Backlash*

Emily Hencken Ritter[†], Jessica S. Sun[‡], Scott A. Tyson[§]

April 9, 2024

This is a working draft. Please do not cite or circulate without author permission.

For the latest version, please follow this link.

Abstract

In this article, we highlight the varied and conflicting ways that scholars conceptualize backlash mobilization in response to repression, drawing meta-lessons from empirical and theoretical research in political science to develop a clarifying formal model as to how repression can cause backlash mobilization. We present the results of a review of published political science articles from 2000-2023 to understand the varied findings and mechanisms considered in this area. The results of that review inform the assumptions of a formal model to examine how each of three posited mechanisms—anger, logistical efficiency, and learning—explain when and how repression can cause mobilization that would not otherwise occur. The model reveals where these mechanisms yield conflicting predictions and also how they may co-occur, highlighting that the empirical observation of backlash mobilization cannot necessarily distinguish the mechanism causing the outcome.

Keywords: Backlash, Repression

[*] Prepared for presentation at the Political Economy of Conflict workshop at Vanderbilt University, April 16, 2024, and the Comparative Politics Annual Conference at Washington University in St. Louis, April 26, 2024. We thank the generous political science faculty and students at Louisiana State University and Emory University for their helpful comments and suggestions. We also thank Pearce Edwards, Zhaotian Luo, Pablo Montagnes, and Chris Sullivan for their thoughts on the project.

[†] Associate Professor, Vanderbilt University; emily.h.ritter@vanderbilt.edu

[‡] Assistant Professor, Emory University; jessica.sun2@emory.edu

[§] Associate Professor, University of Rochester; styson2@ur.rochester.edu

Introduction

Although governments repress populations to *quell* mobilization and dissent actions, they can instead *catalyze* widened participation or intensified efforts to mobilize and challenge the regime. If, after a government represses a movement or a population, there is an increase in the size, frequency, severity, or violence of collective dissent actions, scholars generally label this empirical phenomenon *backlash*. However, the term backlash is used loosely in political science to indicate all sorts of negative responses to a government action, from enraged public opinion to decreased voter support, from media critiques to terror actions. When we limit the search for articles that focus on the backlash to repression, scholars persistently use the term to mean a variety of behaviors and mechanisms. What is backlash to repression? And perhaps more importantly, when and how does repression cause backlash mobilization?

As the first objective in this article, we highlight the varied and conflicting ways that scholars conceptualize backlash mobilization, drawing meta-lessons from empirical and theoretical research in political science to develop a clarifying formal model as to how repression can cause backlash mobilization. We start by cataloging articles studying the correlation between repression and backlash mobilization published over the past two decades in the three general political science journals and related publications in comparative and international relations journals. We code these articles according to (1) the mechanism they describe as connecting repression to backlash, (2) whether the named mechanism is said to cause or deter backlash, and (3) how backlash measured empirically.

This review of scholarship illustrates the wide variation in how scholars think of backlash and relate it to repression. A large portion of this scholarship refers to backlash as an *empirical phenomenon*, where observable government repression is correlated with an increase in negative popular consequences for the government afterward. The backlash in different empirical analyses takes the form of increased participation in dissent actions, increased frequency or severity of dissent actions, or public disapproval of the government measured by attitudes. The actors who join, act, or disapprove are sometimes the dissident group who was repressed, and other times they are bystanders observing that the government repressed dissidents.

Scholars also refer to backlash as the result of a *mechanism*, where the observation of government repression causes a group or person who would not have taken an action against the state to do so. We categorize the mechanisms advocated in published scholarship to explain backlash as three types of explanation: emotion, logistic efficiency, and learning. These different mechanisms pepper the published scholarship on backlash, and none dominates the others as a consensus explanation. Researchers state that these mechanisms lead to backlash outcomes, but the different pathways yield different and sometimes conflicting empirical implications. Additionally, the mechanisms are mainly assumed or asserted without direct testing or logical examinations. Do bystanders become angry when their government represses protesters? We know that their government approval rating decreases, but scholars do not measure their emotions. Is it easier to recruit participants to a movement after repression? We know that participation sometimes increases and sometimes decreases, but we do not know what makes the change possible. Scholars also posit that bystanders and activists can learn things, but they rarely demonstrate empirically what they have learned.

If the phenomenon scholars call backlash is an empirical outcome, we do not know what causes the increased mobilization or action. The explanations are too numerous and not pushed for realistic logic. Moreover, the explanations can yield different results.

If backlash is a mechanism, we cannot identify the mechanism at work without establishing the counterfactual and then predicting the change in mobilization caused by repression. Do these three mechanisms logically cause changed behavior, or would the increased mobilization have happened without repression such that it is not actually backlash? Furthermore, could the three mechanisms coexist such that the observation of backlash cannot distinguish the mechanism at work?

The second objective of this article is to formalize the necessary conditions supporting the three standard arguments that scholars pose as to how repression causes backlash mobilization. We specify a formal model, with both a specific and a general functional form, that allows us to state formally what must logically be true for each mechanism to actually trigger joiners and actions that would not have occurred in the counterfactual without repression. For mobilization to qualify as backlash, there must be participation and action that would not have occurred if activists and bystanders had not observed repression and made an assessment as to what it means for them. We describe the necessary assumptions for that to occur under each potential explanation.

The third objective is to illustrate the conditions and boundaries as to when these mechanisms can plausibly cause backlash. By combining the mechanisms in one framework and model, we can identify when the mechanisms complement or substitute for one another. In particular, we highlight strategic complementarities that underlie an observed increase in mobilized dissent after repression. For example, when a bystander observes repression and becomes angry, this not only makes the bystander want to participate in backlash activities, but it also makes the activist more confident that it will be worthwhile to invest more effort.

The model also allows us to pin empirical observations of backlash to the counterfactual conditions that allow the causal mechanisms to occur...

Backlash Scholarship Trends, 2004–2023

To obtain a picture of the scholarly (lack of) consensus on backlash mobilization, we conducted a coded literature review of articles published on the topic from 2004 to 2023 that were published in the commonly-cited "top three" general-interest political science journals (APSR, AJPS, and JOP)[1] as well as other articles published in political science

[1] These acronyms refer to the *American Political Science Review,* the *American Journal of Political Science,* and the *Journal of Politics.* The inventory of articles includes those published in a volume of the journal from January 1, 2004, to December 31, 2023. It does not include articles that were available online only from the journals during that period.

subfield journals that are frequently cited according to Google Scholar citation trends as relevant to the topic.[2] We classified articles as relevant when the words *repression* and *backlash* were used anywhere in the article, and a review of the abstract revealed that it examines negative political responses to government repression. We also included articles that use repression as an independent variable that precedes a dissent response, regardless of whether the authors classify that pattern as backlash. These inclusion rules yielded XXXX articles, where XXX are from subfield-specific journals. Figure 1 presents a histogram of where the articles were published over time.

[**Descriptive statistics over time**]

We coded these articles as to how they characterize backlash as a mechanism and as an empirical outcome. In all articles in the set, the dependent variable is a response to a government that has repressed some dissidents or potential dissidents from the population, but the studies differ as to what the response is. We categorized the dependent variable of interest, whether conceptualized in a theory or measured in an empirical study, into two types. Scholars describe a response to repression as taking the form of either a dissent action (nonviolent or violent) or a change to public opinion (including government approval ratings and vote patterns). Figure 2 presents the share of articles that study each observable dependent variable as a backlash response to repression.

[**Descriptive statistics by dependent variable & direction—bar chart of total number of dissent / opinion, with percentage bars of predicted effect direction next to the respective bar**]

Figure 2 also presents the directional findings of this body of scholarship as a proportion of the number of articles using each type of dependent variable. XXXX articles study the concept of backlash as dissent responses to repression, such as an increase in dissent

[2] Though the inclusion rule application creates a systematic sample from the top three journals, the subfield journal inclusion rule is not systematically inclusive. There are important studies that we have most certainly missed in our effort to track scholarly trends on the topic.

events, a surge in mobilized participants, or an escalation in violence. XXX% of these studies argue and find that repression incites an increase in the size of the mobilized population or the severity or frequency of dissent actions; XXX% identify conditions where repression instead deters mobilized backlash.[3] A smaller number (XXX) of published articles study how repression affects public attitudes or popular expressed support for the government; XX% find that repression increases public support for the regime, and XX% find a negative effect or decrease in support for the regime, which the authors refer to as a form of backlash.

Actors lists the political decision-makers the authors present as relevant for the response to repression. The list includes the government and its authorized agents (repressor), a dissident or mobilized dissenting group (activist), and a bystander or bystanders deciding how to respond. We characterize the role of each player in the backlash interaction by their contribution to it. The government is the subject or initiator of repression and the activist is its repressed object. The bystander is the subject or initiator of backlash, and the government is the object or recipient of backlash. Figure 3...

[**Descriptive statistics by actor type**]

In all of the articles in our inventory, there is a government principal or agent who represses a target group. Government actors are usually presented as the government, state, regime, or leader, treating the repressive government as a unitary actor that can reliably send an order to repress and agents will carry it out. Nine articles relax the unitary actor assumption by discussing the principal (regime) who orders or condones repression as distinct from the agent who carries it out. In all of these cases, the agent is a security agent, whether police, military, or immigration enforcers. The bystander attributes the responsibility for the repression to the government or its agents, where observing police violence against protesters would lead the bystander to judge the government authorities as a whole to be responsible.

[3] We also include a middle category where the study finds that repression sometimes increases and other times decreases dissent activities.

The object or target of the government's repression is a civilian actor or group who has executed a collective dissent action or represents a potential threat to the government's authority. This reaction is common: When behavior threatens the political system, its authorities, its territory, or its policies, governments respond frequently with repression (Davenport 2007). The object of repression is typically a mobilized dissent or opposition group, where a group has organized to work together and engage in some behavior that threatens the government. Examples include supporters of a political opponent (Tertytchnaya 2023), terror organizations (Freedman and Klor 2023), or activists in a civil society organization or social movement (Chenoweth and Stephan 2011; Steinert and Dworschak 2023). Other scholars characterize backlash as a response to the government repressing an unorganized population identifiable by its ethnicity (Hatz 2019) or behavior (Thachil 2020; Eck et al. 2021), or repressing the citizenry as a whole (Wood et al. 2022; Loewenthal, Miaari, and Abrahams 2023). In other words, sometimes the government represses a small, targeted group, and other times it represses a diffuse population, and repressing with both foci is connected to varieties of backlash mobilization.[4]

Critically, studies of backlash include a bystander or bystanders who observe government authorities repressing a dissenting group and decide whether to take action. In most studies of backlash, the author focuses on an unaffiliated bystander from the general population who is stimulated into a decision upon learning about government repression. This could be a bystander at an event where police use violence against protesters (Reny and Newman 2021), a citizen who learns about government repression from news coverage or an informed source (Tertytchnaya 2023), or a voter reflecting on authorities' behavior when determining their vote choice (Graham and Svolik 2020). In these cases, the bystander(s) were not the direct object of repression, but repression affects them in a way that alters their behavior from what they would have done had they not learned about it. In a smaller subset of backlash mobilization studies, the subject of backlash is the dissent organization that was the target of government repression (Ritter and Conrad

[4] In two of the reviewed articles, backlash occurs when the government has repressed the media.

2016; Petrova 2022; Esberg and Siegel 2023); the experience of being repressed causes a change that leads the organization to increase their efforts to mobilize and enact another event of dissent of greater magnitude. Put differently, most studies assert that the sequence of backlash mobilization is (1) an activist group dissents, (2) a government represses the activist group, and (3) a bystander takes an action to punish the repressive government, but a minority of studies instead examine how the activist group responds to being repressed rather than what an unaffected bystander will do.

We coded each article in the inventory according to the explicit and implicit assumptions each author makes about *how* repression that supposedly leads to an increase in mobilization, negative opinion, or dissenting actions. Most studies explicitly name the mechanism they believe to be at work, but almost all assert the mechanism without directly examining it. The mechanisms that scholars claim to cause backlash mobilization after a repressive action generally fall into one of three categories (illustrated in Figure XXX): Emotion, logistic efficiencies, and information.

[Descriptive statistics by mechanism]

Emotion: Many scholars assert that the backlash results from an emotional reaction that drives a change in behavior. This is the idea that when a government is known to have repressed dissidents, either the dissidents or bystanders become angry and join the backlash or feel fear and leave the movement. Repression can inspire outrage and support for dissident claims, the bystander or activist to invest new or greater efforts to oppose the government. If the public perceives the repressive response to be unjust, illegitimate, or inappropriate, people become outraged. This outrage, directed at the government or its actions, engenders not only sympathy but also support for the repressed group and its claims (Koopmans 1997; Hess and Martin 2006; Aytaç, Schiumerini, and Stokes 2018; Hager and Krakowski 2022).

Logistic efficiencies: This mechanism is the idea that repressing dissent makes it easier for dissidents to mobilize and act with a larger base of support or otherwise increase efforts. Some forms of repression make it more difficult for bystanders and activists to continue to dissent. The state can directly inhibit opportunities to dissent, such as when governments ban protests and make it more risky to join an action (Ellefsen 2021; Tertytchnaya 2023). Repression of civil society organizations and increased barriers to immigrants have a similar limiting effect on the backlash, because closing these opportunities reduces the pool of people who would mobilize (Schon and Leblang 2021; Petrova 2022). Assassinating a leader of the dissent movement can cause the movement to collapse without an entrepreneur to lead it (Sullivan 2016). Other forms of repression make it easier to oppose the government, such as when exiled dissidents have a greater platform to gather resources and supporters once they have left the country (Esberg and Siegel 2023). Imprisoning political dissidents can build new networks for efforts outside of prisons, and their imprisonment serves as a focal point of empathy for dormant dissidents (Steinert and Dworschak 2023). In a simpler sense, government violence against a dissident group can serve as a go-ahead for the dissidents to increase their use of violence in turn (Lichbach 1987; Moore 1998; Chiang 2021).

Information Signaling: The remaining category of mechanisms captures a process of learning: Dissidents or bystanders learn something when they observe the repression of mobilized dissent, which alters what they believe to be the best response. This is the most common type of mechanism in our inventory of articles. They may learn that the government is more resolved than they expected. They may learn the government is willing to take illegal or illegitimate actions to control the population. They may learn the government is willing to repress people like them. And this new information tells them that backlash is the best response to government action.

We want to identify backlash as a mechanism, so that there is an explicit, logical, causal link where an observer sees the government repress a dissenting group and makes a calculation that leads them to join or increase collective dissent activities.We developed a formal model to examine how these mechanisms logically differ from each other and what they each imply about the observable world.

Modeling Mechanisms of Backlash

To understand how and why repression can cause changes in mobilized dissent—its level, intensity, or scale—we develop a model that distills the key elements of the empirical narrative. There are three actors, a government (G), an activist (A), and a bystander (B). The government and activist engage in multiple interactions, in each of which the activist undertakes some anti-government activity and the government represses. The government's relative capacity to repress dissent—relative to the activist's capacity to mobilize observers for anti-government activity—represents the state of the world θ, where higher realizations of θ imply the government has greater capacity relative to the activist. The bystander observes the interactions between the government and the activist and, after their first contest, can choose whether to join in subsequent dissent.

The game begins with a protest by the activist.[5] The government can choose to meet this protest with repression, choosing $v \in \{0, 1\}$, where $v = 1$ means the government represses the initial protest. The level of repression that is realized conditional on the government choosing $v = 1$ is a random variable r_0, which reflects that the government may not have perfect control over the implementation of repression by the security apparatus once repression is ordered. This strategic choice allows us to characterize when the government selects out of using repression, but we primarily focus on the case where $v = 1$ because this is the only scenario in which we can observe backlash—there is no response in the absence of initial repression.

The outcome of the activist's protest depends on both the underlying strength of the government and its repression choice: $\theta + r_0$. Outcomes that are favorable to the government are therefore more likely for more capable governments. The government and activist, as participants in this initial protest stage, observe both the outcome of the protest and the government's relative capacity. Thus, when the activist and government make their subsequent effort and repression choices, respectively, they do so with complete

[5] Allowing the activist to choose whether to protest in the first stage is a straightforward extension of the model. As long as expected costs are not large, protesting is a dominant strategy.

information.

The bystander, on the other hand, observes the initial protest outcome imperfectly. While the bystander can see the government's repression choice, r_0, she remains uncertain of the government's capacity for repression relative to the activist's capacity to mobilize. To capture this uncertainty, the bystander receives a signal of the outcome of the initial protest, $Q = \theta + r_0 + \varepsilon$, where ε is drawn from an absolutely continuous distribution function Ψ, with continuously differentiable density ψ, which satisfies the monotone likelihood ratio property (MLRP). The signal Q captures the *effectiveness* of repression from the perspective of the bystander. Denote her posterior belief by $\pi(\theta \mid Q)$. The bystander then chooses whether to demonstrate on the side of the activist, $d = 1$, or stay home, $d = 0$. If the bystander demonstrates, she pays $c_B > 0$, which reflects the cost of participation.

In the final stage the activist again mounts a challenge to the government, which is met with repression. Specifically, the activist chooses an effort level, e_1, and, at the same time, the government chooses a repression level, r_1. Total effort by the activist, then, is $e = e_1$, given initial protest is exogenous, and total repression is $r = r_0 + r_1$. Effort and repression are costly for the activist and government, where this cost is captured by k_A and k_G, respectively. The activist also suffers an additional cost from repression of the initial protest, paying $c_A(r_0)$, which is increasing in the level of first-stage repression.

The bystander's final-stage protest payoff is given by

$$u_B(d; e_1^*(d;\theta), r_0, r_1^*(d;\theta)),$$

conditional on her demonstration choice d.

The government's payoff is $u_G(d;\theta) - k_G$, and the activist's payoff is $u_A(d;\theta) - k_A - c_A(r_0)$ where $u_{i\in\{A,G\}}$ represents each side's utility from the final challenge stage and k_i is their cost from effort. Further, initial repression reduces the activist's gains from subsequent dissent, and this cost may be different from that born by the bystander.

To summarize, the timing of the game is:

1. Initial protest: activist stages a protest and the government represses;
2. Information: government and activist learn the state of the world and citizen observes the outcome with noise;
3. Demonstration: The bystander elects whether to participate in support of the activist;
4. Challenge: The activist challenges the government, who represses resistance.

In the last stage, the activist solves

$$e_1^*(r_1, d; \theta) \in \operatorname*{argmax}_e \, u_A(e_1, r_0, r_1, d; \theta) - k_A(e_1) - c_A(r_0),$$

and similarly, the government solves

$$r_1^*(e_1, d; \theta) \in \operatorname*{argmax}_r \, u_G(e_1, r_1, d; \theta) - k_G(r_1).$$

Proceeding backwards, the bystander chooses to demonstrate iff $c_B \leq c_B^*$, where c_B^* solves

$$\int u_B(1; e_1^*(1; \theta), r_0 + r_1^*(1; \theta)) \, d\pi(\theta \mid Q, r_0) - c_B \geq \\ \int u_B(0; e_1^*(0; \theta), r_0 + r_1^*(0; \theta)) \, d\pi(\theta \mid Q, r_0), .$$

A subgame perfect equilibrium in our model comprises a threshold strategy for the bystander, characterized by c_B^*, and the pair (e_1^*, r_1^*) which gives optimal effort and repression levels for the activist and government, respectively, that are a Nash equilibrium of the final challenge stage.

Illustrative Example

While our model allows us to make general statements about the link between initial repression and subsequent mobilized dissent, it abstracts from many of the factors that scholars have identified as critical determinants of backlash. To build intuition and to capture some of the key moderators identified in the literature, we build an illustrative example that will be our primary tool for analyzing backlash.

First, we for the purposes of this example, we treat the first stage as fully exogenous. This means we set $v = 1$ and will consider difference values of $r_0 > 0$ to illustrate the effects of initial repression. This allows us to capture the strategic scenario that is our primary focus—when the activist's protest is met with some amount of repression—while limiting the analysis to just the demonstration and challenge stages. The government has either high or low capacity, $\theta \in \{\underline{\theta}, \overline{\theta}\}$ where $\overline{\theta}$ means the government has high repression capacity relative to activist capacity. From the bystander's perspective, the prior probability that the government is type $\overline{\theta}$ is $p = Pr(\overline{\theta})$.

Now, the game begins with the bystander's demonstration choice. The bystander still observes a noisy signal of the government's repression capacity $Q = \theta + r_0 + \varepsilon$ where we draw ε from a standard normal distribution, $\varepsilon \sim N(0, 1)$. We parameterize the bystander's payoff

$$u_B(d; r_0, \theta) = \overbrace{(\lambda r_1 + e_1(1 + \beta d))^{1-\ell}}^{\substack{\text{value from}\\ \text{challenge outcome}}} - \overbrace{\alpha^d r_0}^{\substack{\text{cost from}\\ \text{initial repression}}} - \overbrace{dc_b}^{\substack{\text{cost of}\\ \text{demonstrating}}},$$

where $\ell \in [0, 1]$, which captures the bystander's relative risk aversion, her value from the challenge stage is increasing and concave in both government repression and activist effort, and λ represents the salience of expected repression. For example if λ is high, then the bystander cares quite a bit about how much she thinks the government is going to repress. Finally, β represents a boost participation by the bystander give to the activist. The bystander then internalizes some additional gains of mobilizing together with the activist in the final contest with the government.

The bystander faces two kinds of costs. She again pays the cost of participation if and only if she chooses to demonstrate. Additionally, she pays a cost of initial repression, scaled by $\alpha \in [0, 1]$, which captures the psychological gains from demonstrating. Notice, if the bystander stays home, $\alpha = 1$, so she bears the full cost of initial repression.

Proceeding to the challenge stage, we similarly parameterize the government and activist's payoffs. The government chooses a level of repression that maximizes

$$\max_{r_1} \ -(r_1 - \theta)^2 - r_1(\gamma d + k_g).$$

The government's preferred repression level matches its capacity. Choosing repression that exceeds capacity is costly and may, for example, risk defections by security forces. On the other hand, the government faces a threat from the activist and thus does not want to under-utilize its capacity. Two costs constrain repression, first k_G, which captures the direct cost of employing repression, and second γ, which is an additional cost of repressing the bystander if she has joined the activist in the challenge.

The activist wants aims to match the government's level of repression. It is too costly to mobilize enough to significantly exceed the government's level of repression. The activist maximizes

$$\max_{e_1} \ -(e_1 - (\eta_0 r_0 + \eta_1 r_1 + \beta d))^2 - k_A e_1 - c_A(r_0).$$

Repression has a negative effect on the activist, and may constrain their ability to match the government's effort. The parameters η_0 and η_1, which can differ across stages, reflect how demobilizing repression is for the activist, or how much repression affects the activist's capacity for dissent. The activist also receives a "boost" that comes from the bystander if she has chosen to demonstrate, β. Finally, the activist pays costs from effort, k_A, and initial repression, c_A.

With this illustrative example, we can characterize explicit solutions for optimal repression and activist effort, as well as the bystander's cost cutoff that determines her participation choice. It also allows us to consider how changes in parameters, each of which captures key factors identified in the literature, affect each actors' choice in the challenge stage.

Mobilization in the Challenge Stage

Backlash mobilization, in our model, is captured by a change in the bystander's willingness to demonstrate, c_B^*. caused by an increase in the initial level of repression, r_0. Before we can show how backlash arises, we first characterize the equilibrium repression and effort choices of the government and activist in the final challenge stage.

The government and activist's optimal choice depend on whether the bystander chooses to demonstrate. Given bystander support, $d \in \{0, 1\}$, a Nash equilibrium in the challenge stage is a pair $(e_1^*(d;\theta), r_1^*(d;\theta))$ that solves

$$\begin{aligned} e_1 &= e_1^*(r_0, r_1^*(e_1, d;\theta), d;\theta) \\ r_1 &= r_1^*(e_1^*(r_0, r_1, d;\theta), d;\theta). \end{aligned}$$

These equilibrium quantities have explicit solutions in the context of our illustrative example,

$$\begin{aligned} r_1^* &= \theta - \frac{1}{2}(\gamma d + k_G) \\ e_1^* &= \eta_0 r_0 + \eta_1 r_1 + \beta d - \frac{k_A}{2} \end{aligned}$$

Optimal repression is increasing in government capacity and decreasing in costs. Analogously, activist effort is increasing in repression, activist capacity, and the magnitude of benefits from coordinating with the bystander. These intuitive results suggest that our illustrative model effectively captures expected behavior by the government and activist. The bystander's expectations of this behavior underpin her decision to participate in anti-government activity and allow us to characterize backlash.

Necessary Conditions for Backlash

Backlash, in our model, occurs when an increase in initial government repression makes the bystander more willing to participate in subsequent mobilized dissent. Returning to the bystander's demonstration choice, she will choose to demonstrate, $d = 1$, when the gains from participation (relative to staying home) exceed the costs. Her relative gains depend on both repression and activist effort in the final challenge stage, which also depend on the bystander's choice. We can write these relative gains from participation as

$$\Delta(r_0; \theta) = u_B(1; e_1^*(1; \theta), r_0 + r_1^*(1; \theta)) - u_B(0; e_1^*(0; \theta), r_0 + r_1^*(0; \theta))$$

This allows us to characterize the bystander's equilibrium participation decision.

Lemma 1 *There is a unique c_B^* such that $d = 1$ if and only if $c_B \leq c_B^*$, where c_B^* solves*

$$\int \Delta(r_0; \theta) d\pi(\theta \mid Q, r_0) = c_B^*(Q).$$

To identify when each of the three prominent mechanisms—anger, logistical efficiency, and information—cause backlash mobilization, it is not sufficient to formalize a single path by which repression leads to mobilized dissent. Instead, we identify *necessary conditions* for each of these three mechanisms to generate backlash, or an increase in mobilized dissent following repression. This allows us to show what *must* be true in a model of repression and dissent for backlash mobilization to be observed as an equilibrium outcome.

To identify these necessary conditions, we consider each of the three mechanisms in turn, specify what backlash by each of these pathways looks like in the context of our model, and show what conditions must hold for backlash to materialize. This requires us to effectively shut down the other mechanisms temporarily, isolating the conditions for anger, logistical efficiency, and information to cause backlash independently. Once we have identified the conditions necessary for each individual mechanism to cause backlash mobilization, we combine them in the fully-specified model in the next section.

Backlash Motivated by Emotion

Observing repression has a direct, psychological impact on bystanders. As we highlighted above, past studies have shown that, as bystanders learn about state repression, this can trigger a number of emotional responses [**Examples here**]. While the broader category of psychological responses encompasses many reactions bystanders may have upon witnessing repression, we focus on *anger* as a shorthand for the negative emotions triggered by repression. Specifically, we say backlash arises by an *anger mechanism* when an increase in initial repression causes a negative emotional reaction that increases the bystander's willingness to demonstrate.

In the context of our model, backlash via anger is captured by a direct effect of initial repression, r_0 on the threshold c_B^*

Proposition 1 *Backlash occurs by an anger mechanism if* $\frac{dc_B^*}{dr_0} > 0$.

This result is immediate in the context of our illustrative example. To isolate the anger mechanism, we shut down any indirect effect from initial repression that comes from the final challenge stage. Then, $\frac{dc_B^*}{dr_0} = 1 - \alpha$. Given $\alpha \leq 1$, participating in anti-government activity reduces the direct costs from initial repression and $\frac{dc_B^*}{dr_0} > 0$. We can thus identify the first necessary condition for backlash.

Remark 1 *Backlash motivated by* ***anger*** *requires the psychological costs of repression are mitigated by demonstrating.*

Formally, $u_B(d; e_1^*(d;\theta), r_0, r_1^*(d;\theta))$ *cannot be quasilinear in* r_0.

Participating in anti-government activity can reduce the psychological costs of initial repression for the bystander. Remark 1 highlights that, to observe backlash arising from an anger mechanism, it must be the case that demonstrating is a salve for the anger triggered by observing repression. Should the bystander view participation in anti-government activity as an effective means to channel her anger about repression, backlash follows.

Otherwise, the costs of initial repression are sunk. To see this, rewrite the bystander's utility as

$$u_B(d; r_0, \theta) = (\lambda r_1 + e_1(1 + \beta d))^{1-\ell} \; - \alpha r_0 - dc_b.$$

Then $\frac{dc_B^*}{dr_0} = 0$ and, while repression still upsets the bystander ($\frac{du_B}{dr_0} = -\alpha < 0$), there is no direct effect of initial repression on her demonstration choice. Backlash from an emotional response then requires both anger and *agency.*

Backlash via Logistical Efficiency Gains

Repression imposes costs on the activist. Repression in the initial protest stage may capture both direct consequences for the activist, like costs arising from confrontation with security forces, or indirect consequences, like disruption of communications or freezing assets. Alternatively, as highlighted in [**cites**], repression may have a mobilizing effects on opposition groups, increasing the activists' incentives for, or capabilities to, exert effort in the final challenge. How does a change in the activit's logistical costs of challenging the government after repression affect the bystander?

Backlash mobilization via a logistical efficiency mechanism requires that a change in the activist's expected effort in the final challenge increases the bystander's willingness to participate in anti-government activity; formally, $\frac{dc_B^*}{de_1^*} > 0$. This relies on an indirect effect of initial repression on the bystander, via the activist. Therefore, to identify when backlash may arise via this mobilization mechanism, we first characterize the direct effect of inital repression on activist effort necessary for backlash.

Proposition 2 *Backlash occurs via a logistical efficiency mechanism if activist effort is increasing in both initial repression,* $\frac{\partial e_1^*}{\partial r_0} > 0$*, and bystander participation,* $\frac{\partial e_1^*}{\partial d} > 0$*.*

The activist's optimal effort in the final challenge depends directly on the bystander; $e_1^*(d; \theta)$ is a function of the bystander's demonstration choice. If bystander participation had no or negative effect on activist effort, the bystander's utility from demonstrating may fall below her utility from staying home, precluding backlash. Further, backlash mobilization requires a direct link to the level of initial repression. If initial repression depressed activist effort, this also may deter the bystander from demonstrating. Thus, backlash requires both bystander participation and initial repression mobilize the activist. This allows us to state the total effect of initial repression via a logistical costs mechanism.

Remark 2 *Backlash via* ***logistical efficiency gains*** *follows from a direct effect on activist effort, which makes bystander more willing to demonstrate, which reinforces activist effort.*

Formally, activist effort and bystander participation must be strategic complements.

Absent complementarities between activist effort and bystander participation, initial repression only has the direct anger effect on the bystander. This not only underscores the necessary conditions for backlash via a logistical cost mechanism but also highlights a broader implication of model. Observed backlash may depend on more than one mechanism. When initial repression has both direct and indirect effects on the bystander's willingness to demonstrate, it may be difficult to disentangle which mechanism is driving the response to repression.

Remark 3 *The total effect of initial repression on the bystander's willingness to demonstrate depends on both the* ***anger*** *and* ***logistical efficiency*** *mechanisms. This means observing backlash even absent an informational effect risks conflating multiple mechanisms.*

Though this result suggests some observational equivalence between the anger and logistical costs mechanisms, our model highlights one observable factor that may distinguish between mechanism—activist effort. Observing how repression directly affects the activist can indicate whether backlash is driven by anger or efficiency.

Backlash due to Information

To isolate both the anger and logistical efficiency mechanisms, we presumed complete information. Now, we reintroduce uncertainty over the government's relative capacity for repression to demonstrate how observing initial repression can provide information about the government and potentially generate backlash mobilization. This information channel is the most commonly articulated in the literature, though whether and how learning about the government may cause backlash is still debated [**examples here**].

Backlash via an information mechanism requires the bystander's signal, Q, increases her willingness to demonstrate. Before we can articulate necessary conditions for backlash, we first must identify how an increase in initial repression affects the bystander's posterior belief about the government's capacity.

Lemma 2 *An increase in initial repression signals weakness, reducing the bystander's posterior belief that the government is high capacity.*

$$\frac{\partial}{\partial r_0}\pi(\bar{\theta}|Q, r_0) < 0$$

A higher signal is "good news" about the government's type in the sense that a higher signal Q is more likely for higher θ, or when the government is higher capacity. Therefore, the bystander's posterior belief is increasing in $Q - r_0$ and decreasing in r_0.

To identify backlash via an information mechanism, we must also identify how changes in the bystander's posterior belief affect her willingness to demonstrate. What remains is to show how an increase in Q affects the bystander. We can break down this effect into two components. First, we define a measure the magnitude of "good news," or much an increase in Q increase the odds of a high capacity type. We can write this as,

$$\frac{\pi(\bar{\theta} \mid Q', r_0)}{\pi(\underline{\theta} \mid Q', r_0)} - \frac{\pi(\bar{\theta} \mid Q, r_0)}{\pi(\underline{\theta} \mid Q, r_0)}. \tag{1}$$

Second, we identify the effect of the signal Q on the bystander's participation threshold.

Conjecture 1 *For* $\frac{dc_B^*}{dQ} > 0$*, the difference in Equation* (1) *must decrease in* r_0*.*

Conjecture 1 implies that the bystander's posterior belief must have decreasing differences in Q and r_0. In other words, how much more weight the bystander puts on lower capacity types at $Q' > Q$ decreases for higher levels of initial repression.

Distinguishing Backlash Across Mechanisms

Backlash moblization can arise via each of our anger, logistical efficiency, or information mechanisms, but in each case backlash is characterized by an increase in the bystander's willingness to demonstrate upon observing a higher level of initial repression. In the case of the logistical efficiency mechanism, backlash mobilization arises via an indirect effect, where the direct effect of initial repression is instead on the activist. For the information mechanism, backlash also arises via indirect effect, where the direct effect is both on the activist's and the government's challenge stage actions, conditional on their relative capacity. Observed backlash, then, may depend on the combination of these direct and indirect effects. We now relax the assumption that each mechanism operates indecently to identify how the interaction of multiple mechanisms impacts observed backlash mobilization. However, we again present a complete information benchmark before reintroducing the information mechanism.

To see the combined effect of the anger and logistical efficiency mechanisms, we compute the total derivative of the bystander's tradeoff under complete information.

Proposition 3 *When anger and efficiency backlash mobilization conditions are satisfied, observed backlash is increasing in initial repression.*

Proof:

$$\frac{dc_B^*}{dr_0} = (1-\ell)\left(\left(\lambda r_1 + e_1(1+\beta)\right)^{-\ell} - \left(\lambda r_1 + e_1\right)^{-\ell}\right)\frac{\partial e_1^*}{\partial r_0} + (1-\alpha) > 0$$

■

This highlights a complementary relationship between the logistical costs and anger mechanisms. When the bystander is angered by repression, she is more likely to demonstrate, which increases the activist's optimal effort and reinforces the bystander's willingness for even higher costs of participation.

It is also possible for the logistical efficiency mechanism to contaminate the observed effect of the anger mechanism. To see this, we consider a comparative static on the activist's optimal level of repression. Across contexts, activists face different conditions that may facilitate or hinder the ability to mobilize for anti-government activity. The *activist ecosystem* can significantly impact the likelihood of observing backlash. In a context where the activist has relatively low capacity to match the government's repression, the payoff for the bystander from demonstrating may no longer exceed the costs.

$$\frac{dc_B^*}{d\eta} = \frac{dc_B^*}{de_1}\frac{de_1}{d\eta} = (1-\ell)\left(\left(\lambda r_1 + e_1(1+\beta)\right)^{-\ell} - \left(\lambda r_1 + e_1\right)^{-\ell}\right)\frac{\partial e_1^*}{\partial \eta} > 0,$$

thus when η decreases we are less likely to observe backlash. While the bystander still may find anger motivating, a bad ecosystem may discourage participation, dominating the anger effect. This also has a direct observable implication—backlash in a bad activist ecosystem is more likely to be driven by anger.

Conclusion

Backlash to repression represents both an empirical patterns and a mechanism that links an increase in repression to an increase in mobilized dissent. We identify necessary conditions for backlash mobilization to arise via the three most commonly articulated pathways specified in the extant literature, emotional responses, logistical efficiency gains, and information. Isolating these necessary conditions allows us to show what must be true to see backlash empirically, whether backlash mobilization follows from a single mechanism or a combination of multiple mechanisms. Further, our model highlights the importance of a well-specified theory in explaining backlash. Because these mechanisms can have countervailing effects, observing backlash mobilization implies constraints on each mechanism that are obscured without a theoretical model that clearly articulates these relationships.

References

Aytaç, S. E., L. Schiumerini, and S. Stokes. 2018. “Why Do People Join Backlash Protests? Lessons from Turkey.” Publisher: SAGE Publications Inc, *Journal of Conflict Resolution* 62, no. 6 (July): 1205–1228.

Chenoweth, E., and M. J. Stephan. 2011. *Why civil resistance works: The strategic logic of nonviolent conflict.* Tex.date-added: 2017-03-09 17:26:15 +0000 tex.date-modified: 2017-03-09 17:26:15 +0000. New York, NY: Columbia University Press.

Chiang, A. Y. 2021. “Violence, non-violence and the conditional effect of repression on subsequent dissident mobilization.” *Conflict Management and Peace Science* 38, no. 6 (November): 627–653.

Davenport, C. 2007. “State repression and political order.” Tex.date-added: 2011-06-26 11:08:28 -0500 tex.date-modified: 2011-06-26 11:08:28 -0500, *Annual Review of Political Science* 10:1–23.

Eck, K., S. Hatz, C. Crabtree, and A. Tago. 2021. “Evade and Deceive? Citizen Responses to Surveillance.” Publisher: The University of Chicago Press, *The Journal of Politics* 83, no. 4 (October): 1545–1558.

Ellefsen, R. 2021. “THE UNINTENDED CONSEQUENCES OF ESCALATED REPRESSION*.” *Mobilization: An International Quarterly* 26, no. 1 (March): 87–108.

Esberg, J., and A. A. Siegel. 2023. “How Exile Shapes Online Opposition: Evidence from Venezuela.” *American Political Science Review* 117, no. 4 (November): 1361–1378.

Freedman, M., and E. F. Klor. 2023. “When Deterrence Backfires: House Demolitions, Palestinian Radicalization, and Israeli Fatalities.” Publisher: SAGE Publications Inc, *Journal of Conflict Resolution* 67, nos. 7-8 (August): 1592–1617.

Graham, M. H., and M. W. Svolik. 2020. “Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States.” *American Political Science Review* 114, no. 2 (May): 392–409.

Hager, A., and K. Krakowski. 2022. “Does State Repression Spark Protests? Evidence from Secret Police Surveillance in Communist Poland.” *American Political Science Review* 116, no. 2 (May): 564–579.

Hatz, S. 2019. “Israeli Demolition Orders and Palestinian Preferences for Dissent.” *The Journal of Politics* 81, no. 3 (July): 1069–1074.

Hess, D., and B. Martin. 2006. “Repression, backfire, and the theory of transformative events.” Tex.date-added: 2017-03-13 23:56:08 +0000 tex.date-modified: 2017-03-13 23:57:27 +0000, *Mobilization* 11 (2): 249–267.

Koopmans, R. 1997. “Dynamics of Repression and Mobilization: The German Extreme Right in the 1990s.” *Mobilization* 2 (2): 149–164.

Lichbach, M. I. 1987. “Deterrence or escalation? The puzzle of aggregate studies of repression and dissent.” Tex.date-added: 2011-06-26 11:08:28 -0500 tex.date-modified: 2011-06-26 11:08:29 -0500, *Journal of Conflict Resolution* 31:266–297.

Loewenthal, A., S. H. Miaari, and A. Abrahams. 2023. “How civilian attitudes respond to the state’s violence: Lessons from the Israel–Gaza conflict.” Publisher: SAGE Publications Ltd, *Conflict Management and Peace Science* 40, no. 4 (July): 441–463.

Moore, W. H. 1998. “Repression and dissent: Substitution, context, and timing.” Tex.date-added: 2011-06-26 11:08:28 -0500 tex.date-modified: 2011-06-26 11:08:29 -0500, *American Journal of Political Science* 42 (3): 851–873.

Petrova, M. G. 2022. “Is It All the Same? Repression of the Media and Civil Society Organizations as Determinants of Anti-Government Opposition.” *International Interactions* (May): 1–29.

Reny, T. T., and B. J. Newman. 2021. “The Opinion-Mobilizing Effect of Social Protest against Police Violence: Evidence from the 2020 George Floyd Protests.” *American Political Science Review* 115, no. 4 (November): 1499–1507.

Ritter, E. H., and C. R. Conrad. 2016. “Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression.” *American Political Science Review* 110, no. 1 (February): 85–99.

Schon, J., and D. Leblang. 2021. “Why Physical Barriers Backfire: How Immigration Enforcement Deters Return and Increases Asylum Applications.” Publisher: SAGE Publications Inc, *Comparative Political Studies* 54, no. 14 (December): 2611–2652.

Steinert, C. V., and C. Dworschak. 2023. “Political Imprisonment and Protest Mobilization: Evidence From the GDR.” Publisher: SAGE Publications Inc, *Journal of Conflict Resolution* 67, nos. 7-8 (August): 1564–1591.

Sullivan, C. M. 2016. “Political Repression and the Destruction of Dissident Organizations: Evidence from the Archives of the Guatemalan National Police.” *World Politics* 68, no. 4 (October): 645–676.

Tertytchnaya, K. 2023. “”This Rally is Not Authorized”: Preventive Repression and Public Opinion in Electoral Autocracies.” Publisher: Johns Hopkins University Press, *World Politics* 75 (3): 482–522.

Thachil, T. 2020. “Does Police Repression Spur Everyday Cooperation? Evidence from Urban India.” Publisher: The University of Chicago Press, *The Journal of Politics* 82, no. 4 (October): 1474–1489.

Wood, R., G. Y. Reinhardt, B. RezaeeDaryakenari, and L. C. Windsor. 2022. “Resisting Lockdown: The Influence of COVID-19 Restrictions on Social Unrest.” *International Studies Quarterly* 66, no. 2 (June): sqac015.