrac{s}{n} - 1 \times \mathbb{1}\{\text{i is not a dissident}\}$$

$$\text{No Retaliation} \qquad 0,$$

Where $\lambda \in (0, 1)$, with the additional condition $\frac{1}{\underline{\tau}} \left(\frac{n-1+\lambda}{n} \right) < 1$.

As the posterior belief is unchanged, this payoff structure implies that a type v retaliates if and only if

$$v \geq \frac{1}{\tau} \left(1 - \frac{s(1-\lambda)}{n} \right).$$

This condition yields a severity of repression

$$R(s,\tau) = 1 - H\left(\frac{1}{\tau}\left(1 - \frac{s(1-\lambda)}{n}\right)\right).$$

I can then state the next Proposition that outlines the important trade-off dissidents and civilians face in this model.

Proposition 6 *When insurgent attacks impose injuries to Defender, dissenting makes retaliation severe. Furthermore, if* $G(0) > \underline{\underline{\kappa}}$ *and there is a prevalence of overzealous types in the repressive apparatus, civilians overreact to the threat of repression.*

The intuition relies on the observation that the additional cost of retaliation screened out types who would otherwise observe restraint. This can be observed by noting that the threshold $\frac{1}{\tau}\left(1 - \frac{s(1-\lambda)}{n}\right)$, increases with λ.

Conclusion

Government repression is a traditional response to political dissent, often fostering an environment of fear and pessimism. However, the prevailing analyzes of the failure of deterrence in insurgencies and rebellions are primarily based on psychological explanations. This limitation in the literature underscores the need for a rational model that enhances our understanding of the dynamics between repression and dissent. This paper explores a model of collective action that emphasizes the identification problem in rebellions. A repressive agent who can be overzealous must execute a repressive order with limited information about dissidents. I find that civilians may overreact to the threat of repression if they anticipate security overreach.

The paper underscores the roles of limited information and the actions of security agents in shaping the success of counterinsurgency campaigns. Limited information is a critical component of the model, as dissidents distinguish themselves from civilians by revealing

useful information that intensifies repression. While civilians experience less severe repression, they are the ones most likely to overreact to the threat of it. The frequently observed failure of deterrence in political contentions may be a manifestation of this mechanism.

Other trade-offs specific to this context have been excluded and may serve as topics for future work. Citizens often engage in "in-group policing", as they anticipate that a member action can be detrimental to the group. This channel can be crucial if security forces observe signals at the macro level. Furthermore, future research could also explore the influence of various information structures and incentives for security agents to acquire precise information about dissidents.

Appendix

I start with a clear statement for Remark 1, and provide its proof.

Remark A.1 *Assume Defender perfectly observes the act of dissent.*

1. *Defender retaliates against an individual if and only if the individual is a dissident.*
2. *There exists $\xi_0 > 0$ such that if $g(x) < \xi_0$, the game with perfect information has a unique equilibrium.*
3. *Suppose $g(x) < \xi_0$. The likelihood of dissenting is decreasing in the threat of repression: $\frac{d\alpha^{BM}}{d\tau} \leq 0$.*

Proof

It is clear that Defender's belief is either 1 or 0. If individual i is not a dissident, Defender receives $\tau \times v - 1$ if it retaliates against individual i. If Defender does not retaliate it receives 0. Since $\tau \times v - 1 < 0$, Defender prefers not to retaliate against an individual known to be a civilian. However, if individual i is a dissident, it is optimal to retaliate.

A citizen with private cost c_i has a net expected payoff $\frac{\mathbb{E}[\pi(s+1)-\gamma\pi(s)]}{d(\tau)} - c_i$ from dissenting. The expectation is taken over S^i, the Binomial distribution. It is clear that the dissent strategy is in threshold strategy defined by the cut-point c^{BM}. An individual with private cost c_i chooses to dissent if $c_i \leq c^{BM}$.

Given that others adopt a threshold c^{BM}, the likelihood of dissent is given by $\alpha^{BM} = \Pr[c_i \leq c^{BM}]$. α^{BM} solves the equation

$$\alpha = G\left(\frac{\mathbb{E}[\pi(s+1) - \gamma\pi(s)]}{d(\tau)}\right) \equiv G\left(\frac{P(\alpha)}{d(\tau)}\right). \tag{6}$$

When $\alpha^{BM} = 0$, one obtains $0 < G(P(0)/d(\tau))$; when $\alpha^{BM} = 1$, $1 > G(P(1)/d(\tau))$. Thus, a solution to Equation (6) exists. It is unique if $g(c) < \xi_0 = \min_t \left[\left|\frac{d(\tau)}{P'(t)}\right|\right]$. The Best-response functions must have a slope less than 1 in absolute value.

To be more precised, consider the function $x - G\left(\frac{P(x)}{d(\tau)}\right)$ defined on $(0,1)$. The derivative with respect to x yields $1 - \frac{P'(x)}{d(\tau)} g\left(\frac{P(x)}{d(\tau)}\right)$.[11] The equation $x = G\left(\frac{P(x)}{d(\tau)}\right)$ has a unique solution if for all $c \in [\underline{c}, \overline{c}]$, $g(c) < \min_t \left[\left|\frac{d(\tau)}{P'(t)}\right|\right] \equiv \xi_0$.

Moreover, by the implicit function theorem

$$\begin{aligned}\frac{d\alpha^{BM}}{d\tau} &= \frac{\partial}{\partial\tau}\left[G\left(\frac{P(\alpha^{BM})}{d(\tau)}\right)\right] + \frac{\partial}{\partial\alpha}\left[G\left(\frac{P(\alpha^{BM})}{d(\tau)}\right)\right]\frac{d\alpha^{BM}}{d\tau} \\ &= -P(\alpha^{BM})\left(\frac{d'(\tau)}{(d(\tau))^2}\right) g\left(\frac{P(\alpha^{BM})}{d(\tau)}\right) + \frac{P'(\alpha^{BM})}{d(\tau)} g\left(\frac{P(\alpha^{BM})}{d(\tau)}\right)\frac{d\alpha^{BM}}{d\tau}\end{aligned}$$

$$\frac{d\alpha^{BM}}{d\tau} = -\frac{P(\alpha^{BM})\left(\frac{d'(\tau)}{(d(\tau))^2}\right) g\left(\frac{P(\alpha^{BM})}{d(\tau)}\right)}{1 - \frac{P'(\alpha^{BM})}{d(\tau)} g\left(\frac{P(\alpha^{BM})}{d(\tau)}\right)} \leq 0.$$

This is because $1 - \frac{P'(\alpha^{BM})}{d(\tau)} g\left(\frac{P(\alpha^{BM})}{d(\tau)}\right) > 0$ (if $g(c) < \xi_0$) and $d' \geq 0$. *Q.E.D*

Proof of Lemma 1.

It is important to note that $q(0) = 0$, $q(n) = 1$ and $q(1) = \frac{1}{n}$.

Now suppose Defender observes $s = 2$. One can deduce that

$$q(2) = \frac{\binom{n-1}{1}\alpha^2(1-\alpha)^{n-2}}{\binom{n}{2}\alpha^2(1-\alpha)^{n-2}} = \frac{2}{n}.$$

Citizen i chooses to dissent with either $j \neq i$, or any other $k \neq i$. Moreover, I use the fact that the events $\{(i,j), s=2\}_{\substack{\{i,j\in N\}\\ i\neq j}}$ form a partition of the set of citizens, to apply the law of total probabilities. There are a total of $\binom{n}{2}$ such events and they have equal probability.

Finally, for any observed level of dissent $s \geq 3$,

$q(s) = \frac{\binom{n-1}{s-1}\alpha^s(1-\alpha)^{n-s}}{\binom{n}{s}\alpha^s(1-\alpha)^{n-s}} = \frac{s}{n}$. *Q.E.D*

Proof of Proposition 1 The proof follows from Lemma **??** and the discussion in the text.

11. Note that $P'(x)$ can be negative, if the outcome of the rebellion is a public good.

Q.E.D

Proof of Lemma 2 Given the assumption that $\frac{1}{\underline{\tau}}\left(\frac{n-1}{n}\right) < 1$, it turns out that for any $\tau \in [\underline{\tau}, \overline{\tau}]$ and any dissent signal s such that $1 \leq s \leq n$, $\frac{1}{\tau}\left(1-\frac{s}{n}\right) \in [0,1]$, and $R(s,\tau) \geq 0$. Now, since $s < s+1$, $\frac{1}{\tau}\left(1-\frac{s}{n}\right) > \frac{1}{\tau}\left(1-\frac{s+1}{n}\right)$. A direct implication is that $H\left(\frac{1}{\tau}\left(1-\frac{s}{n}\right)\right) > H\left(\frac{1}{\tau}\left(1-\frac{s+1}{n}\right)\right)$. *Q.E.D*

Before proving Proposition 1 I state the following Lemma along with its proof.

Denote

$$\begin{aligned} M(\alpha;\tau) &\equiv \sum_{s=0}^{n-1} \Pr[S^i = s] \times (R(s+1,\tau) - R(s,\tau)) \\ &= \sum_{s=1}^{n-1} \left[H\left(\frac{1}{\tau}\left(1-\frac{s}{n}\right)\right) - H\left(\frac{1}{\tau}\left(1-\frac{s+1}{n}\right)\right)\right] \binom{n-1}{s} \alpha^s (1-\alpha)^{n-1-s} \\ &+ 1 - H\left(\frac{1}{\tau}\left(\frac{n-1}{n}\right)\right), \end{aligned} \tag{7}$$

$\underline{h} = \inf_{x\in[0,1]} h(x)$ and $\overline{h} = \sup_{x\in[0,1]} h(x)$.

Lemma A.1 *Assume that H is (weakly) convex. If $G(0) > 1 - \left(\frac{\underline{h}}{\underline{h}+(n-1)\overline{h}}\right)^{\frac{1}{n-1}}$, then for any $\tau \in [\underline{\tau}, \overline{\tau}]$, $\frac{\partial M(\alpha;\tau)}{\partial \tau} < 0$.*

Proof: Observe that $\frac{\partial M(\alpha;\tau)}{\partial \tau}$

$$\begin{aligned} &= \frac{1}{\tau^2} \sum_{s=1}^{n-1} \left[\left(1-\frac{s+1}{n}\right) h\left(\frac{1}{\tau}\left(1-\frac{s+1}{n}\right)\right) - \left(1-\frac{s}{n}\right) h\left(\frac{1}{\tau}\left(1-\frac{s}{n}\right)\right)\right] \binom{n-1}{s} \alpha^s (1-\alpha)^{n-1-s} \\ &+ (1-\alpha)^{n-1} \frac{1}{\tau^2} \left(\frac{n-1}{n}\right) h\left(\frac{1}{\tau}\left(\frac{n-1}{n}\right)\right). \end{aligned}$$

In the first term of the previous equality, we have for any $s \geq 1$,

$$h\left(\frac{1}{\tau}\left(1-\frac{s+1}{n}\right)\right) \leq h\left(\frac{1}{\tau}\left(1-\frac{s}{n}\right)\right),$$

for H to be (weakly) convex. In this respect, noting that $\underline{h} = \inf_{x\in[0,1]} h(x)$,

$$\begin{aligned}\sum_{s=1}^{n-1}\left[\left(1-\tfrac{s+1}{n}\right)h\left(\tfrac{1}{\tau}\left(1-\tfrac{s+1}{n}\right)\right)-\left(1-\tfrac{s}{n}\right)h\left(\tfrac{1}{\tau}\left(1-\tfrac{s}{n}\right)\right)\right]\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}&\\ \leq -\tfrac{1}{n}\underline{h}\sum_{s=1}^{n-1}\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}&\\ = -\tfrac{1}{n}\underline{h}(1-(1-\alpha)^{n-1}).&\end{aligned}$$

In addition, since $\overline{h} = \sup_{x\in[0,1]} h(x)$,

$$(1-\alpha)^{n-1}\frac{1}{\tau^2}\left(\frac{n-1}{n}\right)h\left(\frac{1}{\tau}\left(\frac{n-1}{n}\right)\right) \leq (1-\alpha)^{n-1}\frac{1}{\tau^2}\left(\frac{n-1}{n}\right)\overline{h}.$$

I then obtain

$$\frac{\partial M(\alpha;\tau)}{\partial \tau} \leq \frac{1}{n\tau^2}\left(-\underline{h}+\underline{h}(1-\alpha)^{n-1}+(n-1)\overline{h}(1-\alpha)^{n-1}\right).$$

If $(1-\alpha)^{n-1} < \frac{\underline{h}}{\underline{h}+(n-1)\overline{h}}$, or $G(0) > 1-\left(\frac{\underline{h}}{\underline{h}+(n-1)\overline{h}}\right)^{\frac{1}{n-1}}$, then the likelihood of dissent is at least $1-\left(\frac{\underline{h}}{\underline{h}+(n-1)\overline{h}}\right)^{\frac{1}{n-1}}$, and $\frac{\partial M(\alpha;\tau)}{\partial \tau} < 0$. *Q.E.D*

Proof of Proposition 2.

By Lemma 1 for any $s < n$, $R(s+1,\tau) > R(s,\tau)$. Therefore,

$$\sum_{s=0}^{n-1}\Pr[S^i = s]\times(R(s+1,\tau)-R(s,\tau)) > 0.$$

Now suppose that if $G(0) > 1-\left(\frac{\underline{h}}{\underline{h}+(n-1)\overline{h}}\right)^{\frac{1}{n-1}} \equiv \underline{\kappa}$. Then using Lemma A.1,

$$\frac{\partial}{\partial \tau}\left[\sum_{s=0}^{n-1}\Pr[S^i=s]\times(R(s+1,\tau)-R(s,\tau))\right]=\frac{\partial M(\alpha;\tau)}{\partial \tau}<0.$$

Q.E.D

Proof of Proposition 3.

Denote c^* the equilibrium threshold. Further denote $\alpha=\Pr[c_i\leq c^*]$. In equilibrium, we must have the indifference equation

$$\alpha=G\left(\frac{\frac{1}{d(\tau)}\left[\sum_{s=0}^{n-1}(\pi(s+1)-\gamma\pi(s))\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}\right]}{\sum_{s=0}^{n-1}\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}\times(R(s+1,\tau)-R(s,\tau))}\right).$$

Denote $P(\alpha)=\mathbb{E}[\pi(s+1)-\gamma\pi(s)]=\sum_{s=0}^{n-1}(\pi(s+1)-\gamma\pi(s))\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}$. Then the indifference equation can be rewritten as

$$\alpha=G\left(\frac{P(\alpha)}{d(\tau)M(\alpha,\tau)}\right),$$

where $M(\alpha,\tau)>0$ is given by (7).

When $\alpha=0$, $0<G\left(\frac{P(0)}{d(\tau)M(0,\tau)}\right)$; moreover, when $\alpha=1$, $1>G\left(\frac{P(1)}{d(\tau)M(1,\tau)}\right)$, since $\overline{c}$ is large enough. Hence, there exists α^*, solution to the indifference equation.

Under the condition that $g(.)$ is flat enough there exists a unique equilibrium $\alpha^*(\tau)$. In fact there exists $\xi>0$ such that if $g(z)<\xi$ for $z\in[\underline{c},\overline{c}]$, then the game has a unique equilibrium.

Consider the function

$$\mathbf{H}(x;\tau)\equiv\mathbf{H}(x)=x-G\left(\frac{P(x)}{d(\tau)M(x,\tau)}\right)$$

defined on $[0,1]$. The function $\mathbf{H}$ is continuous and differentiable. The derivative with

respect to x is given by

$$\mathbf{H}'(x) = 1 - \left(\frac{P(x)}{d(\tau)M(x,\tau)}\right)\left[\frac{P'(x)}{P(x)} - \frac{\frac{\partial M(x,\tau)}{\partial x}}{M(x,\tau)}\right] \times g\left(\frac{P(x)}{d(\tau)M(x,\tau)}\right).$$

A unique equilibrium exists if the derivative is positive. The condition holds if for any $z \in [\underline{c}, \overline{c}]$,

$$1 > \sup_{x\in[0,1]} \left\{ \frac{P(x)}{d(\tau)M(x,\tau)} \left| \frac{P'(x)}{P(x)} - \frac{\frac{\partial M(x,\tau)}{\partial x}}{M(x,\tau)} \right| \right\} \times g(z),$$

where $|.|$ is the absolute value function.

Since the function

$$x \mapsto \frac{P(x)}{d(\tau)M(x,\tau)} \left| \frac{P'(x)}{P(x)} - \frac{\frac{\partial M(x,\tau)}{\partial x}}{M(x,\tau)} \right|$$

is continuous and positive on the compact $[0,1]$, a supremium exists and is positive. Let $1/\xi$ be the supremium where $\xi > 0$.

Suppose $g(z) < \xi$ for any $z \in [\underline{c}, \overline{c}]$. Then one obtains

$$\begin{aligned} \mathbf{H}'(x) &= 1 - \left(\frac{P(x)}{d(\tau)M(x,\tau)}\right)\left[\frac{P'(x)}{P(x)} - \frac{\frac{\partial M(x,\tau)}{\partial x}}{M(x,\tau)}\right] \times g\left(\frac{P(x)}{d(\tau)M(x,\tau)}\right) \\ &\geq 1 - \frac{1}{\xi} \times g\left(\frac{P(x)}{d(\tau)M(x,\tau)}\right) > 0. \end{aligned}$$

Since the function $x \mapsto \mathbf{H}(x)$ is strictly increasing, with $\mathbf{H}(0) < 0$ and $\mathbf{H}(1) > 0$, the equation $\mathbf{H}(x) = 0$ has a unique solution. *Q.E.D*

I prove Proposition 3 while considering a slightly modified version.

Proposition A.1 *Suppose $g(z) < \xi$ and that aggressive types live in the community. Further suppose that the repressive apparatus is made of overzealous security agents ($H(.)$ is weakly*

convex).

1. *There exists* $\tau^* \in (\underline{\tau}, \overline{\tau})$ *such that the probability an individual participates in dissent activities is increasing for* $\tau \leq \tau^*$*. In order words, the threat of repression has a positive impact on the likelihood of dissent.*

2. *If* $H \sim \mathcal{U}[0,1]$ *then* τ^* *is unique if* $G(0) > 1 - \left(\frac{1}{n}\right)^{\frac{1}{n-1}}$*. Furthermore, Figure 4 in the text represents the relationship between the likelihood of dissent and the threat of repression for* $\tau \in [\underline{\tau}, \overline{\tau}]$*.*

Proof.

Part 1.

By the Implicit Function Theorem, the derivative of the likelihood of dissent α^* with respect to τ must have the same sign as $\frac{\partial}{\partial \tau}\left[G\left(\frac{P(\alpha^*)}{d(\tau)M(\alpha^*;\tau)}\right)\right]$, since $x \mapsto \mathbf{H}(x)$ is an increasing function. In fact,

$$\frac{d\alpha^*}{d\tau} = \frac{\frac{\partial}{\partial \tau}\left[G\left(\frac{P(\alpha^*)}{d(\tau)M(\alpha^*;\tau)}\right)\right]}{\mathbf{H}'(x)\big|_{x=\alpha^*}},$$

where

$$\begin{aligned}\frac{\partial}{\partial \tau}\left[G\left(\frac{P(\alpha^*)}{d(\tau)M(\alpha^*;\tau)}\right)\right] &= \frac{P(\alpha^*)}{d(\tau)M(\alpha^*;\tau)} \times g\left(\frac{P(\alpha^*)}{d(\tau)M(\alpha^*;\tau)}\right) \\ &\times \left[-\frac{d'(\tau)}{d(\tau)} - \frac{\frac{\partial M(\alpha^*;\tau)}{\partial \tau}}{M(\alpha^*;\tau)}\right]. \end{aligned} \tag{8}$$

Since $d'(\underline{\tau}) = 0$, by Lemma A.1, $-\frac{d'(\tau=\underline{\tau})}{d(\tau=\underline{\tau})} - \frac{\frac{\partial M(\alpha^*;\tau=\underline{\tau})}{\partial \tau}}{M(\alpha^*;\tau=\underline{\tau})} > 0$. In addition, since $d'(\overline{\tau})$ is extremely large, the derivative is negative.

The analysis shows that there exists $\tau^\circ \in [\underline{\tau}, \overline{\tau}]$, not necessarily unique, such that $-\frac{d'(\tau)}{d(\tau)} - \left.\frac{\frac{\partial M(\alpha^*;\tau)}{\partial \tau}}{M(\alpha^*;\tau)}\right|_{\tau=\tau^\circ} = 0$. Denote τ^* the smallest τ° (if there are many) such that $\frac{\partial}{\partial \tau}\left[G\left(\frac{P(\alpha^*)}{d(\tau)M(\alpha^*;\tau)}\right)\right] = 0$. By continuity, τ^* exists.

Part 2.

Now suppose that $H(.)$ is the uniform distribution on $[0,1]$. Then,

$$M(\alpha^*, \tau) = \frac{1}{\tau}\left(\frac{1}{n} - (1-\alpha^*)^{n-1}\right) + (1-\alpha^*)^{n-1}.$$

Since $\underline{h} = \overline{h} = 1$, the condition on aggressive types become $G(0) > 1 - \left(\frac{1}{n}\right)^{\frac{1}{n-1}}$. This indicates that $\frac{1}{n} - (1-\alpha^*)^{n-1} > 0$ and $\frac{\partial M(\alpha^*,\tau)}{\partial \tau} < 0$.

$$sign\left(\frac{d\alpha^*}{d\tau}\right) = sign\left(\frac{\partial}{\partial \tau}\left[G\left(\frac{P(\alpha^*)}{d(\tau)\left(\frac{1}{\tau}\left(\frac{1}{n} - (1-\alpha^*)^{n-1}\right) + (1-\alpha^*)^{n-1}\right)}\right)\right]\right).$$

Denote $\Delta \equiv \Delta(\alpha^*;\tau) = \frac{P(\alpha^*)}{d(\tau)\left(\frac{1}{\tau}\left(\frac{1}{n} - (1-\alpha^*)^{n-1}\right) + (1-\alpha^*)^{n-1}\right)}$. By applying Equation (8) where H is the uniform distribution, I obtain

$$\begin{aligned}
\frac{\partial}{\partial \tau}\left[G\left(\Delta(\alpha^*;\tau)\right)\right] &= \Delta(\alpha^*,\tau) \times g(\Delta(\alpha^*,\tau)) \times \left(\frac{1}{\tau}\right) \\
&\times \left[-\tau\frac{d'(\tau)}{d(\tau)} + \frac{\frac{1}{\tau}\left(\frac{1}{n} - (1-\alpha^*)^{n-1}\right)}{\frac{1}{\tau}\left(\frac{1}{n} - (1-\alpha^*)^{n-1}\right) + (1-\alpha^*)^{n-1}}\right] \\
&= \Delta(\alpha^*;\tau) \times g\left(\Delta(\alpha^*;\tau)\right)\left(\frac{1}{\tau}\right)\Sigma(\alpha^*;\tau) \qquad (9)
\end{aligned}$$

where

$$\begin{aligned}\Sigma(\alpha^*;\tau) &= -\tau\frac{d'(\tau)}{d(\tau)} + \frac{\frac{1}{\tau}\left(\frac{1}{n} - (1-\alpha^*)^{n-1}\right)}{\frac{1}{\tau}\left(\frac{1}{n} - (1-\alpha^*)^{n-1}\right) + (1-\alpha^*)^{n-1}} \\ &= -V(\tau) + \frac{L_1(\alpha^*,\tau)}{L_1(\alpha^*,\tau) + L_2(\alpha^*,\tau)}.\end{aligned}$$

I denote $V(\tau) = \tau\frac{d'(\tau)}{d(\tau)}$, $L_1(\alpha^*,\tau) = \frac{1}{\tau}\left(\frac{1}{n} - (1-\alpha^*)^{n-1}\right)$ and $L_2(\alpha^*,\tau) = (1-\alpha^*)^{n-1}$. It is important to note that $L_i > 0$ for $i \in \{1,2\}$.

Since all the terms in Equation (9) are positive except for $\Sigma(\alpha^*;\tau)$ which may be negative, it is clear that

$$sign\left(\frac{d\alpha^*}{d\tau}\right) = sign\left(\frac{\partial}{\partial\tau}\left[G\left(\Delta(\alpha^*;\tau)\right)\right]\right) = sign\left(\Sigma(\alpha^*;\tau)\right).$$

It is then sufficient and necessary to study the function Σ. As a function of τ, Σ is continuous and differentiable on the interval $[\underline{\tau},\overline{\tau}]$. Moreover, $\Sigma(\tau = \underline{\tau},\alpha^*) = \frac{L_1(\alpha^*,\tau=\underline{\tau})}{L_1(\alpha^*,\tau=\underline{\tau})+L_2(\alpha^*,\tau=\underline{\tau})} > 0$, since $d'(\underline{\tau}) = 0$. In addition, at $\tau = \overline{\tau}$, since $d'(\overline{\tau})$ is extremely large, $\Sigma(\tau = \overline{\tau},\alpha^*) < 0$. Therefore, there exists a τ^* such that $\Sigma(\tau = \tau^*,\alpha^*) = 0$.

I show that τ^* is unique. I prove this by showing that $\Sigma(\tau,\alpha^*)$ has a single-crossing. In fact, the single-crossing occurs from above.

Given that α^* is also a function of τ, one obtains

$$\begin{aligned}
\frac{d\Sigma(\tau,\alpha^*)}{d\tau} &= -V'(\tau) + \frac{\left(\frac{\partial L_1}{\partial \tau} + \frac{\partial L_1}{\partial \alpha}\frac{d\alpha^*}{d\tau}\right)(L_1+L_2) - L_1\left(\frac{\partial L_1}{\partial \tau} + \frac{\partial L_1}{\partial \alpha}\frac{d\alpha^*}{d\tau} + \frac{\partial L_2}{\partial \tau} + \frac{\partial L_2}{\partial \alpha}\frac{d\alpha^*}{d\tau}\right)}{[L_1+L_2]^2} \\
&= -V'(\tau) + \frac{L_2\left(\frac{\partial L_1}{\partial \tau} + \frac{\partial L_1}{\partial \alpha}\frac{d\alpha^*}{d\tau}\right) - L_1\left(\frac{\partial L_2}{\partial \tau} + \frac{\partial L_2}{\partial \alpha}\frac{d\alpha^*}{d\tau}\right)}{[L_1+L_2]^2} \\
&= -V'(\tau) + \frac{L_2\frac{\partial L_1}{\partial \tau}}{[L_1+L_2]^2} + \frac{d\alpha^*}{d\tau}\left(\frac{L_2\frac{\partial L_1}{\partial \alpha} - L_1\frac{\partial L_2}{\partial \alpha}}{[L_1+L_2]^2}\right) \quad \text{since } \tfrac{\partial L_2}{\partial \tau} = 0 \\
&= -V'(\tau) - \frac{L_2\left(\frac{1}{\tau^2}\right)\left(\frac{1}{n} - (1-\alpha^*)^{n-1}\right)}{[L_1+L_2]^2} \\
&+ \frac{d\alpha^*}{d\tau}\left(\frac{(n-1)(1-\alpha^*)^{n-2}\left(\frac{1}{\tau}L_2 + L_1\right)}{[L_1+L_2]^2}\right) \qquad (10)
\end{aligned}$$

It is clear that the two terms $-V'(\tau)$ and $-\frac{L_2\left(\frac{1}{\tau^2}\right)\left(\frac{1}{n}-(1-\alpha^*)^{n-1}\right)}{[L_1+L_2]^2}$ are negative. This is because $d'(\tau)/d(\tau)$ is increasing, by log-convexity, and $\frac{1}{n} - (1-\alpha^*)^{n-1} > 0$ since $G(0) > 1 - \left(\frac{1}{n}\right)^{\frac{1}{n-1}}$. Furthermore, $(n-1)(1-\alpha^*)^{n-2}\left(\frac{1}{\tau}L_2 + L_1\right)$ is positive because L_1 and L_2 are positive. Using the expression of $\frac{d\alpha^*}{d\tau}$, one obtains

$$\begin{aligned}
\frac{d\alpha^*}{d\tau} &= \frac{\frac{\partial}{\partial \tau}\left[G\left(\Delta(\alpha^*;\tau)\right)\right]}{\left.\frac{\partial \mathbf{H}(x)}{\partial x}\right|_{x=\alpha^*}} \\
&= \frac{\frac{P(\alpha^*)g(\Delta(\alpha^*;\tau))\left(\frac{1}{\tau}\right)\Sigma(\alpha^*;\tau)}{\left[d(\tau)\left(\frac{1}{\tau}\left(\frac{1}{n}-(1-\alpha^*)^{n-1}\right)+(1-\alpha^*)^{n-1}\right)\right]}}{\left.\frac{\partial \mathbf{H}(x)}{\partial x}\right|_{x=\alpha^*}} \text{ (using equation (9))} \\
&= \frac{P(\alpha^*)g\left(\Delta(\alpha^*;\tau)\right)\left(\frac{1}{\tau}\right)}{\left[d(\tau)\left(\frac{1}{\tau}\left(\frac{1}{n}-(1-\alpha^*)^{n-1}\right)+(1-\alpha^*)^{n-1}\right)\right]\left(\left.\frac{\partial \mathbf{H}(x)}{\partial x}\right|_{x=\alpha^*}\right)}\Sigma(\alpha^*;\tau) \qquad (11)
\end{aligned}$$

Plugging the expression (11) into (10) implies

$$\begin{aligned}\frac{d\Sigma(\tau,\alpha^*)}{d\tau} &= \frac{P(\alpha^*)g\left(\Delta(\alpha^*;\tau)\right)\left(\frac{1}{\tau}\right)}{\left[d(\tau)\left(\frac{1}{\tau}\left(\frac{1}{n}-(1-\alpha^*)^{n-1}\right)+(1-\alpha^*)^{n-1}\right)\right]\left(\frac{\partial \mathbf{H}(x)}{\partial x}\Big|_{x=\alpha^*}\right)} \\ &\times \frac{(n-1)(1-\alpha^*)^{n-2}\left(\frac{1}{\tau}L_2+L_1\right)}{[L_1+L_2]^2}\times\Sigma(\alpha^*;\tau) \\ &- V'(\tau)-\frac{L_2\left(\frac{1}{\tau^2}\right)\left(\frac{1}{n}-(1-\alpha^*)^{n-1}\right)}{[L_1+L_2]^2}\end{aligned}$$

Given that both $\frac{\partial \mathbf{H}(x)}{\partial x}\Big|_{x=\alpha^*}$ and

$$\frac{P(\alpha^*)g\left(\Delta(\alpha^*;\tau)\right)\left(\frac{1}{\tau}\right)}{\left[d(\tau)\left(\frac{1}{\tau}\left(\frac{1}{n}-(1-\alpha^*)^{n-1}\right)+(1-\alpha^*)^{n-1}\right)\right]}$$

are positive, it turns out that $\frac{d\Sigma(\tau,\alpha^*)}{d\tau}\leq 0$ if $\Sigma\leq 0$. Hence, Σ is a decreasing function when negative. Consequently, $\Sigma(\tau;\alpha^*)$ has a single-crossing from above.

To conclude, there exists a unique τ^* such that $\Sigma(\tau^*,\alpha^*)=0$. Moreover, if $\tau\leq\tau^*$ then $\Sigma(\tau^*,\alpha^*)>0$ which implies that $\frac{d\alpha^*}{d\alpha}>0$. Furthermore, if $\tau>\tau^*$, $\Sigma(\tau,\alpha^*)<0$ which implies that $\frac{d\alpha^*}{d\alpha}<0$. *Q.E.D*

Proof of Lemma 3:

Note that

$$\begin{aligned}\sum_x R(x,\tau)\left(f(x\mid a=1)-f(x\mid a=0)\right) &= \left(\frac{3}{2}\sigma-1\right)\left[1-H\left(\frac{1}{\tau}\left(1-q(0)\right)\right)\right] \\ &+ \left(1-\frac{3}{2}\sigma\right)\left[1-H\left(\frac{1}{\tau}\left(1-q(1)\right)\right)\right]+\frac{\sigma}{2} \\ &= \frac{\sigma}{2}+\left(1-\frac{3}{2}\sigma\right)\left[H\left(\frac{1}{\tau}\left(1-q(0)\right)\right)-H\left(\frac{1}{\tau}\left(1-q(1)\right)\right)\right]\end{aligned}$$

Note $\sigma\in(0,\frac{2}{3})$. Since $q(0)<q(1)$, $H\left(\frac{1}{\tau}(1-q(0))\right)>H\left(\frac{1}{\tau}(1-q(1)\right)$, proving the first

part of the Lemma.

Moreover, since $H(.)$ is weakly convex (h is non-decreasing),

$$\begin{aligned}&\frac{\partial}{\partial\tau}\left[\sum_x R(x,\tau)\left(f(x\mid a=1)-f(x\mid a=0)\right)\right]\\&=\frac{1}{\tau^2}\left(1-\frac{3}{2}\sigma\right)\times\left[(1-q(1))h\left(\frac{1}{\tau}\left(1-q(1)\right)\right)-(1-q(0))h\left(\frac{1}{\tau}\left(1-q(0)\right)\right)\right]<0\end{aligned}$$

Q.E.D

To prove Proposition 5, I consider the function

$$N(\alpha_*,\tau)=\sum_x R(x,\tau)\left(f(x\mid a=1)-f(x\mid a=0)\right).$$

Then the indifference condition solved by α_* is given by

$$\alpha_*=G\left(\frac{P(\alpha_*)}{d(\tau)N(\alpha_*,\tau)}\right).$$

I prove Proposition 4, and I also demonstrate the existence of a unique τ_* when the cumulative distribution function $H(.)$ is distributed according to a uniform distribution.

Proof of Proposition 5, along with case the where $H\sim\mathcal{U}[0,1]$.

Proving Part 1 of Proposition 5 is straightforward. The proof is an adaptation of the proof of Proposition A.1 where $N(\alpha_*,\tau)$ serves as $M(\alpha^*,\tau)$.

Similarly, Part 2 of Proposition 5 has a proof that is simply an adaptation of the proof of Proposition A.1, if one replaces the function N with M.

$\underline{H(.)\sim\mathcal{U}[\mathbf{0},\mathbf{1}]}$.

Now suppose $H(.)$ is distributed according to the uniform distribution. The net severity of repression is given by

$$N(\alpha,\tau)=\sum_x R(x)\left(f(x\mid a=1)-f(x\mid a=0)\right)=\frac{\sigma}{2}+\frac{1}{\tau}\left(\frac{\left(1-\frac{3}{2}\sigma\right)\left(2\left(\frac{1-\sigma}{\sigma}\right)-\frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)}{\left(\frac{\alpha}{1-\alpha}+\frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)\left(\frac{\alpha}{1-\alpha}+2\left(\frac{1-\sigma}{\sigma}\right)\right)}\right)$$
$$=\frac{\sigma}{2}+\frac{1}{\tau}\mathbf{Q}(\alpha,\sigma)$$

where $\mathbf{Q}(\alpha,\sigma)=\left(\frac{\left(1-\frac{3}{2}\sigma\right)\left(2\left(\frac{1-\sigma}{\sigma}\right)-\frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)}{\left(\frac{\alpha}{1-\alpha}+\frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)\left(\frac{\alpha}{1-\alpha}+2\left(\frac{1-\sigma}{\sigma}\right)\right)}\right)$.

Denote $\mathbf{J}(x,\tau)\equiv\mathbf{J}(x)=x-G\left(\frac{P(x)}{d(\tau)N(x,\tau)}\right)$.

Suppose $g(c)<\xi_1$ for any $c\in[\underline{c},\overline{c}]$. By the implicit function theorem,

$$\frac{d\alpha_*}{d\tau}=-\frac{\frac{\partial\mathbf{J}(\alpha_*;\tau)}{\partial\tau}}{\left.\frac{\partial\mathbf{J}(x;\tau)}{\partial x}\right|_{x=\alpha_*}}=\frac{\frac{\partial\left[G\left(\frac{P(\alpha_*)}{d(\tau)\left(\frac{\sigma}{2}+\frac{1}{\tau}\mathbf{Q}(\alpha_*,\sigma)\right)}\right)\right]}{\partial\tau}}{\left.\frac{\partial\mathbf{J}(x;\tau)}{\partial x}\right|_{x=\alpha_*}}. \tag{12}$$

Since $\mathbf{J}(x;\tau)$ is strictly increasing as a function of x, $sign\left(\frac{d\alpha_*}{d\tau}\right)=sign\left(\frac{\partial\left[G\left(\frac{P(\alpha_*)}{d(\tau)\left(\frac{\sigma}{2}+\frac{1}{\tau}\mathbf{Q}(\alpha_*,\sigma)\right)}\right)\right]}{\partial\tau}\right)$.

Denote $\Gamma\equiv\Gamma(\alpha_*;\tau)=\frac{P(\alpha_*)}{d(\tau)\left(\frac{\sigma}{2}+\frac{1}{\tau}\mathbf{Q}(\alpha_*,\sigma)\right)}$; where $\mathbf{Q}(\alpha;\sigma)=\left(\frac{\left(1-\frac{3}{2}\sigma\right)\left(2\left(\frac{1-\sigma}{\sigma}\right)-\frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)}{\left(\frac{\alpha}{1-\alpha}+\frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)\left(\frac{\alpha}{1-\alpha}+2\left(\frac{1-\sigma}{\sigma}\right)\right)}\right)$.

$$\begin{aligned}\frac{\partial G(\Gamma(\alpha_*,\tau))}{\partial\tau} &= -\frac{P(\alpha_*)}{\left[d(\tau)\left(\frac{\sigma}{2}+\frac{1}{\tau}\mathbf{Q}(\alpha_*,\sigma)\right)\right]^2}g\left(\Gamma(\alpha_*,\tau)\right)\left[d'(\tau)\left(\frac{\sigma}{2}+\frac{1}{\tau}\mathbf{Q}(\alpha_*,\sigma)\right)-\frac{1}{\tau^2}\mathbf{Q}(\alpha_*,\sigma)d(\tau)\right]\\ &= \Gamma(\alpha_*;\tau)\left(\frac{1}{\tau}\right)g(\Gamma(\alpha_*,\tau))\left[-\tau\frac{d'(\tau)}{d(\tau)}+\frac{\mathbf{Q}(\alpha_*,\sigma)}{\mathbf{Q}(\alpha_*,\sigma)+\tau\frac{\sigma}{2}}\right]\\ &= \Gamma(\alpha_*;\tau)\left(\frac{1}{\tau}\right)g(\Gamma(\alpha_*,\tau))\times\Sigma_1(\alpha_*,\tau).\end{aligned} \tag{13}$$

Where

$$\Sigma_1(\alpha_*,\tau)=-\tau\frac{d'(\tau)}{d(\tau)}+\frac{\mathbf{Q}(\alpha_*,\sigma)}{\mathbf{Q}(\alpha_*,\sigma)+\tau\frac{\sigma}{2}}=-V(\tau)+\frac{L_3(\alpha_*,\tau)}{L_3(\alpha_*,\tau)+L_4(\alpha_*,\tau)},$$

$V(\tau) = \tau\frac{d'(\tau)}{d(\tau)}$, $L_3 = \mathbf{Q}(\alpha_*, \sigma)$ and $L_4(\alpha_*, \tau) = \tau\frac{\sigma}{2}$. It turns out $sign\left(\frac{\partial G(\Gamma(\alpha_*,\tau))}{\partial \tau}\right) = sign\left(\Sigma_1(\alpha_*, \tau)\right)$. Moreover, $\Sigma_1(\alpha_*, \tau = \underline{\tau}) = \frac{L_3(\alpha_*,\tau)}{L_3(\alpha_*,\tau)+L_4(\alpha_*,\tau)} > 0$; $\Sigma_1(\alpha_*, \tau = \overline{\tau}) < 0$, since $d'(\overline{\tau})$ is extremely large. Therefore, there exists τ_* such that for $\tau \leq \tau_*$, $\Sigma_1(\alpha_*, \tau) > 0$, and for $\tau > \tau_*$, $\Sigma_1(\alpha_*, \tau) < 0$.

I show that τ_* is unique. I derive the derivative of Σ_1 with respect to τ, taking into account the fact that α_* is also a function of τ. Since $\frac{\partial L_3}{\partial \tau} = \frac{\partial[\mathbf{Q}(\alpha_*,\sigma)]}{\partial \tau} = 0$, observe that

$$\begin{aligned}
\frac{d\Sigma_1(\alpha_*, \tau)}{d\tau} &= -V'(\tau) + \frac{\frac{\partial L_3}{\partial \alpha}\frac{d\alpha_*}{d\tau}\left[L_3 + L_4\right] - \left[\frac{\partial L_4}{\partial \tau} + \frac{\partial L_3}{\partial \alpha}\frac{d\alpha_*}{d\tau}\right] L_3}{\left[L_3 + L_4\right]^2} \\
&= -V'(\tau) + \frac{L_4\left(\frac{\partial L_3}{\partial \alpha}\right)\frac{d\alpha_*}{d\tau} - L_3\frac{\partial L_4}{\partial \tau}}{\left[L_3 + L_4\right]^2} \\
&= -V'(\tau) - \frac{\left(\frac{\sigma}{2}\right)\mathbf{Q}(\alpha_*, \sigma)}{\left[\tau\left(\frac{\sigma}{2}\right) + \mathbf{Q}(\alpha_*, \sigma)\right]^2} + \frac{\tau\left(\frac{\sigma}{2}\right)\frac{\partial(\mathbf{Q}(\alpha_*,\sigma))}{\partial \alpha}}{\left[\tau\left(\frac{\sigma}{2}\right) + \mathbf{Q}(\alpha_*, \sigma)\right]^2}\frac{d\alpha_*}{d\tau} \qquad (14)
\end{aligned}$$

Now consider the term $\mathbf{Q}(\alpha_*, \sigma)$.

$$\mathbf{Q}(\alpha_*, \sigma) = \left(\frac{\left(1 - \frac{3}{2}\sigma\right)\left(2\left(\frac{1-\sigma}{\sigma}\right) - \frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)}{\left(\frac{\alpha}{1-\alpha} + \frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)\left(\frac{\alpha}{1-\alpha} + 2\left(\frac{1-\sigma}{\sigma}\right)\right)}\right) = \frac{\left(1 - \frac{3}{2}\sigma\right)\left(2\left(\frac{1-\sigma}{\sigma}\right) - \frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)}{\left(\mathbf{A}(\alpha) + \frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)\left(\mathbf{A}(\alpha) + 2\left(\frac{\sigma}{1-\sigma}\right)\right)},$$

with $\mathbf{A}(\alpha) = \frac{\alpha}{1-\alpha}$. Since $\mathbf{A}'(\alpha) > 0$, observe that

$$\frac{\partial(\mathbf{Q}(\alpha_*, \sigma))}{\partial \alpha} = -\frac{\left(1 - \frac{3}{2}\sigma\right)\left(2\left(\frac{1-\sigma}{\sigma}\right) - \frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)}{\left[\left(\mathbf{A} + \frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right)\left(\mathbf{A} + 2\left(\frac{\sigma}{1-\sigma}\right)\right)\right]^2} \times \left[\mathbf{A}'\left(\mathbf{A} + \frac{1}{2}\left(\frac{\sigma}{1-\sigma}\right)\right) + \mathbf{A}'\left(\mathbf{A} + 2\left(\frac{\sigma}{1-\sigma}\right)\right)\right]$$

is negative. Therefore, in expression (14), $\frac{\tau\left(\frac{\sigma}{2}\right)\frac{\partial(\mathbf{Q}(\alpha_*,\sigma))}{\partial \alpha}}{\left[\tau\left(\frac{\sigma}{2}\right)+\mathbf{Q}(\alpha_*,\sigma)\right]^2}$ is negative, as well as $-V'(\tau)$ and $-\frac{\left(\frac{\sigma}{2}\right)\mathbf{Q}(\alpha_*,\sigma)}{\left[\tau\left(\frac{\sigma}{2}\right)+\mathbf{Q}(\alpha_*,\sigma)\right]^2}$.

Furthermore, from expression (12) and (13),

$$\frac{d\alpha_*}{d\tau} = \frac{\frac{\partial\left[G\left(\frac{P(\alpha_*)}{d(\tau)\left(\frac{\sigma}{2}+\frac{1}{\tau}\mathbf{Q}(\alpha_*,\sigma)\right)}\right)\right]}{\partial\tau}}{\left.\frac{\partial\mathbf{J}(x;\tau)}{\partial x}\right|_{x=\alpha_*}} = \frac{\Gamma(\alpha_*;\tau)\left(\frac{1}{\tau}\right)g(\Gamma(\alpha_*,\tau)) \times \Sigma_1(\alpha_*,\tau)}{\left.\frac{\partial\mathbf{J}(x;\tau)}{\partial x}\right|_{x=\alpha_*}}. \tag{15}$$

Replacing expression (15) into equation (14), I obtain

$$\begin{aligned}\frac{d\Sigma_1(\alpha_*,\tau)}{d\tau} &= -V'(\tau) - \frac{\left(\frac{\sigma}{2}\right)\mathbf{Q}(\alpha_*,\sigma)}{[\tau\left(\frac{\sigma}{2}\right)+\mathbf{Q}(\alpha_*,\sigma)]^2} \\ &+ \frac{\tau\left(\frac{\sigma}{2}\right)\frac{\partial(\mathbf{Q}(\alpha_*,\sigma))}{\partial\alpha}}{[\tau\left(\frac{\sigma}{2}\right)+\mathbf{Q}(\alpha_*,\sigma)]^2}\left(\frac{\Gamma(\alpha_*;\tau)\left(\frac{1}{\tau}\right)g(\Gamma(\alpha_*,\tau))}{\left.\frac{\partial\mathbf{J}(x;\tau)}{\partial x}\right|_{x=\alpha_*}}\right) \times \Sigma_1(\alpha_*,\tau).\end{aligned}$$

Based on the analysis above, since $\frac{\Gamma(\alpha_*;\tau)\left(\frac{1}{\tau}\right)g(\Gamma(\alpha_*,\tau))}{\left.\frac{\partial\mathbf{J}(x;\tau)}{\partial x}\right|_{x=\alpha_*}} > 0$, $\frac{d\Sigma_1(\alpha_*,\tau)}{d\tau} \leq 0$ when $\Sigma_1 > 0$. Therefore, Σ_1 has a single-crossing from above. Hence, τ_* exists and is unique. *Q.E.D*

Proof of Proposition 6

The proof is almost similar to the proof of Proposition 2. First of all it is clear that in this framework, $R(s+1,\tau) > R(s,\tau)$. Furthermore, denote

$$\begin{aligned}W(\alpha;\tau) &= \sum_{s=1}^{n-1}\left[H\left(\frac{1}{\tau}\left(1-\frac{s(1-\lambda)}{n}\right)\right) - H\left(\frac{1}{\tau}\left(1-\frac{(s+1)(1-\lambda)}{n}\right)\right)\right]\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s} \\ &+ 1 - H\left(\frac{1}{\tau}\left(1-\frac{1-\lambda}{n}\right)\right),\end{aligned}$$

Then the partial derivative with respect to τ is given by

$$
\begin{aligned}
&= \frac{1}{\tau^2}\sum_{s=1}^{n-1}\left[\left(1-\frac{(s+1)(1-\lambda)}{n}\right)h\left(\frac{1}{\tau}\left(1-\frac{(s+1)(1-\lambda)}{n}\right)\right)\right.\\
&- \left.\left(1-\frac{s(1-\lambda)}{n}\right)h\left(\frac{1}{\tau}\left(1-\frac{s(1-\lambda)}{n}\right)\right)\right]\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}\\
&+ (1-\alpha)^{n-1}\frac{1}{\tau^2}\left(1-\frac{1-\lambda}{n}\right)h\left(\frac{1}{\tau}\left(1-\frac{1-\lambda}{n}\right)\right),
\end{aligned}
$$

which is negative if $G(0) > \frac{(1-\lambda)\underline{h}}{(1-\lambda)\underline{h}+(n-1+\lambda)\overline{h}}$. *Q.E.D*

Online Appendix: Rational Deterrence in Internal Armed Conflicts.

I show that if the value of retaliation is common knowledge, the main result can be preserved. However, a common knowledge value of retaliation makes the analysis less tractable. The threshold of retaliation is now known with certainty, but it is discontinuous. This discontinuity implies that the severity of repression is a step function.

Proposition OA.1 (Common Knowledge Value of Retaliation.) *Suppose Defender's benefit of retaliation, v, is known with certainty. Then, when $v > v^*$ (Defender is overzealous), civilians overreact to the threat of repression if $G(0) > \kappa^\dagger$. Moreover, if v is small, the threat of repression always negates dissent, and deterrence is successful.*

Proof of Proposition OA.1.

I assume that v is common knowledge. It is clear that Defender retaliates if the number of attacks s is greater or equal to $n\left(1-\tau v\right) > 0$. Let S^i be the number of dissidents except player i. A dissident receives

$$\frac{1}{d(\tau)}\left[\sum_{s=0}^{n-1}\pi(s+1)\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}\right]-c_i\Pr\left[S^i\geq k(\tau)-1\right],$$

while a civilian receives

$$\frac{\gamma}{d(\tau)}\left[\sum_{s=0}^{n-1}\pi(s)\binom{n-1}{s}\alpha^s(1-\alpha)^{n-1-s}\right]-c_i\Pr\left[S^i\geq k(\tau)\right].$$

The threshold of retaliation $k^v(\tau) = \lceil n\left(1-\frac{\tau}{\overline{\tau}}v\right)\rceil \geq 1$ is the ceiling of the number $n\left(1-\tau v\right)$, the smallest integer that is greater than or equal to $n\left(1-\tau v\right)$. Thus, the threshold of retaliation is discontinuous in the value of retaliation and the threat of repression. The net expected payoff then reads as

$$\frac{P(\alpha)}{d(\tau)} - c_i \times \Pr\left[S^i = k^v(\tau) - 1\right],$$

where $P(\alpha) = \sum_{s=0}^{n-1} \binom{n-1}{s}(\pi(s+1) - \gamma\pi(s))\alpha^s(1-\alpha)^{n-1-s}$. An individual cares about the probability that their participation changes the retaliation decision. Hence, citizens care about the probability of being pivotal in the success of the collective project (if $\gamma = 1$), and also care about the probability of being pivotal in facing repression. Despite this complication, I show that the main result is preserved.

First, notice that $k^{v=0}(\tau) = n$. Thus, when the value of retaliation is small and the Defender observes enormous restraint, it retaliates if the evidence suggests that the suspect is guilty beyond a reasonable doubt. Defender's belief must be extremely accurate. In this case, the net expected payoff when $v = 0$ reads as follows

$$\frac{P(\alpha)}{d(\tau)} - c_i \Pr\left[S^i = n - 1\right].$$

Hence, deterrence is always successful if Defender observes restraint. In the other extreme case, assume $v = 1$.

It is important to see that the net severity of repression $\Pr\left[S^i = k^{v=1}(\tau) - 1\right]$ is decreasing in τ if the society is hostile $G(0) > \kappa^\dagger$. While abusing notation (since $k^{v=1}(\tau)$ is a step function and is not continuous) we have

$$\frac{\partial}{\partial \tau}\left(\Pr\left[S^i = k^{v=1}(\tau) - 1\right]\right) = \frac{dk^{v=1}(\tau)}{d\tau} \times \frac{\partial}{\partial k}\left(\Pr\left[S^i = k - 1\right]\right)\Big|_{k=k^{v=1}(\tau)}.$$

Note that $S^i \sim \mathcal{B}(n-1, \alpha)$ indicates that $\frac{\partial}{\partial k}\left(\Pr\left[S^i = k - 1\right]\right)$ is positive and negative (due to the log-concavity of the Binomial distribution). If $k^{v=1}(\tau) - 1$ is below the modal value of the Binomial, which is $\lceil n\alpha \rceil$, then

$$\frac{\partial}{\partial k}\left(\Pr\left[S^i = k-1\right]\right)\Big|_{k=k^{v=1}(\tau)} > 0.$$

As the ceiling is a non-decreasing function, $k^{v=1}(\tau) - 1$ is below the mode if and only if $n(1-\tau) - 1 < n\alpha$, which implies $\alpha > \frac{n(1-\tau)-1}{n}$. This condition will be satisfied if $G(0) > \left(\frac{n(1-\tau)-1}{n}\right) \equiv \kappa^{\dagger}$. Moreover, since $\frac{dk^{v=1}(\tau)}{d\tau} < 0$ (again by abusing notation), $\frac{\partial}{\partial \tau}\left(\Pr\left[S^i = k^{v=1}(\tau) - 1\right]\right) < 0$ if the society is hostile. We have shown that

$$\frac{\partial}{\partial \tau}\left[\underbrace{\Pr[S^i + 1 \geq k^{v=1}(\tau)]}_{dissident}\right] < \frac{\partial}{\partial \tau}\left[\underbrace{\Pr[S^i \geq k^{v=1}(\tau)]}_{civilian}\right],$$

and civilians overreact when an overzealous Defender executes the threat of repression. *Q.E.D*

References

Baker, Peter. 2002. "GIs Battle "Ghosts" in Afghanistan. Search for Elusive Enemy Frustrates Americans." *Washington Post, May 16, A1.*

Baliga, Sandeep, Ethan Bueno De Mesquita, and Alexander Wolitzky. 2020. "Deterrence with Imperfect Attribution." *American Political Science Review* 114 (4): 1155–1178.

Bates, Robert. 1983. *Modernization, Ethnic Competition, and the Rationality of Politics in Contemporary Africa. State versus Ethnic Claims :African Policy Dilemmas. Editors, Donald Rothchild and Victor Olunsorola. 152-171.* Boulder,CO: Westview Press.

Bayly, A. Christopher. 1996. *Empire and Information: Intelligence Gathering and Social Communication in India 1780–1870.* Cambridge : Cambridge University Press.

Berman, Eli, N. Jacob Shapiro, and H. Joseph Felter. 2011. "Can Hearts and Minds be bought? Economics of Counterinsurgency in Iraq." *Journal of Political Economy.* 119 (4): 766–819.

Blaydes, Lisa. 2018. *State of Repression: Iraq under Saddam Hussein.* Princeton University Press.

Brandsch, Jürgen. 2020. *Killing Civilians in Civil War : The Rationale of Indiscriminate Violence.* Lynne Rienner Publishers.

Bueno De Mesquita, Bruce, P. David Myatt, Alastair Smith, and A. Scott Tyson. 2024. "The Punisher's Dilemma." *Journal of Politics* 86 (2).

Chassang, Sylvain, and Gerard Padró I Miquel. 2010. "Conflict and Deterrence under Strategic Risk." *The Quarterly Journal of Economics* 125 (4): 1821–1858.

Cooper, Matthew. 1979. *The Nazi War against Soviet Partisans, 1941–1944.* New York: Stein / Day.

Correa, Sofia, Gaétan Nandong, and Mehdi Shadmehr. 2025. "Grievance shocks and coordination in protest." *American Journal of Political Science* 69 (1): 341–354.

Curtice, Travis B., and Behlendorf Brandon. 2021. "Street-level Repression: Protest, Policing, and Dissent in Uganda." *Journal of Conflict Resolution* 65 (1): 166–194.

Davenport, Christian. 2007. "State Repression and Political Order." *Annual Review of Political Science* 10:1–23.

Egorov, Georgy, and Konstantine Sonin. 2011. "Dictators and their Viziers: Endogenizing the Loyalty-Competence Trade-off." *Journal of the European Economic Association* 9 (5): 903–930.

Filkins, Dexter. 2005. "The Fall of the Warrior King." *New York Times Magazine,* 23 October.

Fransisco, Ronald. 1995. "The Relationship between Coercion and Protest: An Empirical Evaluation in Three Coercive States." *The Journal of Conflict Resolution* 39 (2): 263–282.

Gibilisco, Michael. 2021. "Decentralization, Repression, and Gambling for Unity." *The Journal of Politics* 83 (4).

Greitens, Sheena Chesnut. 2016. *Dictators and their Secret Police Coercive Institutions and State Violence.* Cambridge University Press.

Gurr, Ted Robert. 1970. *Why Men Rebel.* Princeton: Princeton University Press.

Kalyvas, N. Stathis. 2006. *The Logic of Violence in Civil War.* New York: Cambridge University Press.

Lagunoff, Roger. 2024. "A Dynamic Model of Authoritarian Social Control." *Review of Economic Studies,* no. 00, 1–37.

Lebas, Adrienne, and E. Lauren Young. 2024. "Repression and Dissent in Moments of Uncertainty: Panel Data Evidence from Zimbabwe." *American Political Science Review* 118 (2): 584–601.

Lewis, I. Janet. 2020. *How Insurgency Begins Rebel Group Formation in Uganda and Beyond.* Cambridge University Press.

Lichbach, Mark Irvin. 1987. "Deterrence or Escalation? The Puzzle of Aggregate Studies of Re-pression and Dissent." *Journal of Conflict Resolution 31:266-97,* no. 31, 266–97.

Liu, Howard, and Christopher M. Sullivan. 2021. "And the Heat Goes On: Police Repression and the Modalities of Power." *Journal of Conflict Resolution* 65 (10): 1657–1679.

Lyall, Jason. 2009. "Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya." *The Journal of Conflict Resolution* 53 (3): 331–362.

McAdam, Dough. 1986. "Recruitment in High-Risk Activism: The Case of Freedom Summer." *American Journal of Sociology* 92 (1): 64–90.

Nandong, Gaétan. 2025. "State Surveillance and Collective Dissent." *Working paper* https://drive.google.com/file/d/1cqCmuqEs10Cvr_yUirq1ZTnXpagf6mJe/view.

Nugent, Elizabeth R. 2020. "The Psychology of Repression and Polarization." *World Politics* 72 (2): 291–334.

Opp, Karl-Dieter, and Wolfgang Roehl. 1990. "Repression, Micromobilization, and Political Protest." *Social Forces* 69 (2): 521–47.

Pearlman, Wendy. 2013. "Emotions and the Microfoundations of the Arab Uprisings." *Perspectives on Politics* 61 (2): 387–409.

Pop-Eleches, Grigore, and Lucan A. Way. 2023. "Toward a Theory of Revolution." *American Journal of Political Science* 67 (2): 456–471.

Rasler, Karen. 1996. "Concessions, Repression, and Political Protest in the Iranian Revolution." *American Sociological Review,* no. 61, 132–52.

Ritter, Emily Hencken, and Courtenay R. Conrad. 2016. "Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression." *American Political Science Review* 110 (1): 85–99.

Rosenfeld, Bryn. 2025. "Risk Attitudes and Political Participation Under Autocracy." *Working paper* https://drive.google.com/file/d/1huSkoLDWaWc6kESoC-2wQpMZlasSIt22/view.

Rozenas, Arturas. 2020. "A Theory of Demographically Targeted Repression." *Journal of Conflict and Rezolution* 64 (7-8): 1254–1278.

Shadmehr, Mehdi. 2014. "Mobilization, Repression, and Revolution: Grievances and Opportunities in Contentious Politics." *Journal of Politics* 76 (3): 621–35.

Shadmehr, Mehdi, and Dan Bernhardt. 2011. "Collective Action with Uncertain Payoffs: Coordination, Public Signals and Punishment Dilemma." *American Political Science Review* 105 (4): 829–859.

Stavro, Martin, and Ryan M. Welch. 2024. "Does Police Militarization Increase Repression?" *Journal of Conflict Resolution* 68 (5): 964–992.

Sullivan, M. Christopher. 2016. "Political Repression and The Destruction of Dissident Organizations: Evidence from the Archives of the Guatemalan National Police." *World Politics* 68 (4): 645–676.

Svolik, Milan W. 2012. *The Politics of Authoritarian Rule.* Cambridge, UK: Cambridge University Press.

Tilly, Charles. 1978. *From Mobilization to Revolution.* Reading, MA: Addison-Wesley.

Trinquier, Roger. 2006. *Modern Warfare: A French View of Counterinsurgency.* Harvard University Press.

Tyson, Scott A. 2018. "The Agency Problem Underlying Repression." *The Journal of Politics* 80 (4): 1297–1310.

U.S.Army. 2007. *Field Manual 3-24, Counterinsurgency Field Manual.* Chicago: Univ. Chicago Press.

Young, Lauren E. 2019. "The psychology of state repression: Fear and dissent decisions in Zimbabwe." *American Political Science Review* 113 (1): 140–155.