Backlash*

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

March 26, 2025

This is a preliminary draft. Please do not circulate without author permission.

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 2003-2023 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 proposed mechanisms–anger, organizational capacity, 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.

*We thank Jessica Maves Braithwaite, Pearce Edwards, Addison Ewig, Nguyen Ha, Chloe Hale, Kai Keltner, Chaelin Kwon, Zhaotian Luo, Pablo Montagnes, Arturas Rozenas, Keith Schnakenberg, Tara Slough, William Spaniel, Chris Sullivan, Stephane Wolton, Joseph Young, Antoine Zerbini, and participants at the Political Economy of Conflict workshop at Vanderbilt University, the Comparative Politics Annual Conference at Washington University in St. Louis, the Comparative Formal Theory conference at LSE, the 2024 Peace Science Society (International) meeting, and the political science faculty and students at Louisiana State University and Emory University for their helpful comments.

†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

Governments repress to quell mobilization and dissent actions, but repression can instead catalyze widened participation or intensified efforts to 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. Even when we restrict attention to backlash *to repression,* scholars persistently use the term to mean a variety of behaviors. This conceptual breadth leaves ambiguous which actions taken in response to repression count as backlash.

The conceptual ambiguity in the definition of backlash is mirrored by the mixed empirical evidence for increases in dissent following repression. While repression and changes in subsequent patterns of dissent are both widely observed, explanations for exactly how and why these phenomena are linked vary considerably across studies. Backlash is a theoretical concept—it is the meaning we give to a positive correlation between government repression and subsequent mobilized dissent. The concept of backlash, however, has emerged and evolved from an empirical literature that studies this correlation. This has perhaps led to the proliferation of explanations and evidence that make it hard to know backlash when we see it in data.

To understand backlash, we answer two fundamental questions. What is backlash to repression? And perhaps more importantly, when, how, and why does repression cause backlash mobilization? In this article, we define backlash using a general model of dissidence. We then identify two key challenges to studying backlash. First, backlash captures multiple different pathways by which repression can increase dissent. We describe four mechanisms

where backlash can follow as an observable implication of each of these channels. Theoretically, this introduces a problem of *observability*. To identify backlash our theories must allow us to ...TK Second, since backlash follows from multiple mechanisms, studying backlash requires a research design that is sufficiently unambiguous to *attribute* backlash to a specific mechanism.

Our framework is deliberately built to reflect findings from the literature on backlash to repression. To do so, we systematically inventory political science articles studying the correlation between repression and backlash mobilization published over the past two decades in the three general political science journals and other well-cited publications in comparative and international relations subfield journals. This allows us to highlight the varied and conflicting ways in which scholars conceptualize backlash mobilization. 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, (3) how backlash is operationalized as a dependent variable, and (4) what methodologies are used to study it.

This scholarship review illustrates the variation in the ways scholars define backlash, measure it, and explain it as related to repression. A large portion of scholarship refers to backlash as an *empirical phenomenon*, where observable government repression is correlated with an increase in negative political consequences for the government. The backlash in 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 repressed group and other times 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 a dissent action to do so. We categorize the mechanisms of backlash into three types: emotion, strategic feedback, and learning. These 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.

Furthermore, 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 are rarely supported by realistic logic. Moreover, the different explanations can yield different results without clarity as to when or if one should dominate.

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 a 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 of what it means for them. We describe the necessary assumptions for this to occur under each possible 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 each other. 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, not only does this make 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...

Trends in Backlash Scholarship, 2004–2023

General Patterns

To obtain a picture of the body of scholarly knowledge on backlash to repression, we conducted a coded literature review of articles published on the topic from 2004 to 2023 that were published in the top-ranked general-interest political science journals (the *American Political Science Review,* the *American Journal of Political Science,* and the *Journal of Politics*).[1] as well as articles published in political science subfield journals that are frequently cited according to Google Scholar citation trends as relevant to the topic.[2] We classified ar-

1. The inventory includes articles 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.

2. We include a few articles from economics and sociology journals when they are commonly cited by political scientists on the topic. Although the inclusion rule creates a systematic sample from the top three political science journals, the subfield journal inclusion rule is not systematically inclusive. There are

ticles as relevant when the words *repression* and *backlash* or *backfire* 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 91 articles, where 26 are from subfield-specific journals.[3] Figure 1 presents a histogram of where the articles were published over time, suggesting a significant increase in the scholarly use of the term backlash or studies of repression and dissent in the last decade in both subfield and general political science journals.

[Figure 1 here.]

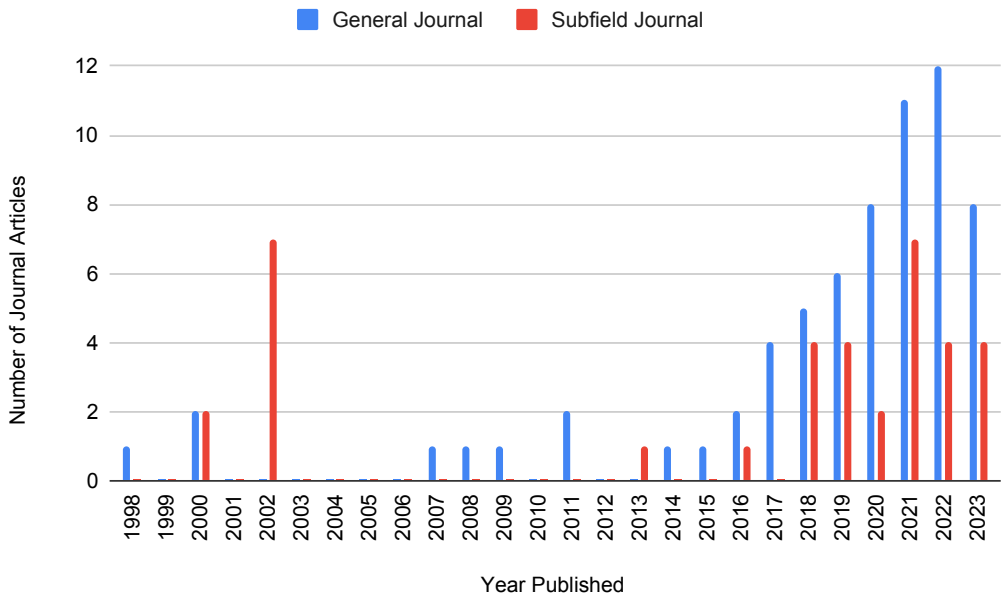

Figure 1: Number of Articles Over Time

We coded these articles as to how they characterize backlash as a mechanism and as an empirical outcome. For all articles in the set, the dependent variable is an activated response

important studies that we have certainly missed in our effort to track scholarly trends on the topic.

3. The full inventory of the coded articles are reported in the Supplemental Materials.

to a government that has repressed some dissidents or potential dissidents, but the studies differ on what the response is. We classified 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 a dissent action (nonviolent or violent) or a change in 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 in a stacked histogram.

[Figure 2 here.]

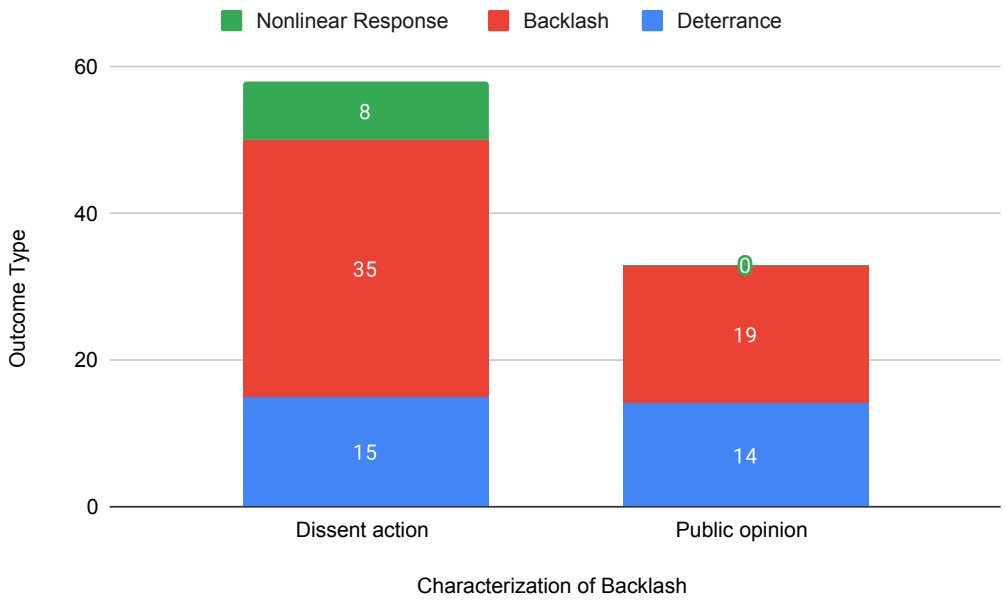

Figure 2: Backlash Characterization and Empirical Outcome

Figure 2 also presents the directional findings as a proportion of the number of articles using each type of dependent variable. 58 articles study the concept of backlash as dissent actions in response to repression, such as an increase in protest events, a surge in mobilized participants, or an escalation in violence. 60.3% 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; 25.9% identify conditions where repression instead deters mobilized backlash.[4] A smaller but meaningful number (33) of published articles study how repression affects public attitudes or popular expressed support for the government; 42.4% of these find that repression maintains or increases public support for the regime, and 57.6% find a negative effect on or decrease in support for the regime, which the authors refer to as a form or driver of backlash.

Despite the variety of approaches and findings presented in scholarly research on the topic of backlash, there are core elements that scholars consistently include in their theories and concepts of backlash, which we rely on to build a general model of dissidence.

Actors and Explanations

Actors are the political decision makers that scholars discuss as relevant for the response to repression. The list includes the government and its authorized agents (referred to collectively as the government), a dissident or mobilized dissenting group (activist), and bystanders deciding how to respond. We characterize the role of each actor in the backlash interaction by their contribution to it. The government is the subject or initiator of repression. 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 presents patterns in how actors are described in the research inventory.

[Figure 3 here.]

In all of the articles surveyed, there is a government 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

4. We also include a middle category (*partial deter/cause*) where the study finds that repression sometimes increases and other times decreases dissent activities.

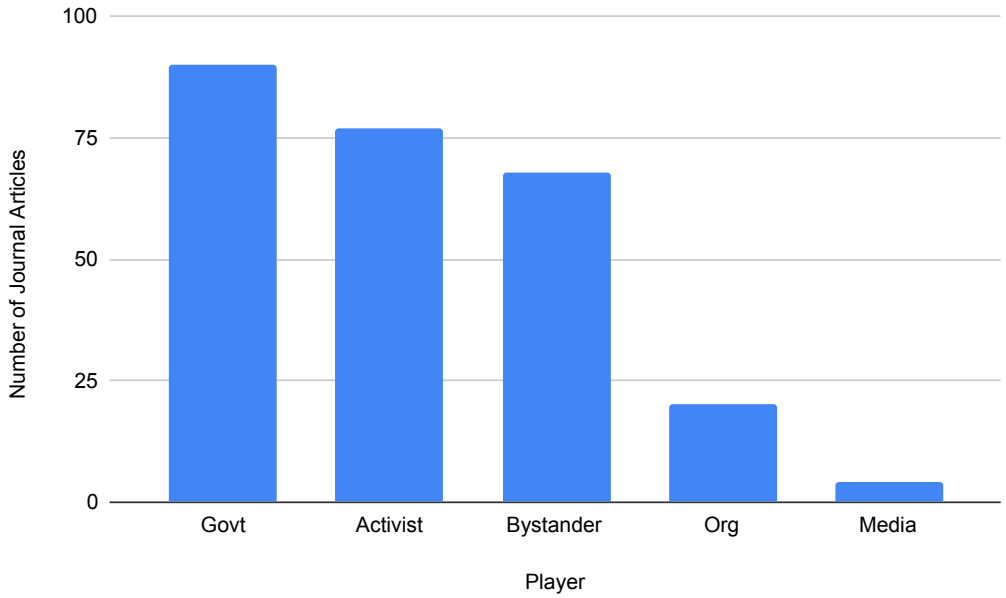

Figure 3: Types of Actors in Backlash Articles

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 security agent who carries it out, 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 as a whole to be responsible.

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; Komisarchik, Sen, and Velez 2022) 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.[5]

Critically, studies of backlash include bystanders who learn that government authorities repress a dissenting group and decide whether to take action. In most, the authors focus on "an uninvolved witness to contentious politics (Strauss 2018)" 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 who is reflecting on the behavior of the authorities when determining their vote choice (Graham and Svolik 2020). In these cases, the bystanders 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 it not occurred. In a smaller subset of studies, the subject of backlash is the dissenting group that was the target of government repression (Ritter 2014; Ritter and Conrad 2016; Esberg and Siegel 2023); the experience of being repressed causes a change that leads the organization to increase their efforts to mobilize and carry out another event of dissent of greater magnitude. In sum, most studies assert that the sequence of backlash mobilization is (1) an activist group dissents (or threatens to), (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.

5. In two of the reviewed articles, backlash occurs when the government has repressed the media.

We coded each article in the inventory according to explicit and implicit *explanations* about how repression leads to increased mobilization, negative opinion, or dissent actions. Most studies 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 4): Emotion, organizational capacity, and information.

[Figure 3 here.]

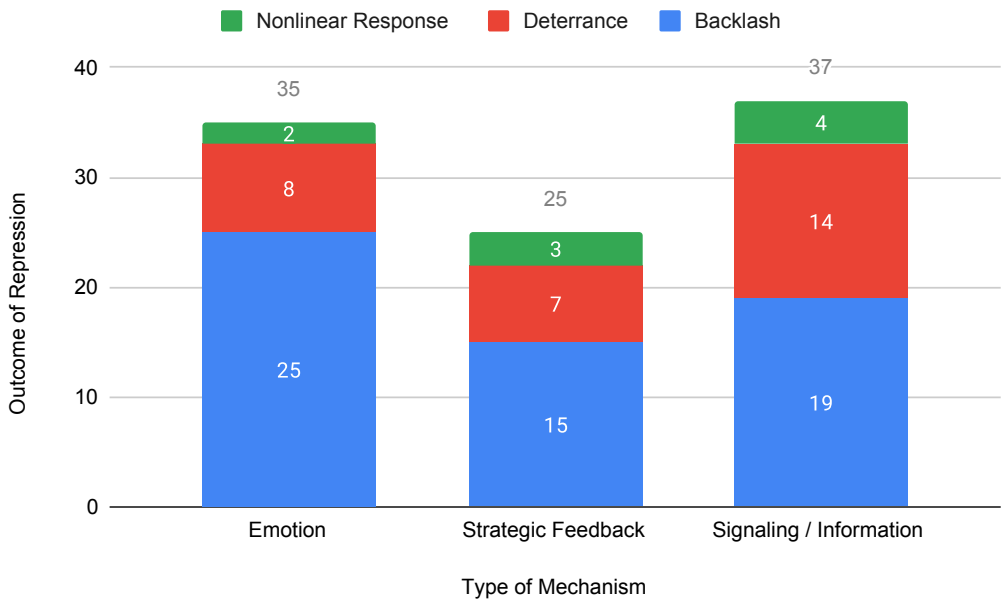

Figure 4: Backlash Mechanism Types and Predicted Outcomes

Emotion: When a government represses dissidents, either the dissidents or bystanders experience an emotion that is powerful enough to motivate their behavior. They may become angry and join the dissent effort. Repression can inspire outrage and support for dissident claims, leading 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).[6] This can be especially common among people whose families or identity groups experienced political violence in the past, because the legacy of the historical violence leaves descendants primed and resourced to react to new violations (Rozenas, Schutte, and Zhukov 2017; Haffert 2022; Thaler, Mueller, and Mosinger 2023). Other scholars conceptualize emotion through honor and morals, explaining how threats to ones' morals and desire for self-respect drives backlash (Pearlman 2018; Dafoe, Hatz, and Zhang 2021).

Strategic Feedback: 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 their efforts against the government. 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 dissidents to increase their use of violence in turn (Lichbach 1987; Moore 1998; Chiang

6. We include in this category arguments that grievances evoke backlash or revolt, as in most cases it is the psychological experience of the grievance, loss, or relative deprivation that motivates the group to act (Collier and Hoeffler 2004; Shadmehr 2014; al-Anani 2019; Gibilisco 2021).

2021).

Learning: The remaining explanation category is defined by a learning process: Activists or bystanders learn something when they observe the repression of mobilized dissent, which alters what they choose as the best response. This is the most common type of mechanism in our inventory of articles. When the general population becomes aware that the government has violated their rights, they are more likely to punish government actors with dissent (Simmons 2009) or at the polls (Conrad and Moore 2010; Pop-Eleches and Way 2021). Overt repression makes it known that dissidents are under surveillance, leading activists to change their tactics (Davenport 2015; Sullivan and Davenport 2018; Gohdes 2020; Eck et al. 2021). Observers may also learn that the government is willing to take illegal or illegitimate actions to control the population (Rozenas and Stukal 2019; Curtice and Behlendorf 2021). In contrast, the government can portray activists in a negative light, so that the bystander will support repressive tactics and not join the dissent effort (Edwards and Arnon 2021; Pop-Eleches and Way 2021).

Activists and bystanders may learn that the government is more resolved to repress than they previously understood, so they expect repression in the next period and decline to participate in expectation (Ritter and Conrad 2016). Government resolve is a concept that captures the state's capacity to repress (E. B. Carter and B. L. Carter 2022), or what costs authorities accept to carry it out (Carroll and Pond 2021), and scholars sometimes label these same ideas as government strength (Thachil 2020). In short, observers learn that the government is more willing or able to repress dissidents than they expected, leading them to anticipate increased repression if they join or increase dissent efforts.

Bystanders also learn about the activist group through observing the dissent–repression interaction. For those who are not sure whether activists are better leaders for the bystander than the incumbent government, patterns of dissent and repression can clarify the valence of the group and its leaders (Shadmehr and Boleslavsky 2022). Many bystanders have a

tendency to free-ride, but observing a collective action and the government response informs them of the likelihood that others will join them for a successful challenge (Shadmehr and Bernhardt 2011). Repression reveals the joint investment equilibrium (De Jaegher and Hoyer 2019), so the bystander expects to be contributive and pivotal, leading them to join the activist (Kuran 1991; Lohmann 1993, 1994).

As illustrated in Figure 4, the three types of mechanism are scattered across the studies of backlash; scholars argue very different things about how backlash occurs. Also importantly, the mechanisms connecting repression to backlash are very often assumed, rather than examined with careful logic or tested directly. Do people get angry or afraid when they learn that the government represses protesters? We know that their opinion of the government decreases, but we do not measure their emotions. Is it actually easier or more difficult to recruit participants after repression? We only know if dissent or mobilization increases or decreases, not why. Scholars posit that bystanders could be learning, but we do not know what they are learning. The literature treats backlash as an empirical phenomenon that occurs in a variety of correlated contexts.

Rather than adding another model of backlash in context to the field, we use the consistent unifying ideas found in this scholarship to build a theoretical model of dissidence, where activists dissent, government represses, and a dissident bystander make a set of choices in equilibrium that can follow the empirical pattern of backlash mobilization. The three-actor model sets up clear counterfactual behaviors to identify when observable repression causes either the activists to increase their efforts or the bystander to join the activists. We specify each mechanism type in its simplest forms: repression causes (a) a change in the bystander's emotional utility, (b) an increase in the activist's efficiency of effort, and/or (c) a change in the bystander's posterior beliefs. This setup allows us to examine how these mechanisms logically differ from each other, when they actually can cause backlash behaviors, and what each implies about the observable world.

Bridging Theory and Empirics

Backlash is an empirical phenomenon—but it is not purely an empirical phenomenon. Scholars ultimately care not only whether backlash occurs, but also *why it happens*. Scholars generally want to attribute backlash to a theoretically-motivated mechanism. Although the occurrence of backlash is typically assessed empirically by investigating the correlation between initial repression and subsequent bystander participation, answering questions like "what causes backlash?" involves identifying the theoretical mechanism responsible and activating it, experimentally or as if experimentally (Woodward 2005). This concern gives rise to the problems of causal identification which we investigate in this article.

To measure and judge how much of the participation of bystanders is motivated by backlash, the objective is to empirically identify the relationship between initial repression and participation. To be a cause, must be the case that repression causes the bystander to act in a way that they would not have done in the absence of repression. They would have stayed home, but observing repression motivated them to change their behavior. This means that there must be something about joining (or deciding to opt out) that responds to the mechanism at work. Causal identification also requires an understanding of the process through which repression alters the counterfactual. The mechanism tells us what experimental treatments to apply to activate the process leading to backlash, as well as what observable patterns we should expect to see to falsify the theory.

The relationship between theory and experiment is not a simple one and it is important to understand how that relationship manifests in different contexts (Bueno De Mesquita and Tyson 2020). Connecting a theoretical effect with the estimand of a research design involves building a bridge. First, the theoretical mechanism must give rise to a phenomenon with enough reliability that one should expect that the treatment will reliably activate the mechanism to produce the phenomenon of interest on average. Second, the research design

should be sufficiently focused and powerful so that the researcher can measure an empirical difference from the counterfactual with reasonable accuracy. Together, these considerations give rise to two dimensions of the identification problem. *Observability* requires that the theoretical mechanism that causes the backlash be expressed in such a way that a sufficiently focused research design would produce evidence in favor of it when it is active. In other words, the theoretical mechanism must be sufficiently clear that the researcher can correctly identify the conditions under which the mechanism should be at work and replicate those conditions in their design. *Attribution* requires that the empirical correlation between initial repression and dissent participation be attributable to a specific mechanism.

Our goal is to draw meaning from a varied and informative corpus of scholarly studies to understand why and when repression will lead to the presence or absence of backlash; this is a question of empirical and theoretical identification. In what follows, we present a series of formal investigations, built from the unifying assumptions of the backlash scholarship, to consider and challenge the expectation that we can *observe* backlash and know it to be backlash and *attribute* backlash to a logical theoretical explanation.

A Formal Definition of Backlash

We present a general formal model that distills the key elements of the narrative common to the studies described above. Formally, our game features three actors: an activist (A), a government (G), and a bystander (B), and our analysis focuses primarily on the bystander.[7] In the first stage of the game, the bystander chooses whether to demonstrate on the side of the activist, $d = 1$, or stay home, $d = 0$. We normalize the bystander's payoff for abstaining to zero. If she demonstrates, she pays $c_B > 0$, which reflects the cost of participation.

In the second stage, the activist undertakes some anti-government activity, $e \in [0, \overline{e}]$,

7. We refer to the bystander as she, the activist as they, and the government as it.

and the government represses, $r \in [0, \overline{r}]$. Because we focus on backlash, our analysis begins after an exogenous initial level of repression, $\rho_0 \in \Omega$, which could target the activist, the bystander, or others, where Ω is a compact subset of $\mathbb{R}$.[8] The government's capacity for repression (relative to the activist) is represented by $\theta \in \Theta$, which is a compact subset of $\mathbb{R}$.

The initial level of repression, ρ_0, and the government's capacity, θ, are drawn from an absolutely continuous joint distribution function p, with continuously differentiable density. Before making any strategic decisions, the government and the activist both learn the state of the world θ, but the bystander does not. Therefore, the state of the world θ captures the potential uncertainty between the bystander and other actors about the relative capacity of the government to repress. The initial repression, ρ_0, potentially tells the viewer something about how likely the government is to repress in the next period and how much effort the activist will put in. Let the smooth density function $\pi_p(\theta \mid \rho_0)$ reflect the bystander's posterior belief about the state of the world, θ, conditional on the level of initial repression, ρ_0, and the prior, p. We suppose that the posterior, π, is *responsive*, meaning that its derivative has full rank.[9]

The government payoff is

$$u_G(e, r, d; \rho_0, \theta),$$

and the activist payoff is

$$u_A(e, r, d; \rho_0, \theta).$$

The final stage payoff for the bystander is given by

$$d \cdot \big(u_B(e, r, \rho_0; \theta) - c_B\big).$$

The functions are smooth and strictly concave. The timing of the dissidence game is:

8. It is straightforward to extend the model to endogenize the initial level of repression.
9. This feature is true if π is derived from Bayes's rule.

0. The state of the world, θ, and initial repression, ρ_0, are drawn and revealed to G and A , and only ρ_0 is revealed to B;

1. **Participation**: B elects whether to join (demonstrate in) dissidence;

2. **Contention**: A exerts effort and G represses, after which the game ends and payoffs are realized.

An equilibrium to our game is where each actor chooses a sequential best response given the information they have when they make their choice, and which is consistent with how they update information.

Lemma 1 *There exists an equilibrium, characterized by the triple* $(c_B^*(\rho_0), e^*(d, \theta; \rho_0), r^*(d, \theta; \rho_0))$, *where activist effort is*

$$e^*(d, \theta; \rho_0) \in \underset{e \in [0, \bar{e}]}{\operatorname{argmax}}\ u_A(e, r, d; \rho_0, \theta);$$

government repression is

$$r^*(d, \theta; \rho_0) \in \underset{r \in [0, \bar{r}]}{\operatorname{argmax}}\ u_G(e, r, d; \rho_0, \theta);$$

and where Bystander participates if and only if $c_B \leq c_B^*(\rho_0)$, *where*

$$c_B^*(\rho_0) = \int u_B(e^*(1, \theta; \rho_0), r^*(1, \theta; \rho_0), \rho_0; \theta) \pi_p(\theta \mid \rho_0)\ d\theta. \tag{1}$$

An equilibrium to the dissidence game corresponds to an activist effort level, government repression, and participation by the bystander, which all potentially depend on the initial level of repression (for a variety of reasons we will return to). The bystander will demonstrate when the gains from participation exceed the costs, reflected by the cost threshold $c_B^*(\rho_0)$, characterizing the bystander's equilibrium participation decision in terms of the set of costs, $c_B \leq c_B^*(\rho_0)$, that she is willing to incur to demonstrate.

Because we are interested in the connection between the theoretical outcome of backlash and the measurement and empirical identification of backlash as a phenomenon, we require a few additional features that we would not require in a purely theoretical analysis. The equilibrium characterized in Lemma 1 is not necessarily unique (without stronger assumptions), and as a consequence, changes in the initial level of repression could, in principle, facilitate a change between different equilibria. Such changes would appear as a kind of backlash, i.e., increased ρ_0, while reducing $c_B^*(\rho_0)$, could also facilitate a switch to an equilibrium where the lowered $c_B^*(\rho_0)$ is nevertheless larger than before. Though interesting, we focus on more substantively motivated forms of backlash, and so our analysis presumes a unique equilibrium.[10]

Backlash is ultimately about the relationship between the bystander's willingness to demonstrate, represented in our model by $c_B^*(\rho_0)$, and the initial level of repression, ρ_0. It is reflected in our model by the change in the bystander's equilibrium willingness to demonstrate, $c_B^*(\rho_0)$, caused by an increase in the initial level of repression, ρ_0. Specifically, increases in $c_B^*(\rho_0)$ imply that bystander is willing to pay larger demonstration costs, or equivalently, the set of costs for which the bystander demonstrates is larger. Our analysis of backlash thus corresponds to an analysis of the comparative static on the cutoff $c_B^*(\rho_0)$ with respect to the initial level of repression. This is because when $\frac{\partial}{\partial \rho_0} c_B^*(\rho_0) < 0$, or when $\frac{\partial}{\partial \rho_0} c_B^*(\rho_0) = 0$, then increased initial repression implies that bystander is willing to demonstrate for fewer costs, or where initial repression does not influence participation by the bystander (in total), respectively. When $\frac{\partial}{\partial \rho_0} c_B^*(\rho_0) > 0$, then increased initial repression leads to an increase in bystander's incentive to demonstrate.

Imagine an experiment that exogenously manipulates initial repression, ρ_0, and that this manipulation is known only to the analyst—to avoid issues of commensurability (**?**). Then,

10. This is stronger than necessary. All we require is that changes in ρ_0 do not lead to changes between different equilibria.

the associated treatment effect is reflected by the total derivative:

$$\frac{dc_B^*(\rho_0)}{d\rho_0} = \int \frac{\partial u_B(e^*(1,\theta;\rho_0), r^*(1,\theta;\rho_0), \rho_0; \theta)}{\partial \rho_0} \pi_p(\theta \mid \rho_0)\, d\theta \tag{2}$$

$$+ \int \frac{\partial u_B(e^*(1,\theta;\rho_0), r^*(1,\theta;\rho_0), \rho_0; \theta)}{\partial e} \cdot \frac{de^*(1,\theta;\rho_0)}{d\rho_0} \pi_p(\theta \mid \rho_0)\, d\theta \tag{3}$$

$$+ \int \frac{\partial u_B(e^*(1,\theta;\rho_0), r^*(1,\theta;\rho_0), \rho_0; \theta)}{\partial r} \cdot \frac{dr^*(1,\theta;\rho_0)}{d\rho_0} \pi_p(\theta \mid \rho_0)\, d\theta \tag{4}$$

$$+ \int u_B(e^*(1,\theta;\rho_0), r^*(1,\theta;\rho_0), \rho_0; \theta) \cdot \frac{d\pi_p(\theta \mid \rho_0)}{d\rho_0}\, d\theta. \tag{5}$$

This expression represents the total effect on mobilized dissent, measured by $c_B^*(\rho_0)$, resulting from an exogeonous change in initial repression, ρ_0. Equations (2)-(5) correspond formally to the different channels by which initial repression can influence bystander participation: anger, strategic feedback with the activist, strategic feedback with the government, and information, respectively. Because of the presence of uncertainty over the government's capacity, θ, each of these channels is averaged over θ according to the bystander's posterior belief, π.

How does one use an exogenous change in initial repression, ρ_0, to identify and measure backlash? Equations (2)-(5) illustrate aspects of the identification problem. First, when all channels do not have the same sign, or when they only share the same sign for some levels of initial repression, then the correlation between initial repression and participation reflects substantive factors potentially associated with backlash as well as features of the particular sample of cases that went into constructing that correlation.

Because the different channels by which backlash could arise, reflected by Equations (2)-(5), identifying backlash requires a more focused empirical approach. In particular, besides just exogenous variation in initial repression, identification requires a more focused research design that manipulates initial repression in a way that activates only one potential channel of backlash at a time.

Observability: Bringing Theory to Data

We begin our analysis from a theoretical perspective and consider when—if backlash is present—a sufficiently focused research design would be expected to detect it. Although an account of when the comparative static associated with backlash, $\frac{dc_B^*(\rho_0)}{d\rho_0} > 0$, can be articulated by a particular theory, to connect such a theory to a research design backlash needs to be produced for a sufficiently broad set of parameters.

Suppose that a researcher posits a theory of dissidence (perhaps by writing a model), where the comparative static $\frac{\partial c_B^*}{\partial \rho_0} > 0$, whether formally expressed or not, holds only when some (observable) parameter, x, is low, and for higher values of x the comparative static flips, and becomes $\frac{\partial c_B^*}{\partial \rho_0} \leq 0$. If the researcher knows the value at which the comparative static flips, $\overline{x}$, then she can adjust her approach to account for cases that fall into the $x < \overline{x}$, and those that fall into $x > \overline{x}$. However, while many theories can produce such an $\overline{x}$ within the theory, they cannot produce an actual value of $\overline{x}$ that would apply to concrete cases. Consequently, it is exceedingly rare that the researcher would observe $\overline{x}$—even with a strong empirical proxy for the theoretical object x, it is not clear how a researcher could observe an endogenously (e.g., strategically) determined cutoff. But, when x is observable, and since backlash is a phenomenon that is produced in the researcher's theory only when x is sufficiently small, the researcher should look for a negative interaction between x and the comparative static associated with backlash, i.e., $\frac{\partial c_B^*}{\partial \rho_0 \partial x} < 0$, reflected, for example, by a negative interaction term in a regression. In this case, backlash is more difficult to observe, but may not be impossible to observe.

Suppose instead that a researcher's theory is that backlash only happens for intermediate values of the observable variable, x.[11] In this case, at low values and high values of x,

11. One could motivate this scenario with the prior distribution, p, where, when the prior expected government capacity for repression, θ, is low, $\frac{\partial c_B^*}{\partial \rho_0} < 0$, when θ is expected to take an intermediate value, $\frac{\partial c_B^*}{\partial \rho_0} > 0$, and if the bystander expects θ is large, $\frac{\partial c_B^*}{\partial \rho_0} < 0$. Such a theory of dissidence then must contend with the question of how a researcher could observe the prior distribution, p, that reflects bystander beliefs.

bystanders are discouraged from participation, but at intermediate values, those within some range $[\underline{x}, \overline{x}]$, bystanders are encouraged to participate. For example, the researcher may posit that bystanders are more willing to participate in dissidence when activist groups have intermediate levels of external funding, but are deterred when activists are either particularly well or poorly funded. This, substantively, may reflect a theory of dissidence where the bystander sometimes free rides on activist effort, not bothering to demonstrate when the government is expected to be (relatively) weak and fearing repression should she demonstrate when the government is expected to be strong. While activist funding may be observable to the researcher, this example highlights how the thresholds $\underline{x}$ and $\overline{x}$, potentially depend not only on funding but also on the bystander's expectations of the activists' strength relative to the government. Based on this alternative dissidence model, backlash can only arise for combinations of funding levels *and* prior distributions, p, where the marginal distribution over θ has certain characteristics. While funding can proxy for activist strength generally, it is not precise enough to pinpoint these strategically determined thresholds and allow for reliable observation. This theoretical model—and a comparative static that depends non-monotonically on some observable parameter—may be justifiable representation of the interaction between a government, an activist, and a bystander, such a theory does not produce empirical implications because its comparative statics simply cannot be identified even with the best of research designs.

These examples illustrate that observability is not purely, or even primarily, a problem of research design. Instead, observing backlash requires a researcher's theory of dissidence has certain features that allows for bridging theory and research design. Observability is satisfied whenever the set where backlash holds, has full measure, meaning that the set $\frac{dc_B^*(\rho_0)}{d\rho_0} \leq 0$ has measure zero. To see the implications of this condition, if one were drawing a prior distribution from $\Delta(\Theta, \Omega)$ at random, then observability requires that the probability one pulls a distribution from the set where backlash holds must be 1. Otherwise, backlash

would not be a reliable enough phenomenon to be empirically assessed since whether it is present. If there are some combinations of parameters such that sometimes $\frac{dc_B^*(\rho_0)}{d\rho_0} > 0$ and other times $\frac{dc_B^*(\rho_0)}{d\rho_0} \leq 0$, whether backlash is present, or whether the reverse is true, relies on features unavailable to the analyst.

Pain

Experiencing or witnessing repression may have a direct, psychological impact on bystanders. This is reflected in the model by the direct influence of ρ_0 on u_B. Although the broader category of psychological responses encompasses many reactions, we focus on *anger* as the predominant form of negative emotions triggered by repression (cf. Aytaç, Schiumerini, and Stokes 2018; Hager and Krakowski 2022). Specifically, we say that backlash arises from an *anger mechanism* when an increase in initial repression ρ_0 increases the threshold $c_B^*(\rho_0)$, through the direct influence of ρ_0 on the bystander's payoff to demonstrating u_B.

To focus on the anger mechanism in isolation, we hold fixed the other potential channels of backlash, which here requires that we fix anticipated activist effort, e, anticipated repression, r, and the state of the world, θ, which, to shutdown information effects, is known by the bystander (i.e., her posterior belief is degenerate). Imagine an experiment that varies the level of initial repression (exogenously) but where the influence of this change can only impact the bystander through her direct experience with initial repression, i.e., how initial repression might impact activist effort, government responsive repression, or the bystander's posterior belief are all shut down. Consequently, any observed variation between ρ_0 and $c_B^*(\rho_0)$ can be attributed to anger.

Proposition 1 *Backlash results as an observable implication from an* ***anger mechanism*** *if and only if when u_B is fixed in activist effort, e, repression, r, and government repression capacity θ, then the function u_B is strictly increasing in ρ_0 for almost all e, r, and θ.*

This result essentially has two parts. First, it gives sufficient conditions for when the bystander's incentives yield backlash after seeing initial repression. That u_B is strictly increasing in ρ_0 implies that an increase in initial repression must make the bystander angry, which means ρ_0 must have a *direct effect* on the bystander's payoff from demonstrating. Moreover, the bystander treats demonstrating as a means to channel her anger about the government's repression. Backlash that emerges from an emotional response requires both *anger* (direct effect) and *agency* (participation offsets direct negative effects). Otherwise, if repression upsets or angers the bystander, regardless of her choice to demonstrate, and if costly demonstrating would not alleviate that anger, and the bystander would not incur a cost to do so.

The second part of Proposition 1 outlines what feature of a theory of backlash motivated by anger must have for it to reflect anger-driven backlash as an empirical implication. It is not enough that $\frac{\partial c_B^*}{\partial \rho_0} > 0$ for some region of parameters, since in this case backlash would not be a reliable enough phenomenon for empirical investigation. Instead, it must be the case that for essentially any values of parameters, we would expect to observe backlash. In practice, this implies that across observations in a sample of potential backlash instances, if backlash follows from an anger mechanism, an experiment that isolates the direct effect of initial repression on subsequent bystander participation in dissent would detect this relationship. A theory that indicates backlash would not hold for some subset of parameters implies those parameters would then confound the relationship between ρ_0 and $c_B^*(\rho_0)$. Proposition 1 thus derives a theoretical implication that an empirical study requires.

Strategic Feedback

The anger mechanism is a *direct* effect of government repression: G represses A, and the repression directly activates the emotional element of the bystander's utility function, to which she responds. Many studies in our review explain backlash through an *indirect* mech-

anism, where repression directly affects the activist's utility function, and then their altered behavior indirectly affects the bystander's likelihood of joining the activist effort.[12] This could be an increase in the efficiency of mobilization efforts or a more compelling reward for participation. We refer to these indirect mechanisms as *strategic feedback* processes.

To isolate the strategic feedback between the bystander and the activist (or government) we need to consider two things. First, how does the bystander's incentive to participate change in the action choice of the activist (government)? Second, how does the action choice of the activist (government) depend on the initial level of repression?

Beginning with the feedback from the activist to the bystander, the level of effort the activist chooses depends on whether the bystander has participated, d, and the bystander's participation decision depends on the effort they anticipate will be provided by the activist. To study this channel in isolation, we hold fixed reactive repression, r, and the state of the world, θ, evaluated at a degenerate posterior belief; this means the actors are optimizing their behavior under complete information and not learning from the choices in the game. We also fix u_B in ρ_0 to shut down the direct (anger) effect analyzed previously. This structure mimics an experiment that isolates the effect of initial repression (ρ_0) on bystander participation through the mediator of (anticipated) activist effort.

Proposition 2 *Backlash results as an observable implication from* ***strategic feedback with the activist****, fixing r, ρ_0, and θ in u_B, if and only if for almost all r and θ, activist effort is strictly increasing (decreasing) in ρ_0 and bystander participation is increasing (decreasing) in e.*

Proposition 2 demonstrates that backlash arises as a result of a strategic feedback with activist effort in two different ways, each of which requires that initial repression has an

12. Some scholars categorize a pattern where repression is correlated with increased activist effort as another form of backlash, which can occur in our model when repression leads the activist to put in more effort, but the bystander choice is fixed. We discuss this pattern below.

indirect effect on the bystander through its direct effect on the activist. First, if initial repression, ρ_0, causes the activist to provide more effort e (a different kind of backlash that we discuss below) the bystander's willingness to demonstrate increases when her participation will complement the effort of the activist. In this case, the effort of the activist is complementary to the participation of the bystander, i.e., higher effort by the activist makes the bystander want to demonstrate more.[13] This could occur, for example, if repression creates a focal point for mobilizing efforts, garners media attention to the activist's cause, or generates donations, all of which enable the activist to increase their level of effort in the second contention stage.

Second, we may still see backlash (that is, a bystander demonstrating when they otherwise would not have) even when initial repression, ρ_0, causes the activist to provide *less* effort. When a higher level of initial repression is demobilizing for the activist, then backlash requires that the bystander must want to participate more as the activist puts in less effort, i.e., $\frac{dc_B^*}{de} < 0$. In this case, the bystander must be more willing to demonstrate under circumstances that hinder the activist, taking on a relatively larger role to compensate for the demobilizing effect of initial repression on activist effort—a form of substitution.

Finally, Proposition 2 establishes that the combination of how the activist responds to initial repression and how the bystander responds to activist effort affect the likelihood of observing backlash together. It must be reliably the case that a research design that treated an activist/bystander pair with repression would yield evidence that the action or anticipated action of the activist triggered the bystander's behavior. In other words, researchers need to expect this strategic feedback to occur so reliably as to expect consistent evidence other for (or against) the presence of the interconnection.

Similar to the strategic feedback induced by activist effort, the government's choice of

13. This is similar to strategic complementarities between bystander participation and activist effort, but differs in that the bystander and activist move sequentially.

reactive repression (r) may also be influenced by initial repression (ρ_0) and thus can also be a source of strategic feedback that leads to backlash. To study this channel in isolation and as distinct from strategic feedback with the activist, we now hold fixed activist effort (e) and the state of the world, θ, evaluated at a degenerate posterior belief. We also fix u_B in ρ_0 as we did above to shut down the direct emotional effect analyzed previously. This structure corresponds to an experiment that isolates the effect of initial repression (ρ_0) on bystander participation through the mediator of (anticipated) reactive government repression.

Proposition 3 *Backlash results as an observable implication from* ***strategic feedback with the government****, fixing e, ρ_0, and θ in u_B, if and only if for almost all r and θ, government repression is strictly increasing (decreasing) in ρ_0 and bystander participation is increasing (decreasing) in r.*

Under Proposition 3, initial repression can alter the government's incentives to repress the activist, which can then have an indirect effect on the bystander's incentive to demonstrate. Backlash can follow from this strategic feedback two ways. First, if an increase in initial repression leads the government to increase repression in the second stage, and if increased anticipated repression in the second period increases the bystander's incentive to demonstrate, then the combined effect is backlash. Similar to above, in this case anticipated government repression and demonstrating are complementary. Second, if the government's repression choice in the final stage decreases with increased initial repression, perhaps because of resource constraints, then backlash requires that as government repression reduces the willingness the bystander to participate. Hence, increased initial repression, ρ_0, implies a lower level of anticipated repression, r, which increases bystander participation, i.e., $c_B^*(\rho_0)$ increases.

Lastly, Proposition 3 shows how the relationship between reactive repression, and how the bystander responds to anticipated reactive repression, must come together within a the-

ory so that the theory produces backlash as an empirical implication. One of these relationships—where repression and bystander participation are complementary or substitute—must hold for virtually all parameters to avoid issues of confounding when the theory is evaluated empirically.

This section highlights two different ways that initial repression alters the anticipated behavior of an activist and the government, each taken in isolation, which then facilitates a strategic feedback on the bystander's participation decision. In certain circumstances, backlash can result from one or both of these strategic channels, an issue we return to later.

Information

Information is the most commonly articulated mechanism in the literature, although whether and how learning about the government and activist can cause backlash is still debated. To isolate and analyze the mechanisms of emotional costs and strategic feedback, we temporarily assumed away any uncertainty for the bystander, holding fixed the government's capacity θ and considering a degenerate posterior belief, which is akin to temporarily asserting complete information. This was necessary to ensure that the effects studied in each of those cases were not dependent on distributional assumptions that are not typically associated with the substantive mechanism of interest.

Now, we reintroduce uncertainty over the government's relative capacity for repression to demonstrate how observing initial repression can provide information that potentially generates backlash. The last channel we study is where the bystander learns about the government's capacity, θ, from initial repression, ρ_0— which is possible whenever there is some correlation between government capacity and initial repression.

We are interested in isolating the informational channel between initial repression and bystander participation, as though we manipulated this mechanism in isolation. To do so we need to hold fixed the direct influence of initial repression on the bystander's incentives

as well as the strategic feedback that follows from activist effort and government reactive repression.

Proposition 4 *Backlash results as an observable implication from an* ***information mechanism****, fixing r, e, and ρ_0 in u_B, if and only if for for almost all r, e, and p, u_B is strictly increasing (decreasing) in θ and $\pi_p(\theta \mid \rho_0)$ is strict monotone likelihood increasing (decreasing) in ρ_0.*

The first part of this result (sufficiency) outlines how backlash can arise through an informational mechanism. In particular, it details what aspects of the informational environment can give rise to backlash, i.e., what signals are produced, and what the bystander is learning about, via the relationship between the state of the world, θ, and the bystander's payoff, u_B.

The bystander learns under one of two conditions that define the relationship between the government's relative capacity to repress and the realization of the level of ρ_0 she observes (that is, her posterior beliefs about θ can update in one of two directions). First, a higher ρ_0 implies a stronger government relative to the activist, i.e., ρ_0 is "good news" about θ. Second, a higher ρ_0 implies a stronger activist relative to the government (lower θ), in which case ρ_0 is "bad news" about θ. The bystander's understanding of the signal of ρ_0 implies one of these relationships between ρ_0 and θ, and this critically informs her choice to demonstrate. Backlash via an information mechanism requires that the signal the bystander observes, ρ_0, affects her willingness to demonstrate monotonically, which is determined by the relationship of θ to u_B. If initial repression makes the bystander believe the government is relatively stronger (the good news case), to observe backlash it must be that the bystander is more willing to participate in dissent, and the reverse must hold if initial repression makes the bystander believe the government is weak.

The second part of Proposition 4 (necessity) shows the link between features of the informational environment and backlash as an empirical phenomenon. In particular, the

relationship between initial repression, ρ_0, and the state of the world, θ, must be stable enough to ensure that initial repression is either always good news, or always bad news, about the state of the world. Moreover, the bystander's payoff must be strictly monotonic in the state of the world. Otherwise, although backlash could result from an informational mechanism, but would not do so in a way that could be empirically detected, i.e., an observed positive correlation between initial repression and participation would be a fluke.

Attribution: Connecting Data to Theory

The ambition of backlash studies is not only to find a particular correlation but to attribute that correlation to a specific substantive mechanism. Backlash is potentially the result of several different substantive channels, highlighted above in Propositions 1-4, each one corresponding to a different mechanism responsible for increased initial repression producing higher participation. But because an experimental manipulation, even an exogenous one, may activate more than one channel at a time, the actual substantive mechanisms responsible for backlash might not be empirically determined or measured. An experimental manipulation that activates multiple channels does not allow a researcher to connect the measured outcome to any particular mechanism. Instead, each potential channel should be studied *in isolation* so that their individual effects can be studied. This is akin to considering an experiment that isolates the effect of repression on the bystander's choice, shutting down all other channels that may affect the demonstration choice. In other words, as we describe when establishing observability, an experiment should be sufficiently focused such that it can be connected to a theory that reliably produces backlash.

What does it mean to have a sufficiently focused research design that activates only one backlash channel? First, a researcher must be able to isolate only one term of the total effect of ρ_0 on bystander participation, either (2), (3), (4), or (5). Doing so distinguishes a single

direct effect of an increase in initial repression on bystander participation. However, our theoretical analysis highlights that a researcher must also be able to hold fixed any indirect channels that affect bystander participation to attribute a change in mobilized dissent to a particular mechanism of backlash. Failing to do so yields a research design that can produce a negative correlation between repression and participation when, in reality, a more focused research designed would have detected a correct positive backlash effect. Therefore, we present an example of how to use our general model to hold mechanisms fixed and derive what the model would predict as backlash with only one mechanism activated.

To illustrate how to isolate a mechanism and identify it in experimental or observed data, consider an example where an analyst aims to identify backlash via strategic feedback with the activist. First, to isolate this strategic feedback channel, it is important to theoretically motivate why initial repression increases activist effort and, therefore, bystander participation. We can do so with the following parameterized model, which is a special case of our general model above for purposes of this illustration. Let the bystander's utility function be:

$$u_B = d\left[e - r + \theta + \rho_0 - c_B\right].$$

To isolate each parameter in turn as we would in a focused experiment, $\frac{\partial u_B}{\partial e} = 1$, $\frac{\partial u_B}{\partial r} = -1$, $\frac{\partial u_B}{\partial \theta} = 1$, $\frac{\partial u_B}{\partial \rho_0} = 1$. This utility function satisfies all the assumptions on the bystander's utility necessary for backlash via each of the individual mechanisms outlined in Propositions 1-4. A research design that focused on any one of these channels could possibly observe backlash and attribute it to one of the four backlash mechanisms.

The activist and government, in this example, choose effort and repression, respectively, in part to match the other's choice; their payoffs are decreasing in the distance between effort

and repression. The activist's utility function is

$$u_A = (e - \theta)^2 - (e - r)^2 + e\rho_0 - (1 - d)c_A,$$

while the government's utility function is

$$u_G = -(r - \theta)^2 - (r - e)^2 - r\rho_0 - (1 + d)c_G.$$

There are three key differences between these functions. First, the government chooses repression to match its capacity to repress, θ. When the government is relatively more capable, the government chooses more repression in the final contention stage. The activist chooses an effort to mismatch the state, so when the government is relatively less capable, the activist exerts more effort. Second, the effect of initial repression on their utilities differs. Initial repression is motivating for the activist but disincentivizing for the government.[14] This captures a scenario where the government has a budget for repression and, having expended substantial resources repressing prior to the bystander's choice, can only afford a lower level of repression in the final contention stage. This satisfies the other conditions necessary to identify backlash via strategic feedback mechanisms. Finally, in this illustration, we introduce a cost to the activist and the government in the final contention stage.[15] For the activist, when the bystander demonstrates, this reduces their cost of the final contention stage. The opposite is true for the government. When the bystander demonstrates, government costs increase. These assumptions fix the parameters in the two functions to feature how bystander choice feeds back to improve the activist's strategic advantage.

Imagine that, for whatever reason, the analyst is unable to fix r (the government's contribution) to conduct a focused experiment that can identify the backlash through strategic

14. For the activist, $\frac{\partial u_A}{\partial \rho_0} = e$, and for the government $\frac{\partial u_G}{\partial \rho_0} = -r$.
15. Incorporating these costs in the general model are straightforward and have no effect on results.

feedback with the activist; Holding r fixed is a requirement to observe the backlash and attribute it to strategic feedback in Proposition **??**. If r is not fixed, instead of considering how an increase in ρ_0 affects bystander participation only through its effect on activist effort, we must now consider the effect of a change in ρ_0 on e^*, accounting for how e^* responds to the government's choice of repression. In other words, we must evaluate the effect of ρ_0 on (e^*, r^*) together instead of just e^*.

The Nash equilibrium of the contention stage of our illustrative example is the pair $(e^*, r^*) = (\frac{1}{2}(2\theta - \rho_0), \frac{1}{2}(2\theta - \rho_0))$. Considering an increase in the level of initial repression, $\frac{\partial e^*}{\partial \rho_0} = -\frac{1}{2}$ and $\frac{\partial r^*}{\partial \rho_0} = -\frac{1}{2}$. In equilibrium, an increase in ρ_0 leads to a decrease in both e^* and r^*. What is the effect of reducing both e^* and r^* on the bystander? In this example, backlash results from strategic feedback with the government because *bystander participation increases as expected repression decreases.* This continues to hold because the result, in this case, did not rely on fixing r.

When activist effort decreases, we see a *decrease* in the bystander's willingness to demonstrate via strategic feedback with the activist. This implies we do not see backlash via this mechanism and, moreover, we may not be able to sign the total effect of an increase in initial repression, given by Equations (2)-(5). However, our illustrative model shows us that this conclusion is explicitly an artifact of failing to isolate the strategic feedback channel. If we were able to keep r fixed, we would correctly identify the backlash through this strategic feedback mechanism.

As this example illustrates, strategic feedback channels pose the most pernicious problems for a researcher unable to implement a focused research design.

Proposition 5 *A research design cannot attribute a correlation between initial repression and bystander participation to a single backlash mechanism if e or r are not held fixed.*

To attribute backlash to the anger or information mechanisms, the researcher must hold fixed both activist effort, e, and reactive repression, r. Attributing backlash to strategic

feedback with the activist (government) requires holding repression (effort) fixed. When an individual mechanism—anger, strategic feedback, or information—can be isolated, it is possible to both attribute a correlation between initial repression and subsequent bystander participation to backlash, and to determine which theoretical mechanism drove the change in the bystander's behavior.

In many circumstances, it is challenging to hold fixed effort and repression. This naturally raises the question of what can be learn about backlash via an imperfect, only partially focused experiment. Our approach of marrying theory and research design allows us to articulate both how to proceed in conducting an imperfect experiment and precisely why it is important to be cautious in making conclusions about backlash once we have done so.

Returning to our illustrative example where the researcher is unable to fix reactive repression, r, to identify backlash, theory can help make up for the gap between the ideal focused experiment and the feasible unfocused experiment. In our illustrative example, activist effort and government repression in the final stage are strategic complements, and it is this feature of the model that in part generates contradictory findings about backlash. Imagine we rewrote the activist's utility function as

$$\tilde{u}_A = (e-\theta)^2 - (e-r)^2 + e\rho_0 - (1-d)c_A,$$

and the government's utility function as

$$\tilde{u}_G = -(r-\theta)^2 - (e+r)^2 - r\rho_0 - (1+d)c_G.$$

With these new utility functions, optimal effort and repression are given by the pair $(e^*, r^*) = (\frac{\rho_0}{2} - \theta, \frac{1}{2}(2\theta - \rho_0))$, and in equilibrium, an increase in ρ_0 increases activist effort ($\frac{\partial e^*}{\partial \rho_0} = \frac{1}{2}$) and decreases repression ($\frac{\partial r^*}{\partial \rho_0} = -\frac{1}{2}$). A model with these utility functions for the activist and government therefore can identify backlash resulting from strategic feedback even when

we cannot conduct the ideal experiment. What is left for the analyst is to show why this model, with $\tilde{u}_A$ and $\tilde{u}_G$, is the correct model of dissidence in the context of study. With supplemental evidence that the activist and government act as defined in this new illustrative model, e.g., where effort and repression are substitutes and the government aims to minimize the total contention in the final stage, we may be more confident that a correlation between initial repression and bystander participation can be attributed to backlash via strategic feedback with the activist.

Alternate Forms of Backlash

Our main analysis focuses on backlash as a response to changes in the initial level of repression, measured by the comparative static $\frac{dc_B^*(\rho_0)}{d\rho_0} > 0$. Our model of dissidence, however, contains two alternate outcomes that might also be referred to as "backlash." First, how does ρ_0 influence the effort choice of the activist, i.e., when is $\frac{de^*(\rho_0)}{d\rho_0} > 0$? Second, as a theoretical concern, does the increased anticipation of repression, i.e., increased r^*, also change c_B^*?

Beginning with the former, backlash of the form of a direct increase in effort by the activist following initial repression has been studied in the literature. Our framework can straightforwardly accommodate this type of backlash, as it is related to the strategic feedback mechanisms we identify for the bystander.

Proposition 6 *Backlash results from a* ***direct change in activist effort*** *if and only if, when u_A is fixed in repression, r and the bystander's choice, d, then the function u_A is strictly increasing in ρ_0 for almost all r and d.*

Comparing Proposition 6 to Proposition **??**, identifying backlash resulting from a change in activist effort requires fewer assumptions than identifying backlash resulting from strategic

feedback with the activist. In both cases, we must hold fixed the level of repression, r. Identifying activist-driven backlash only requires also holding fixed the bystander's demonstration choice, an observable action in many real-world contexts. By contrast, to identify backlash resulting from strategic feedback with the activist, we must also fix θ and shut down the direct effect of ρ_0, which generates backlash via the anger mechanism. Short of the ideal experiment, it is less demanding to identify backlash resulting from activist effort than any backlash resulting from a change in the bystander's behavior.

Turning to backlash driven by anticipation of repression, this is something which is not easily addressed empirically as it relies on the bystander's response to her perception of increased future repression (which may never materialize). When the bystander expects a high level of future repression, she may be deterred from demonstrating. In this case, we would not observe backlash as we have conceptualized it—where the bystander is more willing to demonstrate—because the bystander stays home. Anticipating future repression, therefore, introduces selection bias. Theoretically, we can capture a deterrent effect of anticipated future repression through the direct effect of r^* on c_B^*, and we can separate this from backlash, operating through the downstream effects of a change in ρ_0. When we observe bystander behavior, however, anticipated future repression deterring participation is indistinguishable from (negative) strategic feedback with the government, where an increase in ρ_0 decreases anticipated repression, and therefore decreases the bystander's willingness to demonstrate, as in Proposition **??**.

Conclusion

Backlash to repression represents both an empirical pattern 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 existing literature, emotional responses, gains in organizational capacity 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. Furthermore, 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.

Overall, richer theory can, in part, compensate for deviations from an ideal experiment. It is then incumbent on the analyst to validate the additional assumptions required in the richer theoretical model. By providing supplemental evidence that supports the validity of the refined theory, we can be more confident in any empirical observation of backlash. It is not the case, however, that all richer theoretical models facilitate attribution of backlash to one of the four mechanisms. In fact, in some cases a richer theoretical model can undermine any effort to identify backlash.

References

al-Anani, K. 2019. "Rethinking the Repression-Dissent Nexus: Assessing Egypt's Muslim Brotherhood's Response to Repression since the Coup of 2013." *Democratization* 26, no. 8 (November): 1329–1341.

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

Bueno De Mesquita, E., and S. A. Tyson. 2020. "The commensurability problem: Conceptual difficulties in estimating the effect of behavior on behavior." *American Political Science Review* 114 (2): 375–391.

Carroll, R., and A. Pond. 2021. "Costly signaling in autocracy." *International Interactions* 47 (4): 612–632.

Carter, E. B., and B. L. Carter. 2022. "When Autocrats Threaten Citizens with Violence: Evidence from China." *British Journal of Political Science* 52, no. 2 (April): 671–696.

Chenoweth, E., and M. J. Stephan. 2011. *Why civil resistance works: The strategic logic of nonviolent conflict.* 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.

Collier, P., and A. Hoeffler. 2004. "Greed and grievance in civil war." Tex.date-added: 2011-06-26 11:08:28 -0500 tex.date-modified: 2011-06-26 11:08:28 -0500, *Oxford Economic Papers* 56 (4): 563–595.

Conrad, C. R., and W. H. Moore. 2010. "What stops the torture?" *American Journal of Political Science* 54 (2).

Curtice, T. B., and B. Behlendorf. 2021. "Street-level Repression: Protest, Policing, and Dissent in Uganda." *Journal of Conflict Resolution* 65, no. 1 (January): 166–194.

Dafoe, A., S. Hatz, and B. Zhang. 2021. "Coercion and Provocation." Publisher: SAGE Publications Inc, *Journal of Conflict Resolution* 65, nos. 2-3 (February): 372–402.

Davenport, C. 2007. "State Repression and Political Order." *Annual Review of Political Science* 10, no. 1 (June): 1–23.

———. 2015. *How Social Movements Die: Repression and Demobilization of the Republic of New Africa.* New York: Cambridge University Press.

De Jaegher, K., and B. Hoyer. 2019. "Preemptive Repression: Deterrence, Backfiring, Iron Fists, and Velvet Gloves." *Journal of Conflict Resolution* 63, no. 2 (February): 502–527.

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

Edwards, P., and D. Arnon. 2021. "Violence on Many Sides: Framing Effects on Protest and Support for Repression." *British Journal of Political Science* 51, no. 2 (April): 488–506.

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." *Journal of Conflict Resolution* 67, nos. 7-8 (August): 1592–1617.

Gibilisco, M. 2021. "Decentralization, Repression, and Gambling for Unity." *The Journal of Politics* 83, no. 4 (October): 1353–1368.

Gohdes, A. R. 2020. "Repression Technology: Internet Accessibility and State Violence." *American Journal of Political Science* 64, no. 3 (July): 488–503.

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.

Haffert, L. 2022. "The Long-Term Effects of Oppression: Prussia, Political Catholicism, and the Alternative für Deutschland." *American Political Science Review* 116, no. 2 (May): 595–614.

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." *Mobilization* 11 (2): 249–267.

Komisarchik, M., M. Sen, and Y. R. Velez. 2022. "The Political Consequences of Ethnically Targeted Incarceration: Evidence from Japanese American Internment during World War II." *The Journal of Politics* 84, no. 3 (July): 1497–1514.

Koopmans, R. 1997. "Dynamics of repression and mobilization: The german extreme right in the 1990s." Tex.date-added: 2017-03-09 18:25:41 +0000 tex.date-modified: 2017-03-09 18:26:38 +0000, *Mobilization* 2 (2): 149–164.

Kuran, T. 1991. "Now out of never: The element of surprise in the east european revolution of 1989." *World Politics* 44 (1): 7–48.

Lichbach, M. I. 1987. "Deterrence or escalation? The puzzle of aggregate studies of repression and dissent." *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." *Conflict Management and Peace Science* 40, no. 4 (July): 441–463.

Lohmann, S. 1993. "A signaling model of informative and manipulative political action." *American Political Science Review* 87 (2): 319–333.

———. 1994. "The dynamics of informational cascades: The Monday Demonstrations in Leipzig, East Germany, 1989-1991." *World Politics* 47 (1): 42–101.

Moore, W. H. 1998. "Repression and Dissent: Substitution, Context, and Timing." *American Journal of Political Science* 42 (3): 851–873.

Pearlman, W. 2018. "Moral Identity and Protest Cascades in Syria." *British Journal of Political Science* 48, no. 4 (October): 877–901.

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.

Pop-Eleches, G., and L. A. Way. 2021. "Censorship and the Impact of Repression on Dissent." *American Journal of Political Science* (September): ajps.12633.

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. 2014. "Policy disputes, political survival, and the onset and severity of state repression." Tex.date-added: 2014-06-19 20:53:35 +0000 tex.date-modified: 2014-06-19 20:53:35 +0000, *Journal of Conflict Resolution* 58 (1): 143–168.

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.

Rozenas, A., S. Schutte, and Y. M. Zhukov. 2017. "The Political Legacy of Violence: The Long-Term Impact of Stalin's Repression in Ukraine." *The Journal of Politics* 79, no. 4 (October): 1147–1161.

Rozenas, A., and D. Stukal. 2019. "How Autocrats Manipulate Economic News: Evidence from Russia's State-Controlled Television." *The Journal of Politics* 81, no. 3 (July): 982–996.

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

Shadmehr, M. 2014. "Mobilization, Repression, and Revolution: Grievances and Opportunities in Contentious Politics." *The Journal of Politics* 76, no. 3 (July): 621–635.

Shadmehr, M., and D. Bernhardt. 2011. "Collective action with uncertain payoffs: Coordination, public signals, and punishment dilemmas." Tex.date-added: 2015-12-09 22:56:09 +0000 tex.date-modified: 2015-12-09 22:56:09 +0000, *American Political Science Review* 105 (4): 829–851.

Shadmehr, M., and R. Boleslavsky. 2022. "International Pressure, State Repression, and the Spread of Protest." *The Journal of Politics* 84, no. 1 (January): 148–165.

Simmons, B. A. 2009. *Mobilizing for human rights: International law in domestic politics.* Tex.date-added: 2011-06-26 11:08:28 -0500 tex.date-modified: 2011-06-26 11:08:29 -0500. Cambridge, MA: Cambridge University Press.

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

Strauss, C. 2018. "Engaged by the spectacle of protest: How Bystanders became invested in Occupy Wall Street." In *Political Sentiments and Social Movements: The Person in Politics and Culture,* edited by C. Strauss and J. R. Friedman, 33–60. Beverly Hills: Springer International Publishing.

Sullivan, C. M., and C. Davenport. 2018. "Resistance is mobile: Dynamics of repression, challenger adaptation, and surveillance in US 'Red Squad' and black nationalist archives." *Journal of Peace Research* 55, no. 2 (March): 175–189.

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

Tertytchnaya, K. 2023. ""This Rally is Not Authorized": Preventive Repression and Public Opinion in Electoral Autocracies." *World Politics* 75 (3): 482–522.

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

Thaler, K. M., L. Mueller, and E. Mosinger. 2023. "Framing Police Violence: Repression, Reform, and the Power of History in Chile." *The Journal of Politics* 85, no. 4 (October): 1198–1213.

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.

Woodward, J. 2005. *Making Things Happen: A Theory of Causal Explanation.* Oxford University Press.

Supporting Information for Backlash

Contents

A Proofs

Proof of Lemma 1: Follows by direct application of Glicksberg's Theorem. ■
Proof of Proposition 1: By smoothness of u_B and the Implicit Function Theorem, the function $\frac{\partial c_B^*}{\partial \rho_0} : \mathcal{P} \equiv [0, \overline{e}] \times [0, \overline{r}] \times \Theta \times \Omega \to \mathbb{R}$ is a smooth function. Define the set

$$E_+ = \left\{ x = (e, r, \theta, \rho_0) \;\middle|\; \frac{\partial c_B^*}{\partial \rho_0} > 0 \right\} \subset \mathcal{P},$$

which is the preimage of the positive reals in $\mathcal{P}$. We begin with sufficiency, i.e., if u_B is strictly increasing in ρ_0, then E_+ has full measure. Application of the Implicit Function Theorem to the characterization of the bystander's cutoff, (1), establishes that, since u_B is strictly increasing, c_B^* is strictly increasing in ρ_0 everywhere, and hence, E_+ has full measure.

Second, to establish necessity, suppose that E_+ has full measure and that there exists a positive measure set, Z, such that for $x \in Z$, u_B is nonincreasing in ρ_0 at x. By the Implicit Function Theorem and the characterization of the bystander's cutoff, (1), it must be that $\frac{\partial c_B^*}{\partial \rho_0} \leq 0$ at x, and hence, $x \in \mathcal{P} \setminus E_+$, implying that $Z \subset \mathcal{P} \setminus E_+$, which has measure zero, contradicting that Z has positive measure. ■
Proof of Proposition ??: (Sketch) Define the set

$$F_+ = \left\{ x = (r, \rho_0, \theta) \;\middle|\; \frac{\partial c_B^*}{\partial e} \cdot \frac{de^*}{d\rho_0} > 0 \right\} \subset \mathcal{P} = [0, \overline{r}] \times [0, \overline{\rho}] \times \Theta.$$

We begin with sufficiency, i.e., if u_B is strictly increasing in e, and e^* is strictly increasing in ρ_0, then F_+ has full measure. This follows by application of the chain rule and the Implicit Function Theorem to the characterization of the bystander's cutoff, (1).

Second, to establish necessity, suppose that F_+ has full measure and that there exists a positive measure set, Z, such that for $x \in Z$, u_B is nonincreasing in e at x when e^* is strictly increasing in ρ_0. By the chain rule and the Implicit Function Theorem, with the characterization of the bystander's cutoff, (1), at x, it must be that $\frac{\partial c_B^*}{\partial e} \cdot \frac{de^*}{d\rho_0} \leq 0$, and hence, $x \in \mathcal{P} \setminus F_+$, implying that $Z \subset \mathcal{P} \setminus F_+$, which has measure zero, contradicting that Z has positive measure. ■
Proof of Proposition ??: (Sketch) Define the set

$$G_+ = \left\{ x = (e, \rho_0, \theta) \;\middle|\; \frac{\partial c_B^*}{\partial r} \cdot \frac{dr^*}{d\rho_0} > 0 \right\} \subset \mathcal{P} = [0, \overline{e}] \times [0, \overline{\rho}] \times \Theta.$$

We begin with sufficiency, i.e., if u_B is strictly increasing in r, and r^* is strictly increasing in ρ_0, then G_+ has full measure. This follows by application of the chain rule and the Implicit Function Theorem to the characterization of the bystander's cutoff, (1).

Second, to establish necessity, suppose that G_+ has full measure and that there exists a positive measure set, Z, such that for $x \in Z$, u_B is nonincreasing in r at x when r^* is strictly increasing in ρ_0. By the chain rule and the Implicit Function Theorem, with the characterization of the bystander's cutoff, (1), at x, it must be that $\frac{\partial c_B^*}{\partial r} \cdot \frac{dr^*}{d\rho_0} \leq 0$, and hence,

$x \in \mathcal{P} \setminus G_+$, implying that $Z \subset \mathcal{P} \setminus G_+$, which has measure zero, contradicting that Z has positive measure. ■

Proof of Proposition 4: (Sketch) Define the set

$$H_+ = \left\{ x = (e, r, \rho_0, \theta) \;\middle|\; \frac{\partial c_B^*}{\partial r} \cdot \frac{dr^*}{d\rho_0} > 0 \right\} \subset \mathcal{P} = [0, \overline{e}] \times [0, \overline{r}] \times [0, \overline{\rho}] \times \Theta.$$

We begin with sufficiency, i.e., if u_B is strictly increasing in θ, and if π is strictly increasing in the monotone likelihood ratio order, then by **?**, Proposition 2, H_+ has full measure. This follows by application of the Implicit Function Theorem to the characterization of the bystander's cutoff, (1).

Second, to establish necessity, suppose that H_+ has full measure and that there exists a positive measure set, Z, such that for $x \in Z$, u_B is nonincreasing in θ, or π is strictly increasing in the monotone likelihood ratio order. By the Implicit Function Theorem, with the characterization of the bystander's cutoff, (1), at x, it must be that $\frac{\partial c_B^*}{\partial r} \cdot \frac{dr^*}{d\rho_0} \leq 0$, and hence, $x \in \mathcal{P} \setminus G_+$, implying that $Z \subset \mathcal{P} \setminus H_+$, which has measure zero, contradicting that Z has positive measure. ■

Proof of Proposition 5: When e and r are held fixed at their equilibrium levels, then $\frac{de^*}{dr} = \frac{dr^*}{de} = 0$, recovering (3) and (4) and satisfying attribution to strategic feedback. Relaxing this assumption, define the set

$$J_+ = \left\{ x = (e, r, \rho_0, \theta) \;\middle|\; \frac{dc_B^*}{d\rho_0} > 0 \right\} \subset \mathcal{P} = [0, \overline{e}] \times [0, \overline{r}] \times [0, \overline{\rho}] \times \Theta.$$

Suppose J_+ has full measure. Given u_A and u_G define the positive measure set Z such that, for $x \in Z$, $\frac{de(1,\theta;\rho_0)}{d\rho_0} > 0$ and $\frac{dr(1,\theta;\rho_0)}{d\rho_0} < 0$ at x. Then, take the derivative

$$\begin{aligned}
\frac{dc_B^*(\rho_0)}{d\rho_0} = &\int \frac{\partial u_B(e^*(1,\theta;\rho_0), r^*(1,\theta;\rho_0), \rho_0; \theta)}{\partial \rho_0} \pi_p(\theta \mid \rho_0)\, d\theta \\
&+ \int \frac{\partial u_B(e^*(1,\theta;\rho_0), r^*(1,\theta;\rho_0), \rho_0; \theta)}{\partial e^*} \cdot \left(\frac{de^*(1,\theta;\rho_0)}{d\rho_0} + \frac{de^*}{dr^*} \frac{dr^*(1,\theta;\rho_0)}{d\rho_0} \right) \pi_p(\theta \mid \rho_0)\, d\theta \\
&+ \int \frac{\partial u_B(e^*(1,\theta;\rho_0), r^*(1,\theta;\rho_0), \rho_0; \theta)}{\partial r^*} \cdot \left(\frac{dr^*(1,\theta;\rho_0)}{d\rho_0} + \frac{dr^*}{de^*} \frac{de^*(1,\theta;\rho_0)}{d\rho_0} \right) \pi_p(\theta \mid \rho_0)\, d\theta \\
&+ \int u_B(e^*(1,\theta;\rho_0), r^*(1,\theta;\rho_0), \rho_0; \theta) \cdot \frac{d\pi_p(\theta \mid \rho_0)}{d\rho_0}\, d\theta.
\end{aligned}$$

where the additional terms follow by application of the chain rule. Then for some for $x \in Z$, $\frac{dc_B^*(\rho_0)}{d\rho_0} \leq 0$, contradicting that J_+ has full measure. For fixed e and r such that, $\frac{de^*}{dr} = \frac{dr^*}{de} = 0$, Propositions **??** and **??** provide necessary and sufficient conditions for J_+ to have full measure for all $x \in Z$. ■