Revealed and Concealed Repression:

Measurement, Deterrence, and Backlash

Maria Titova[1] Emily Hencken Ritter[2] Mehdi Shadmehr[3]

[1]Department of Economics, Vanderbilt University. E-mail: motitova@gmail.com

[2]Department of Political Science, Vanderbilt University. E-mail: emily.h.ritter@vanderbilt.edu

[3]Department of Public Policy and Department of Economics, UNC Chapel Hill. E-mail: mshadmeh@gmail.com

Abstract

Regimes routinely conceal acts of repression. We show that observed repression may be negatively correlated with total repression—which includes both revealed and concealed acts—across time and space. This distortion implies that policy interventions aimed at reducing repression by incentivizing regimes can produce perverse effects. It also poses challenges for research evaluating the efficacy of repression—its deterrent and backlash effects. To address this, we develop a model in which regimes choose both whether to repress and whether to conceal repression. We leverage equilibrium relationships to propose a method for recovering concealed repression using observable data. We then provide an informational theory of deterrence and backlash effects, identifying the conditions under which each arises and intensifies. Finally, we show that comparing protest probabilities in the presence and absence of repression provides an upper bound on the size of the backlash effect, overstating its magnitude and thereby underestimating the efficacy of repression.

Keywords: Protest, Repression, Deterrence, Backlash, Signaling, Disclosure Games

Word count: 10,929

1. Introduction

Regimes routinely conceal acts of repression (Guriev and Treisman 2022). For example, the Iranian government has tried to hide the 1988 prison executions (Abrahamian 1999; Amnesty International 2018), and the Chinese government has similarly tried to conceal the 1989 Tienanmen Square massacre (Reporters Without Borders 2008). Regimes remove social media posts about government arrests (King, Pan, and Roberts 2013), punish dissidents for fabricated nonpolitical crimes when they could punish them for actual political offenses (Pan, Xu, and Xu 2025), and employ tactics such as stealth torture that are difficult to detect (Conrad, Hill, and Moore 2018; Rejali 2007). For example, during the Dirty War, the Brazilian military regime instructed its torturers to "press hard without leaving marks" (Dassin 1998, p. 180). Regimes also sometimes reveal acts of repression. Examples include the Saudi regime's mass execution of Shi'i dissidents (Human Rights Watch 2022), the Iranian regime's execution of protesters in the aftermath of a public challenge in 2022 (Amnesty International 2023), and the Chinese regime's execution of Uyghur activists (Guardian 2009). Each case was reported, informing the public about the use of violence against political opponents.

Because regimes sometimes conceal repression, observed repression is generally a distorted measure of total repression and its trend. Total repression, consisting of both concealed and revealed repression, can be significantly higher than observed repression alone. Importantly, the trends of total and observed repression can even go in *opposite* directions. If a government shifts tactics from revealed repression to concealed repression, it may produce the appearance of declining repressive events, even as the total extent of repression increases. Consequently, policy interventions based on observed repression and its trend can be misguided and ineffective. The international community spends sig-

nificant resources, in the form of sanctions, aid, or soft power, to mitigate repression.[1] However, governments and international organizations commonly use measures of observed repression (e.g., Human Rights Watch reports, Freedom House indices, and UN Universal Periodic Reports) to inform policies aimed at mitigating human rights abuses (Andirin et al. 2022; Lebovic and Voeten 2009; Poe 1992).[2]

Regimes' strategic decisions to conceal repression also create challenges for research on state–society interactions (Davenport 2007; Earl 2011; Guriev and Treisman 2022; Hassan, Mattingly, and Nugent 2022; Moore 1998, 2000). Concealed repression complicates efforts to identify the consequences of repression, specifically its deterrence and backlash effects. The potential for concealment shapes how the public interprets and responds both to observed government repression and its absence. Researchers, in turn, must identify the causal effects of repression even when both they and the public remain uncertain about the occurrence of unobserved, concealed repression.

We develop a model to analyze strategic environments in which regimes choose both whether to repress and whether to conceal repression. Our contribution is two-fold. First, we identify conditions under which observed and total repression may be negatively correlated and show how concealed repression can be recovered from observable variables by leveraging equilibrium relationships. Second, we highlight how concealed repression com-

[1]Examples of U.S. legislation include the Global Magnitsky Human Rights Accountability Act of 2016, Foreign Assistance Act of 1961, section 502B, as amended, and the Leahy laws. Examples of international organizations resources include the World Bank's Human Rights trust fund (`https://www.worldbank.org/en/programs/humanrights`) and the European Union's Human Rights and Democracy Thematic Programme, and its predecessor, the European Instrument for Democracy and Human Rights.

[2]Lebovic and Voeten (2009) find the World Bank reduces aid to those countries shamed by the UN Human Rights Commission/Council for rights violations. Andirin et al. (2022, p. 7) note that "[T]he US Millennium Challenge Corporation incorporates Freedom House indices into its criteria to determine a country's eligibility for assistance (Millennium Challenge Corporation 2020). Canada's Country Indicators for Foreign Policy project integrates Freedom House indicators into data aimed at providing guidance to development-agency staff (Carment 2010). The Open Government Partnership Global Report cites Freedom House data in the context of identifying potential areas for future work and improvement (Open Government Partnership 2019, pp. 72, 78, 96)."

plicates the estimation of deterrent and backlash effects, leading to incorrect conclusions about the efficacy of repression. We then employ our model to provide an informational theory of deterrence and backlash, identifying the conditions under which each effect tends to arise and intensify. We further show that comparing the probability of protest in the presence and absence of observed repression yields an upper bound on the magnitude of the backlash effect, overstating the backlash and therefore underestimating the true efficacy of repression.

In our framework, dissidents may have an opportunity to organize and make demands on the state to change the status quo. These dissidents include activists, members of the opposition, or more broadly the "vanguards" or "early risers" in a movement (Shadmehr and Bernhardt 2019; Tarrow 2011). We refer to them collectively as *activists*. If the activists organize, the regime decides whether to concede or repress the activists and whether to reveal repression or to conceal it at a cost. The public observes concession, and it observes repression if it is revealed. The public then decides whether to protest.

The public is uncertain about the legitimacy of the state's use of coercive force.[3] In particular, as in Shadmehr and Boleslavsky (2022), the public is uncertain whether activists' policy demands are beneficial or harmful to them, reflecting the uncertainties involved in major policy changes. Our model has two novel ingredients: the state's decision to conceal its own repression, and the public's inference about the presence of organized activists. Because the government may conceal repression, the public remains uncertain about whether activists have organized unless it observes repression or conces-

[3]The public expects the state to use coercion to protect it when necessary (Almond 1956; Mansbridge 2012, 2014; Opp and Ruehl 1990), and therefore does not universally view coercion as illegitimate. Moreover, public opinion about activists, protesters, opposition groups, and the state itself shifts in response to observed interactions between them (Chen and Huo 2025; Chenoweth 2023; Chenoweth and Stephan 2011; Hager and Krakowski 2022; Tertytchnaya 2023; Wasow 2020). As captured in the phrase "the revolution devours its children," regime changes can produce outcomes far worse than anticipated (Shadmehr and Bernhardt 2011). These observations underscore the public's uncertainty.

sion. If the public protests when the activists have organized, the protest succeeds and the activists' demands are implemented. Otherwise, the protest fizzles and the status quo remains, reflecting the role of leadership and organizational resources in sustaining social movements (Davenport 2015; Sullivan 2016; Tarrow 2011; Tilly 1978).

The public does not take the desirability of change, the legitimacy or illegitimacy of state coercion, or the organizational capacity of the opposition as given. Instead, the public makes inferences about these features. Revealed repression—or its absence—provides the public with information about two key aspects: (1) whether there are organized activists and a mobilization is underway, and (2) the beneficial or harmful nature of the activists' demands. In making its decision, the regime weighs the costs of concession, concealment, and the potential spread of mobilization from activists to the general public. The public, in turn, weighs the downside risk of protest if the activists' demands would worsen the status quo, the upside if their demands would improve it, the likelihood of success, and the direct costs of protesting.

Our first set of results concerns the measurement of repression when the regime strategically conceals it. We show that lower concealment costs increase total repression but reduce revealed repression—creating the appearance that the regime is becoming more tolerant. Greater ease of concealment leads the regime to substitute away from revealed and toward concealed repression, while also making repression more attractive overall. Conversely, as concealment becomes more costly, total repression may decline even as revealed repression increases—making the regime appear less tolerant. Similar patterns emerge in response to changes in protest costs. In sum, observed and total repression can move in opposite directions in response to environmental changes. Thus, by focusing on observable repression, existing measures not only overlook concealed acts but may also lead to opposite conclusions about the true trend in total repression. We address this problem by leveraging the equilibrium relationships between unobserved total repression

and observable parameters. We show how total repression can be recovered from observed repression and the public's beliefs, which can be measured through surveys.[4]

Our second set of results is about the deterrence and backlash effects of repression (Aytaç, Schiumerini, and Stokes 2018; Davenport 2007; Earl 2011; Francisco 2004; LaFree, Dugan, and Korte 2009; Martin 2007). We define backlash (deterrence) as occurring whenever revealed repression causes an increase (decrease) in the likelihood of public protest relative to the public protest that would have occurred when the public knows that the regime has not engaged with the activists—that is, under the prior. We distinguish between two settings: (1) when the regime and the public have low conflict of interest, so the regime is more reluctant to concede to activists with demands that are *harmful* to the public, and (2) when they have high conflict of interest, so the regime is more reluctant to concede to activists with *beneficial* demands.

We show that when the conflict of interest with the public is low, observed repression tends to have a deterrence effect unless the public believes that activists are unlikely to mobilize, so the likelihood of protest absent observed repression is already low. However, when the conflict of interest between the regime and the public is high, observed repression tends to have a backlash effect, because the public updates to believe the demands of the repressed activists are likely to benefit them.

In settings with low conflict of interest, where deterrence and backlash effects can occur, we show that deterrence arises in a larger subset of parameters and that the deterrence (backlash) effect is larger (smaller) when (i) the public believes activists are more capable of mobilization, (ii) the public believes activists are less likely to demand harmful

[4]We follow the structural estimation approach, in which unobserved parameters are estimated based on their equilibrium relationships with observed parameters in a proposed model. This method has been used in ideal point estimation (McCarty, Poole, and Rosenthal 2006), bargaining models (Diermeier, Eraslan, and Merlo 2003; Signorino 1999), third-party intervention in conflicts (Gibilisco and Montero 2022), and the measurement of repression (Andirin et al. 2022).

policy changes, or (iii) protest costs are lower (in the first-order stochastic dominance sense).[5] These conditions tend to increase the likelihood of protest when the public believes that the regime has not engaged with the activists—the relevant counterfactual. However, they do not affect the likelihood of protest following revealed repression. The latter is less sensitive to environmental changes because the regime acts strategically and it can always choose to concede. It adjusts its behavior in response to environmental shifts so that the probability of protest following revealed repression is determined by the cost of concession—any higher, and the regime would prefer to concede. Finally, we show that the difference in the probability of protest in the presence and absence of observed repression is an upper bound on the size of the backlash effect. Therefore, estimates based on this difference tend to overestimate the backlash effect and, consequently, underestimate the efficacy of repression.

1.1. Literature and Contributions

Our paper contributes to studies of measurement problems in political economy when the variable of interest is partially unobserved under conditions that are *a priori* unknown. An example is when actors strategically opt out of an interaction. Heckman-type selection models offer procedures to reduce bias in inference through statistical modeling of the selection process (Wooldridge 2010). Ritter (2014) and Conrad and Ritter (2019) take this approach to analyze the determinants of repression. In contrast, Andirin et al. (2022) take a structural approach to measuring unobserved repression, mapping unobserved preventive repression into a model parameter that they estimate. We also take a structural approach, using the equilibrium relationship between observable variables to recover the

[5] In settings with high conflict of interest, where only the backlash effect occurs, we show that the size of the backlash effect is smaller when the public believes activists are more capable of mobilization or less likely to demand harmful policy changes.

unobserved repression. When a one-to-one mapping is not available, we show that the standard approach of comparing protest probabilities in the presence and absence of repression yields an upper bound on the size of the backlash effect—thereby overstating its magnitude and underestimating the efficacy of repression.

Recovering unobserved total repression from observable parameters is especially important when observed and unobserved repression are negatively correlated across space and time. In such cases, relying solely on observed repression can lead to perverse policy interventions that produce effects opposite to those intended. Specifically, interventions may mistakenly be deemed successful when they are not, and vice versa, and inadvertently reward regimes that have increased total repression and penalize those that have reduced it. The problem of identifying the true trend of repression over time has been the subject of recent academic debates (Cingranelli and Filippov 2018; Cope, Crabtree, and Fariss 2020; Fariss 2019). Fariss (2014) points out that time trends in repression measures may be misleading because the standards used by experts and reporters to classify actions as repression can change over time. In contrast, the problem we focus on concerns the regime's strategic decision to conceal repression, so that observed and unobserved repression can move in opposite directions as the environment evolves. However, there is an overlap: If the ability of observers to detect repression improves more than the ability of the regime to conceal it, this is equivalent to an increase in concealment costs in our model. In that case, observed (revealed) repression may rise even as total repression falls, potentially leading to misguided and damaging policy interventions.

Our paper also contributes to the scholarship on the backlash and deterrence effects of repression. Many case studies (Martin 2007) and quantitative analyses (Francisco 2004) suggest that repression can provoke backlash protests. The empirical literature identifies instances of backlash across a range of conflicts, including Northern Ireland (LaFree, Dugan, and Korte 2009), Israel/Palestine (Dugan and Chenoweth 2012), Iran

(Rasler 1996), Vietnam (Dell and Querubin 2018; Kocher, Pepinsky, and Kalyvas 2011), and Turkey (Aytaç, Schiumerini, and Stokes 2018). At the same time, substantial evidence shows that repression often reduces mobilization (Davenport 2007; Downes 2008; Earl 2011; Kalyvas 2006; Tarrow 2011; Tilly 1978). Studies document deterrence effects even for indiscriminate repression, including in Guatemala (Stoll 1993), Iraq (Condra and Shapiro 2012), Chechnya (Lyall 2009), and Ukraine (Rozenas and Zhukov 2019). Davenport (2007) refers to these conflicting findings as the "punishment puzzle."

Scholars attribute the deterrence effect of repression primarily to its ability to increase the costs of dissent (Opp and Ruehl 1990; Tilly 1978). With few exceptions (Gibilisco 2021; Rozenas 2020; Shadmehr and Boleslavsky 2022), the formal literature has focused on the deterrence effect.[6] The broader literature offers a wide range of explanations for backlash, including the generation of moral and emotional outrage (Aytaç, Schiumerini, and Stokes 2018; González and Prem 2024; Lebas and Young 2024; Pearlman 2018; Rosenfeld 2025), radicalization of regime opponents (Dugan and Chenoweth 2012; Siqueira and Sandler 2006), and defensive opposition to protect against indiscriminate repression (Rozenas 2020). Importantly, a growing literature seeks to reconcile the punishment puzzle by conditioning the relative strength of deterrence and backlash effects on various mediating factors. Scholars have examined how the efficacy of repression varies with identity factors (Lawrence 2017), rebel tactics (Chenoweth 2023; Chenoweth and Stephan 2011), reliance on or integration with the broader population (Opp and Ruehl 1990; Toft and Zhukov 2015), psychological versus material motivations (Bueno De Mesquita and Shadmehr 2023), and regime type and government institutions (Aytaç and Stokes 2019; Boleslavsky, Ginsburg, and Shadmehr 2019; Ritter 2014).

[6] In the formal literature, repression raises the costs of anti-regime actions (Boix and Svolik 2013; Bueno de Mesquita 2010; Casper and Tyson 2014; Chau, Hassan, and Little 2025; Chen and Suen 2017, 2021; Egorov and Sonin 2021, 2024; Guriev and Treisman 2020; Little, Tucker, and LaGatta 2015; Persson and Tabellini 2009; Shadmehr and Bernhardt 2011), or directly reduces the probability of revolt (Acemoglu and Robinson 2006; Besley and Persson 2011; Boix 2003).

We show that concealed repression poses fundamental challenges for identifying repression's deterrence and backlash effects. Using our model, we then develop an informational theory of these effects, characterizing the conditions under which each arises and intensifies. An important implication is that the common approach—comparing protest probabilities in the presence and absence of observed repression—tends to overstate backlash and thus understate the efficacy of repression. We build on the literature's insight that (i) the public views some alternatives to the status quo as damaging, and therefore regards certain instances of state coercion as legitimate; and (ii) the public remains uncertain about the activists' motivations and the state's objectives, and thus about the legitimacy of specific instances of coercion—see fn. 3. As a result, whether repression hinders or promotes further protest depends on whether the public perceives the alternative sought by the repressed activists as worse or sufficiently better than the status quo. Early social movement scholars emphasized variations of this insight under the rubric of the legitimacy or illegitimacy of repression (Opp and Ruehl 1990). Importantly, observing repression is informative to the public about the beneficial or harmful nature of the activists' demands. We extend this insight by highlighting that repression is also a signal of the presence of dissent that is threatening enough to provoke repression.[7] The presence of organized activists, in turn, raises the public's perceived likelihood of overturning the status quo should they find change desirable. Anticipating that learning about organized dissent may encourage protest, the regime may strategically choose to conceal repression.

Methodologically, our model falls into an unexplored class of disclosure games (Hart, Kremer, and Perry 2017; Verrecchia 2001) with private non-disclosure costs and a type-dependent outside option for the sender. We discuss this methodological contribution after the formal presentation of the model.

[7]For example, Pan and Siegel (2020) find that Saudi repression of online dissidents led to "a very large increase in the popularity of Saudi Google searches for imprisoned opinion leaders", reflecting the "Backlash Effects on the Saudi Public" (p. 118).

2. Model

We begin by outlining the model informally before turning to its formal exposition. Consider a society consisting of a regime, the general public, and activists. To maintain the status quo, the regime needs some support from the general public. If activists challenge the regime and the public supports them, the status quo is overturned. Some activists pursue policies that, if implemented, will benefit the public (e.g., Solidarity in Poland); we refer to them as "good" activists. Others pursue policies that will harm the public (e.g., Mojahedin in Iran); we refer to them as "bad" activists. There is a conflict of interest between the regime and the public: It prefers to maintain the status quo over any change.

With probability γ, activists have a window of opportunity to organize and exert pressure on the regime for policy change (McAdam 1999; Tarrow 2011; Tilly and Tarrow 2007). With the complementary probability $1-\gamma$ there is no window of opportunity or they cannot organize, so there is no initial pressure or protest. With probability γq, the organized activists are the type to demand good policies for the public and, with the remaining probability $\gamma(1-q)$, they are the type to demand harmful ones. If activists organize, the regime can either concede to their demands or repress them, and either conceal or reveal that repression.

The public then decides whether to protest against the regime. The protest succeeds if and only if some activists have organized but the regime has not conceded to their demands. If the public's protest succeeds, the activists' demands are implemented, and the regime pays a cost. Importantly, the public does not know whether there is a window of opportunity and activists have made an organized challenge nor whether they seek beneficial or harmful policies (i.e., whether activists are good or bad) when making its protest choice (Shadmehr and Bernhardt 2011; Shadmehr and Boleslavsky 2022). When the public observes concessions, it knows that protest will be futile because the activists

are satisfied or co-opted, and hence the public does not protest. When the public observes repression, it knows that its protest will succeed if it is undertaken, but it must assess whether it would want to side with potentially bad activists against the regime. By contrast, when the public does not observe either concession or repression, it remains uncertain whether there is a window of opportunity and the activists organized or the regime repressed organized activists and concealed that repression. These uncertainties and trade-offs are at the core of our analysis.

Formally, there is a regime, a public, and a nonstrategic activist. There are three types of activists, denoted by $\theta \in \Theta := \{G, B, N\}$, representing good ($G$), bad ($B$) and unorganized ($N$). We can think of θ as part of the state of the world that reflects whether activists are good or bad for the public and whether they are organized or not. The players share a common prior $\mu_0(G, B, N) := (\gamma q, \gamma(1-q), 1-\gamma)$. The game proceeds in three stages.

In stage 1, the regime and the public each learn their respective private information. The regime privately observes θ, drawn from μ_0, and the cost of concealing repression c, drawn independently from a distribution H. The public privately observes its protest cost ρ, drawn from a distribution G, independently from θ and c.[8] We assume that H is twice continuously differentiable with full support on $[\underline{c}, \overline{c}]$, where $0 \leq \underline{c} < \overline{c}$, and that G is also twice continuously differentiable, with full support on $[\underline{\rho}, \overline{\rho}]$, where $0 \leq \underline{\rho} < \overline{\rho}$.

In stage 2, the regime chooses its action. If there is an organized activist (that is, if $\theta \in \{G, B\}$), the regime decides whether to concede to the activist, to repress-and-reveal (public repression), or to repress-and-conceal, and we move to stage 3. If there is no organized activist (i.e., $\theta = N$), the regime has no decision and the players move to stage 3. We can think of $(\theta, c) \in \Theta \times [\underline{c}, \overline{c}]$ as the regime's two-dimensional type.

[8]Because protest costs are private information, the results are identical if ρ is drawn and the public is informed at the beginning of stage 3, after the regime's decision.

The term repress-and-conceal refers to the combination of carrying out repression and concealing it from the public. It takes two broad forms. First, some repression tactics are inherently less observable to the public: These include stealth torture (Rejali 2007), covert surveillance, the weakening of activist organizations (Nalepa and Pop-Eleches 2022; Sullivan 2016), subtle intimidation, and the use of fabricated non-political charges (Pan, Xu, and Xu 2025). As Li (2018) notes, "a strong regime, possessing an effective and loyal internal security sector... can usually prevent the occurrence of transgressive protests in the first place and thus can avoid shooting protesters on the streets" (p. 25). Second, regimes actively suppress information about repression through media control. This includes co-opting or threatening journalists (Mazzaro 2023; Reporters Without Borders 2023), flooding social media with misinformation or irrelevant content to drown out news of repression, establishing state-controlled media, and controlling access to the internet and digital platforms (Roberts 2018). The costs of concealment reflect the opportunity costs of the expertise, coordination, resources, and infrastructure required to hide repression. These costs underscore the high levels of sophistication, organization, and technology needed to carry out concealed repression (Beraja et al. 2023; Greitens 2016; Guriev and Treisman 2022). In fact, economic windfalls to regimes have been shown to increase the prevalence of less observable forms of repression (Andirin et al. 2022).

In stage 3, the public chooses its action. The public observes concession if it occurred, and it observes repression if and only if the regime chose to repress-and-reveal. Otherwise, the public observes no news, so it cannot tell whether no activist has challenged the regime, or whether the regime has chosen to repress-and-conceal. The public then decides whether to protest. The protest succeeds if and only if there is an organized activist (that is, $\theta \in \{G, B\}$) and the regime has not conceded. If the protest succeeds, the activist's demands are implemented, and the regime pays a cost. The payoffs are realized, and the game ends.

The regime's payoff following a successful protest by the public is normalized to 0, and its payoff from the status quo is normalized to 1. Its payoff from conceding to an activist (and avoiding a successful protest) is $1-\alpha_G$ when the activist is good, and $1-\alpha_B$ when the activist is bad, with $\alpha_G, \alpha_B \in (0,1)$. The regime pays the concealment cost c if it decides to repress-and-conceal. The public's payoff from the status quo is normalized to 0. Its payoff from the implementation of good and bad policies are β_G and β_B, respectively, with $\beta_G > \max\{\beta_B, 0\}$. The public pays the cost ρ if it protests.

We focus on perfect Bayesian equilibria with the D1 refinement specified in the appendix.[9] The regime's strategy is a mapping

$$\sigma : \{G, B\} \times [\underline{c}, \overline{c}] \to \Delta\{\text{concede}, \text{repress-and-reveal}, \text{repress-and-conceal}\}$$

from the regime's type (θ, c) to a probability distribution over the regime's actions. The public's strategy is a mapping $\phi_\omega : [\underline{\rho}, \overline{\rho}] \to [0,1]$ from its protest costs ρ to a probability of protesting upon observing $\omega \in \{R, NN\}$, where R indicates revealed repression and NN means observing neither repression nor concession, i.e., no news.

Methodological Contribution Before proceeding with the analysis, we discuss the paper's methodological contribution. Without the concealment option, our game is a standard signaling game (see Part 2 of Proposition 2 and Corollary 1). Shadmehr and Boleslavsky (2022) and Boleslavsky, Ginsburg, and Shadmehr (2019) study the role of social norms and different forms of commitment (exogenous international pressure or endogenous institutions) in variations of this game, analyzing public beliefs, the spread of protest, and welfare. Our model builds on this signaling framework by enabling the regime to conceal repression. This extension is both substantively and methodologically

[9] D1 refinement rules out perfect Bayesian equilibria where repress-and-reveal is off the equilibrium path—see Remark 2 in the appendix.

significant. Substantively, it allows us to study revealed and concealed repression and their consequences. Methodologically, as we will describe, it transforms the model into a disclosure game that remains unexplored in the literature.

Absent concession and concealment costs, our model is one of disclosure of verifiable information; a special case of Hart, Kremer, and Perry (2017) and Gieczewski and Titova (2024).[10] We use the insight of Dye (1985) that no news may reflect that the sender is potentially uninformed. Without concession but with concealment costs, the regime (sender) conceals if the cost is below a threshold and does not conceal otherwise. The public then learns whether there is an organized activist but does not learn whether the activist is good or bad. Our model is different from censorship models (Besley and Prat 2006; Dye 1985; Shadmehr and Bernhardt 2015) in which the sender chooses between fully revealing the state of the world and revealing nothing. Here, the sender chooses between revealing that there is an activist (thus pooling states G and B) and revealing nothing.

Our model with concession and concealment costs corresponds to a verifiable disclosure model with both nondisclosure costs and, importantly, a type-dependent outside option (concession). This class of models has remained unexplored in the literature. As we will show, the public (receiver) learns about the nature of the activists' demands (good/bad) upon observing repression. Upon observing no news, the public (receiver) always updates its belief about whether the activists are organized; and updates its belief about the nature of activists' demands (good/bad) if and only if the regime (sender) and the public (receiver) have a large conflict of interest ($\alpha_G > \alpha_B$).

[10] Using the notation of Hart, Kremer, and Perry (2017), $T = \{G, B, N\}$, $E = \{$repress-and-conceal, repress-and-reveal$\}$, $E_G = E_B = E$, $E_N = \{$repress-and-conceal$\}$, $L_G = L_B = T$, and $L_N = \{N\}$. That is, types B and G can pretend to be any type, but type N can pretend to be itself only. Using the notation of Gieczewski and Titova (2024), $M(B) = M(G) = \{$repress-and-conceal, repress-and-reveal$\}$ and $M(N) = \{$repress-and-conceal$\}$.

3. Revealed and Concealed Repression and Their Measurement

We maintain the following assumption to focus on settings in which, in equilibrium, the regime engages in all three actions (repress-and-conceal, repress-and-reveal, and concede) and observing repression does not fully reveal whether activists are good or bad.

Assumption 1. *Let $\beta^e := q\beta_G + (1-q)\beta_B$ be the public's expected policy payoff from successful revolt under the prior. Then,*

1. $\alpha_G < \alpha_B$
2. $\alpha_G < G(\beta^e)$
3. $\alpha_G < \overline{c}$
4. $\alpha_G > \underline{c}$
5. $\alpha_G > G(\beta_B)$
6. $\underline{\rho} < \left(1 + \frac{1-\gamma}{\gamma}\frac{1}{H(\alpha_G)}\right)^{-1} \beta^e$

The first part, $\alpha_G < \alpha_B$, is substantive, reflecting that the regime dislikes conceding to organized bad activists more than conceding to organized good activists, i.e., the conflict of interest between the regime and the public is mild.[11] Consequently, the public remains uncertain about the activist's type, whether $\theta = G$ or B, in equilibrium after observing repression—see part 2 of Proposition 2. Parts 2 to 5 simplify the exposition by ensuring that the regime engages with a strictly positive probability in all three choices of repress-and-conceal, repress-and-reveal, and concede, in equilibrium. Part 6 ensures that the equilibrium probability of protest absent revealed repression is not 0, so that

[11] In the next section, we will analyze the setting with $\alpha_G > \alpha_B$, in which observing repression reveals the activist's type—see Proposition 5.

the model's implications remain consistent with empirical patterns. Part 6 is satisfied automatically if $\underline{\rho} = 0$ and Part 5 is satisfied if $\beta_B \leq 0$.

We begin analysis with the public's decision. Let $\mu = (\mu(G), \mu(B), \mu(N)) \in \Delta(\Theta)$ be the public's belief. We denote by μ_R the public's posterior belief upon observing repression and by μ_{NN} its posterior belief when the public observes no news. Given a belief μ, the public protests if and only if its (posterior) expected payoff from protest exceeds its costs: $r(\mu, \rho) := \mu(G)\beta_G + \mu(B)\beta_B - \rho > 0$. It follows that the public protests whenever direct protest costs are below a threshold that depends on the public's belief about the activist.

LEMMA 1. *In any equilibrium, the public protests if and only if $\rho \leq \widetilde{\rho}(\mu)$, where $\widetilde{\rho}(\mu) = \mu(G)\beta_G + \mu(B)\beta_B$ and $\mu \in \Delta(\Theta)$ is the public's belief.*

It is also intuitive that the regime chooses to repress-and-conceal an activist of type $\theta \in \{G, B\}$ whenever the concealment cost is below some threshold $\widetilde{c}_\theta$. Remarkably, Proposition 1 shows that these thresholds are the same for good and bad activists in any equilibrium: $\widetilde{c}_G = \widetilde{c}_B = \widetilde{c}$. That is, if the regime plans to conceal repression, it represses and conceals both types of activists equally.

PROPOSITION 1. *Suppose Assumption 1 holds. Then, there exists a concealment cost threshold $\widetilde{c} \leq \alpha_G$ such that, in every equilibrium, the regime strictly prefers to repress-and-conceal if $c < \widetilde{c}$, and repress-and-reveal or concede otherwise. Consequently, when the public does not observe any news (of concession or repression), it believes that the presence of an organized opposition is less likely, but it does not update its beliefs about whether the activist is good or bad: no news is good news for the regime. Formally, $\mu_{NN} = (\gamma' q, \gamma'(1-q), 1-\gamma')$ where $\gamma' := \frac{H(\widetilde{c})\gamma}{H(\widetilde{c})\gamma + 1 - \gamma} < \gamma$.*

The intuition hinges on the assumption that it is less costly for the regime to concede to good activists than to bad ones (Assumption 1.1). Moreover, upon observing a regime action, the public reacts the same way regardless of the true activist type ($\theta = G$ or B)

because the public does not observe the true activist type. Together, these imply that the regime engages in repress-and-conceal of both types of activists equally. Suppose it did not. The regime's alternatives to repress-and-conceal are better when the activists are good. It follows that (1) if the regime engages in repress-and-conceal of some type of activist less frequently, it will be the good activists, and (2) the regime concedes to the activists any time the regime does not repress-and-conceal. This, in turn, would imply that the regime will never repress-and-reveal when faced with good activists. But if the regime never observably represses the good activists, the public will never protest upon repress-and-reveal. Then, the regime will repress-and-reveal the good activists instead of conceding to them. Therefore, the regime chooses repress-and-conceal of all activist types equally, and it prefers (observably) repressing the good activists at least as much as conceding to them.

An implication of Proposition 1 is that the public does not learn anything about the nature of the activists' demands following no news: The posterior likelihood ratio that the activists are good versus bad remains the same as the prior. When the regime plans to conceal repression, it always represses all activists, good and bad, so the public does not learn about the activist's type following no news. To further clarify the intuition, let us decompose the decision of a type-(θ, c) regime into two steps. In Step 1, the regime decides whether to repress or concede. In Step 2, if it has repressed, it decides whether to reveal or conceal that repression. The Step 2 decision compares the payoff from concealing repression, $1 - G(\widetilde{\rho}(\mu_{NN})) - c$, and revealing repression, $1 - G(\widetilde{\rho}(\mu_R))$, so that the regime conceals repression whenever

$$c < \widetilde{c} := G(\widetilde{\rho}(\mu_R)) - G(\widetilde{\rho}(\mu_{NN})). \tag{1}$$

Anticipating this in Step 1, the regime knows that it will not conceal when $c > \widetilde{c}$, so the choice would be between conceding and getting $1 - \alpha_\theta$, and revealing repression and

getting $1 - G(\widetilde{\rho}(\mu_R))$. Under Assumption 1, the regime will always repress bad activists in Step 1.[12] When facing good activists, the regime is indifferent between repress and concede:

$$1 - G(\widetilde{\rho}(\mu_R)) = 1 - \alpha_G. \tag{2}$$

When, instead, $c < \widetilde{c}$, the same logic applies, but because concealment is costly, the indifference breaks in favor of repression: $1 - G(\widetilde{\rho}(\mu_{NN})) - c > 1 - G(\widetilde{\rho}(\mu_R)) = 1 - \alpha_G$, so $1 - G(\widetilde{\rho}(\mu_{NN})) > 1 - \alpha_G$. That is, when the regime plans to conceal, it always represses good activists. Because the regime always represses all organized activists when it plans to conceal repression, the absence of repression is not informative about whether the activists are good or bad.

However, the absence of news of concession or repression can be informative about the presence of organized activists. Because activists only sometimes organize to challenge the regime ($\gamma < 1$) and because the regime sometimes conceals repression $\widetilde{c} < \overline{c}$,[13] following no news, the public updates its belief that it is less likely that activists have organized, $\mu(N) < \mu_0(N)$:

$$\Pr(\theta \in \{G, B\} \mid \text{no news}) = \frac{\gamma\ H(\widetilde{c})}{\gamma\ H(\widetilde{c}) + 1 - \gamma} < \gamma. \tag{3}$$

Proposition 2 summarizes these arguments and details their consequences for equilibrium behavior and the probability of repression.

---

[12] Otherwise, i.e., if the regime only sometimes concedes to bad activists, then it will always concede to good activists (since $\alpha_G < \alpha_B$). But then, the public learns that the activists are bad when observing repression, so the regime's payoff from repression is $1 - G(\beta_B)$, which exceeds the payoff from conceding to good activists (since $\alpha_G > G(\beta_B)$), a contradiction.

[13] By Proposition 1, $\widetilde{c} \leq \alpha_G$, and by Assumption 1.3, $\alpha_G < \overline{c}$.

PROPOSITION 2. *In every equilibrium:*

1. *If $c < \widetilde{c}$, the regime represses all organized activists and conceals its repression, where $\widetilde{c} \in (\underline{c}, \alpha_G)$ is the unique solution to*

$$\frac{\gamma H(\widetilde{c})}{\gamma H(\widetilde{c}) + 1 - \gamma}\ \beta^e = G^{-1}(\alpha_G - \widetilde{c}). \tag{4}$$

2. *If $c > \widetilde{c}$, the regime publicly represses organized bad activists and represses organized good activists with probability less than 1. In particular, the likelihood ratio of publicly repressing good versus bad organized activists is*

$$\kappa := \frac{\Pr(\textit{repress-and-reveal} \mid G)}{\Pr(\textit{repress-and-reveal} \mid B)} = \frac{1-q}{q}\frac{G^{-1}(\alpha_G) - \beta_B}{\beta_G - G^{-1}(\alpha_G)} \in (0, 1).$$

To understand the role of concealment, it is useful to consider two extreme cases: when the costs of concealment are prohibitively high, or negligible. The next result follows from Equation (4) in Proposition 2.

COROLLARY 1. *If $H(x)$ converges pointwise almost everywhere to $\mathbf{1}_{\{x \geq 1\}}$, so that concealment costs are almost always prohibitively high, then $H(\tilde{c}) \to 0$.*

If $H(x)$ converges pointwise almost everywhere to $\mathbf{1}_{\{x \geq 0\}}$, so that the concealment costs are almost always negligible, then

$$H(\tilde{c}) \to \begin{cases} 1 & ;\ G(\gamma\beta^e) < \alpha_G; \\ \frac{1-\gamma}{\gamma}\frac{G^{-1}(\alpha_G)}{\beta^e - G^{-1}(\alpha_G)} & ;\ G(\gamma\beta^e) \geq \alpha_G. \end{cases}$$

Intuitively, when the concealment costs are prohibitively high, the regime almost never chooses to repress-and-conceal. When costs are negligible, the likelihood of con-

cealment increases in the concession cost α_G; once the payoff from conceding to the good activists $(1-\alpha_G)$ falls below the payoff from repressing-and-concealing *all* activists $(1-G(\gamma\beta^e))$, the regime always represses-and-conceals.

To better understand the role of concession, let us simplify the model and remove that option. Without the option to concede, revealed repression and absence of news convey no information about whether the activists are good or bad. As in the full model, the regime represses and conceals whenever (1) holds: $c < \widetilde{c} := G(\widetilde{\rho}(\mu_R)) - G(\widetilde{\rho}(\mu_{NN}))$. Otherwise, it represses and reveals the repression. The threshold $c = \widetilde{c}$ is the unique solution to the regime's indifference condition: $\widetilde{c} = G(\beta^e) - G(\gamma'(\widetilde{c})\ \beta^e)$, where γ' is given in Proposition 1. Intuitively, when the regime pays the concealment cost, it pays to reduce the public's posterior that there is an organized activist. Remark 1 summarizes.

REMARK 1. *Suppose the regime does not have the option to concede. There is a unique threshold* $\widetilde{c} \in (\underline{c}, G(\beta^e))$*, such that the regime represses all organized activists and conceals it, if* $c < \widetilde{c}$*; otherwise it represses all organized activists and reveals its repression.*

Proposition 2 allows us to characterize the ex ante probabilities of revealed and concealed repression conditional on organized activists demanding a policy change. Recall that $\sigma(\text{repress-and-reveal} \mid \theta, c)$, the probability of repress-and-reveal by a type-(θ, c) regime, depends on the realization of concealment costs c. Therefore, the ex ante probability of the revealed repression of a type-θ opposition is

$$\Pr\left(\begin{smallmatrix}\text{revealed}\\ \text{repression}\end{smallmatrix} \mid \theta\right) = \int_{\underline{c}}^{\bar{c}} \sigma(\text{repress-and-reveal} \mid \theta, c) dH(c) = \begin{cases} 1 - H(\widetilde{c}), \text{ if } \theta = B, \\ \kappa(1 - H(\widetilde{c})), \text{ if } \theta = G. \end{cases}$$

From Proposition 2, the equilibrium probability of revealed repression (conditional on

organized activists demanding a policy change) is

$$\begin{aligned}\Pr\left(\text{revealed repression}\right) &= q\Pr(\text{repress-and-reveal} \mid G) + (1-q)\Pr(\text{repress-and-reveal} \mid B)\\ &= (\kappa q + 1 - q)(1 - H(\widetilde{c}))\\ &= \frac{\beta_G - \beta_B}{\beta_G - G^{-1}(\alpha_G)}(1-q)(1-H(\widetilde{c})),\end{aligned} \tag{5}$$

while the corresponding equilibrium probability of concealed repression (repress-and-conceal) is

$$\Pr\left(\text{concealed repression}\right) = H(\widetilde{c}). \tag{6}$$

Finally, from (5) and (6), the probability of total repression is

$$\begin{aligned}\Pr(\text{total repression}) &= \Pr(\text{revealed repression}) + \Pr(\text{concealed repression})\\ &= 1 - q(1-\kappa)(1-H(\widetilde{c}))\\ &= 1 - \frac{\beta^e - G^{-1}(\alpha_G)}{\beta_G - G^{-1}(\alpha_G)}(1-H(\widetilde{c})).\end{aligned} \tag{7}$$

Naturally, $\Pr(\text{concession}) = 1 - \Pr(\text{total repression})$. A formal literature studies how regimes combine repression with information manipulation when the information is unrelated to repression (Egorov and Sonin 2021, 2024; Gehlbach et al. 2025; Gitmez, Molavi, and Sonin 2025; Guriev and Treisman 2020). Our focus is instead on the regime's censorship of information about repression itself. Interpreting $H(\tilde{c})$ as the probability of censorship (of repression), (7) implies that repression and its censorship are complements.

3.1. Measuring Repression

Revealed acts of repression are observable, so their probability can be empirically estimated. In contrast, concealed acts of repression are not observable. This creates a *measurement problem*: if researchers or policymakers focus only on observable repression,

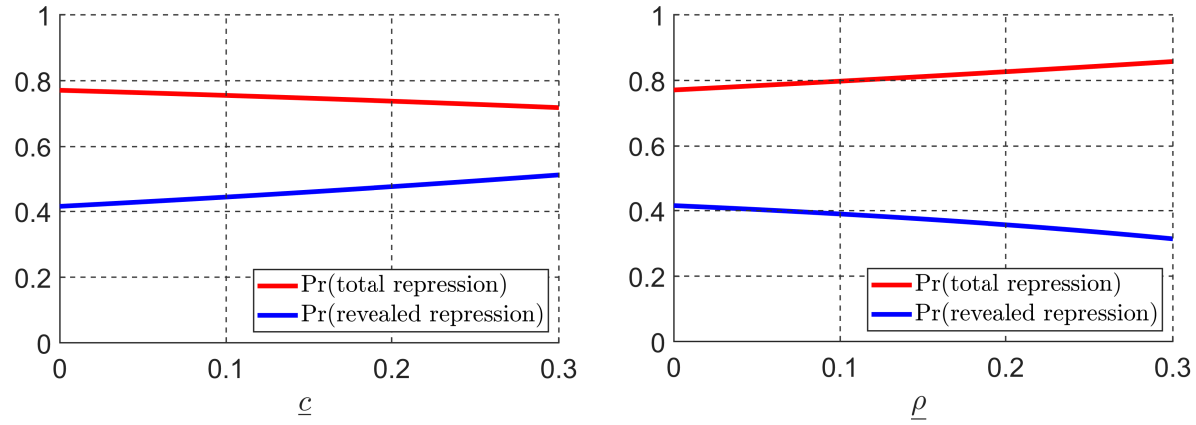

Figure 1. *Negative correlation between total and revealed repression. Basing analysis or policy on the observable revealed repression can lead to perverse outcomes. Left Panel:* $G = U[0,1]$ *and* $H = U[\underline{c},1]$*; Right Panel:* $G = U[\underline{\rho},1]$ *and* $H = U[0,1]$*. Parameters:* $q = 0.65$*,* $\gamma = 0.4$ $\beta_G = 2.5$*,* $\beta_B = -1$*,* $\alpha_G = 0.6$*, and* $\alpha_B > 0.6$*.*

they risk drawing misleading inferences about the overall prevalence of repression. To illustrate this, we show in Proposition 3 that reductions in concealment costs, such as efficiencies due to technological progress, increase unobservable—and even total—repression while reducing observable revealed repression. As a result, observers may mistakenly infer that the regime is becoming more tolerant.

PROPOSITION 3. *Let* $H_1(c)$ *and* $H_2(c)$ *be two distributions of concealment costs such that* $H_1(c) < H_2(c)$ *for all* c *in the interior support of the distributions (so that* $H_1 \succ_{FOSD} H_2$*). Then,* $\Pr_1(\textit{revealed repression}) > \Pr_2(\textit{revealed repression})$*, but* $\Pr_1(\textit{total repression}) < \Pr_2(\textit{total repression})$*, where* $\Pr_i(A)$ *is the probability of event* A *when* $c \sim H_i$*.*

Conversely, when concealment costs increase, total repression can fall while revealed repression simultaneously increases, making it *appear* to observers that the regime is becoming less tolerant. These patterns are not special to concealment costs. As Figure 1 illustrates, similar negative correlations and opposite trends arise in response to variations in protest costs.

To circumvent this measurement problem, we leverage the relationships between observable and unobservable equilibrium objects of interest, which allows us to write down total (i.e., observable and unobservable) repression as a function of observable repression. Let $q' = \Pr(G \mid \text{revealed repression})$ denote the public's belief, upon observing repression, that the activists are good. Since the regime observably represses bad organized activists more often (Proposition 2, Part 2), the public updates negatively about the type of organized activists that are observably repressed, so $q' < q$. By Bayes Rule, the public's posterior belief about the activists is given by $\frac{q'}{1-q'} = \frac{q}{1-q}\kappa$. Combining this with (5), we have

$$\Pr(\text{revealed repression}) = \frac{1-q}{1-q'}(1 - H(\widetilde{c})). \tag{8}$$

We can interpret q and q' as *measures of public opinion* about the activists, the opposition, or more broadly, the "vanguards" or "early risers" (all of whom we have collectively referred to as the activist). Specifically, they reflect the public's belief that the activists' demands would constitute an improvement over the status quo if implemented. Measures of public opinion q and q' can be estimated from population surveys that measure the public's belief about the activists after revealed repression (for q') and absent revealed repression (for q).[14] Pr(revealed repression) can be estimated from the frequency of the observable acts of revealed repression. Let $\widehat{q}$, $\widehat{q'}$, and $\widehat{p}$ be the corresponding consistent estimates. Similarly, let $\widehat{H}$ be a consistent estimate of $H(\widetilde{c})$, which is not directly observed. From (8), we can estimate $H(\widetilde{c})$: $\widehat{H} = 1 - \frac{1-\widehat{q'}}{1-\widehat{q}}\ \widehat{p}$.

Combining these results, Proposition 4 provides a method to estimate the probability

[14] In addition to institutionalized surveys such as AmericasBarometer, AfroBarometer, ArabBarometer, AsianBarometer, World Values Survey, and country-specific surveys such as GAMAAN, researchers routinely conduct surveys even in authoritarian regimes (Bursztyn, González, and Yanagizawa-Drott 2020; Cantoni et al. 2019; Huang 2015, 2018; Lebas and Young 2024; Pop-Eleches, Robertson, and Rosenfeld 2022; Rosenfeld 2017, 2025; Tertytchnaya 2023; Tertytchnaya and Lankina 2020). To estimate q and q', one could ask questions like the following at regular intervals, capturing public opinion before and after observed repression: "Suppose the current opposition movement were to gain political power. On a scale from 0 to 100, how likely is it that your personal condition would improve as a result?"

of total repression—covering both observable (revealed) and unobservable (concealed) forms—in terms of primitive and equilibrium objects that can be directly measured or estimated (i.e., $\widehat{q}$, $\widehat{q'}$, and $\widehat{p}$).

PROPOSITION 4. *In every equilibrium,*

$$\Pr(\textit{total repression}) = 1 - \frac{q - q'}{1 - q} \Pr(\textit{revealed repression}),$$

where $q > q'$. *Consequently,* $1 - \frac{\widehat{q} - \widehat{q'}}{1 - \widehat{q}} \widehat{p}$ *is a consistent estimate for the probability of total repression.*

Because preventive repression tends to be "less visible, more deniable" (Andirin et al. 2022, p. 28), our analysis of concealed and revealed repression is closely related to the literature on preventive and reactive repression (Carey et al. 2022; Danneman and Ritter 2014; Esberg 2021; Ritter and Conrad 2016). Although regimes may sometimes conceal reactive repression and not all preventive repression is concealed, preventive repression that is deliberately concealed falls within our definition of concealed repression. Proposition 4 then provides a method for accounting for such preventive repression, in conjunction with other forms of concealed repression—using public opinion data and observed repression.

Focusing on preventive repression, Andirin et al. (2022) argue that more repressive regimes tolerate a lower probability of protest, which in turn makes protest less predictable. They leverage this robust insight to construct a high-frequency global dataset of preventive repression, based on the distribution of predicted protest probabilities that they estimate. Our model, by contrast, addresses concealed repression more broadly, and our proposed solution to recover concealed repression relies on public inference rather than predicted protest probabilities.

4. Deterrence and Backlash

Concealed repression complicates the identification of repression's deterrence and backlash effects. The public can protest after observing (revealed) repression or without observing any repression. The probability of protest after revealed repression is $G(\widetilde{\rho}(\mu_R))$. The probability of protest without observing repression is $G(\widetilde{\rho}(\mu_{NN}))$. If the protest is observable, these probabilities can be estimated from the protest frequencies. Moreover, under Assumption 1, $G(\widetilde{\rho}(\mu_{NN})) < G(\widetilde{\rho}(\mu_R))$ in equilibrium.[15] Thus, the estimated probability of protest is higher following revealed repression than when no repression is observed.

However, this empirical pattern does not necessarily imply that repression causally increases protest, i.e., that there is a backlash effect. The core difficulty lies in defining the appropriate counterfactual. In settings where regimes sometimes conceal repression, the absence of observed repression does not imply that repression did not occur. Therefore, comparing protest probabilities in the presence and absence of *observed* repression does not identify the causal effect of repression. A more meaningful counterfactual for assessing the causal effect of repression is the probability of protest under the prior, that is, before the public updates its beliefs about the occurrence of concealed repression. We now formally define deterrence and backlash effects.

Definition 1. *Let $D := G(\widetilde{\rho}(\mu_0)) - G(\widetilde{\rho}(\mu_R))$. We say that repression has a*

- *deterrence effect if the public is less likely to protest after observing repression, $D > 0$, in which case D is the deterrence effect, and a*
- *backlash effect if the public is more likely to protest after observing repression, $D < 0$, in which case $-D$ is the backlash effect.*

[15] From Equation (1), $G(\widetilde{\rho}(\mu_R)) - G(\widetilde{\rho}(\mu_{NN})) = \widetilde{c}$, and from Proposition 2, $\widetilde{c} > \underline{c} \geq 0$.

The deterrence and backlash effects of repression hinge on the public's inferences after observing repression. These inferences, in turn, depend on the conflict of interest between the regime and the public. We thus extend our model to settings with severe conflict of interest, in which the regime prefers to concede to organized activists whose demands harm the public over the activists with beneficial policy demands.

Assumption 2. *Suppose that:*

1. $\alpha_B < \alpha_G$
2. $\alpha_B < G(\beta^e)$
3. $\alpha_G > G(\beta_G)$
4. $\alpha_G < \overline{c}$
5. $G(\beta_G) > \underline{c}$

The first part, $\alpha_B < \alpha_G$, is substantive, and reflects that the regime dislikes conceding to organized good activists more than conceding to organized bad activists. The remaining assumptions simplify equilibrium exposition. Parts 2 to 5 simplify the exposition by ensuring that the regime engages with a strictly positive probability in all three choices, in equilibrium. Proposition 5 characterizes the equilibrium behavior in settings with severe conflict of interest.

Proposition 5. *Under Assumption 2, an equilibrium exists. In any equilibrium, there are two thresholds $\widetilde{c}_B < \widetilde{c}_G$ such that:*

- *If $c < \widetilde{c}_B$ ($c < \widetilde{c}_G$), the regime represses organized bad (good) activists and conceals its repression, where $\widetilde{c}_B \in [\underline{c}, \alpha_G]$ and $\widetilde{c}_G \in (\underline{c}, G(\beta_G)]$.*
- *If $c > \widetilde{c}_B$, the regime concedes to organized bad activists.*
- *If $c > \widetilde{c}_G$, the regime represses organized good activists and reveals its repression.*

Proposition 5 implies that the regime chooses to repress-and-conceal differentially

across good and bad activists: Because the regime dislikes bad activists less, it is more willing to concede to them than to good activists and less willing to incur costs to conceal their repression. As a result, the absence of news becomes informative not only about the presence of organized activists, but also about their type. The difference stems from the severe conflict of interest. When the regime dislikes bad activists less, it can always concede to them without having an incentive to deviate to revealed repression, since the public protests upon observing repression. This sustains differential concealment of repression across activist types, in contrast to the setting of Proposition 1, where the regime dislikes bad activists more.

Combining Propositions 2 and 5 (mild and severe conflict of interest), we characterize the conditions under which revealed repression causes deterrence or backlash effects.

PROPOSITION 6. *Revealed repression has either a deterrence or a backlash effect except on a measure-zero subset of parameters. Moreover,*

- *Under the mild conflict-of-interest conditions of Assumption 1, revealed repression has a deterrence effect if* $G(\gamma\beta^e) > \alpha_G$*, but has a backlash effect if* $G(\gamma\beta^e) < \alpha_G$*. In particular, if* γ *is sufficiently large, then there is a deterrence effect.*
- *Under the severe conflict-of-interest conditions of Assumption 2, there is always a backlash effect.*

First, consider settings with mild conflict of interest (Assumption 1). When the public observes repression, it infers that there are organized activists and updates negatively about their type, since the regime publicly represses the bad activists more often because conceding to them is more costly. If the regime's cost of conceding to good activists is relatively low, the regime has less incentive to repress them. In this case, when repression is observed, the public updates its beliefs more negatively about the activists' type and is less likely to protest. Thus, a deterrence effect arises when the regime's concession costs

to good activists are low. Conversely, a backlash effect occurs when these costs are high, indicating a greater conflict of interest.[16]

This logic extends when the conflict of interest becomes severe (Assumption 2). Now, upon observing repression, the public learns that some activists have organized and that the organized activists are good. This raises the public's incentives to protest, creating a backlash effect.

These intuitions extend to the sizes of the deterrence and backlash effects $|D|$. Under Assumption 1, from Proposition 2, $D = G(\gamma\beta^e) - \alpha_G$, where we recall that $\beta^e = q\beta_G + (1-q)\beta_B$. Under Assumption 2, from Proposition 5, $D = G(\gamma\beta^e) - G(\beta_G)$.

COROLLARY 2. *The size of the deterrence (backlash) effect is increasing (decreasing) in q, β_B, and γ, and under Assumption 1, also in $G(\cdot)$.*

It may be difficult to provide direct estimates for the sizes of deterrence and backlash effects $|D|$. Instead, we provide bound estimates for $|D|$ under Assumption 1 in settings where revealed repression and protest are both observable. In particular, let $\widehat{p}_R$ be a consistent estimate of the probability of protest after observed repression, $G(\widetilde{\rho}(\mu_R))$, and let $\widehat{p}_{NN}$ be a consistent estimate of the probability of protest that does not follow observed repression, $G(\widetilde{\rho}(\mu_{NN}))$. These probabilities can be estimated from the corresponding frequencies of protest incidents. Although we cannot estimate $G(\widetilde{\rho}(\mu_{NN}))$, we can still provide a lower bound empirical estimate for D using the fact that $G(\widetilde{\rho}(\mu_{NN})) \leq G(\widetilde{\rho}(\mu_0))$ under Assumption 1.[17]

$$\widehat{p}_{NN} - \widehat{p}_R \approx G(\widetilde{\rho}(\mu_{NN})) - G(\widetilde{\rho}(\mu_R)) \leq G(\widetilde{\rho}(\mu_0)) - G(\widetilde{\rho}(\mu_R)) = D. \tag{9}$$

[16] A deterrence effect is also more likely when the public is already inclined to protest, such as when it believes activists are likely to be organized (higher γ) or good (higher q).

[17] The inequality follows because no news could be due to concealment by the regime or due to absence of organized activists.

The estimated bound becomes tighter as γ or $H(\widetilde{c})$ increase, so that $G(\widetilde{\rho}(\mu_{NN}))$ becomes closer to $G(\widetilde{\rho}(\mu_0))$. From (1) and Proposition 2, $G(\widetilde{\rho}(\mu_{NN})) - G(\widetilde{\rho}(\mu_R)) < 0$. This implies an upper bound for the size of a backlash effect.

PROPOSITION 7. *Let* $\underline{D} := G(\widetilde{\rho}(\mu_{NN})) - G(\widetilde{\rho}(\mu_R))$, *and* $\widehat{\underline{D}}$ *be its estimate. Under Assumption 1: (1)* $\underline{D} \leq D$*; (2)* $\underline{D} < 0$*; (3) The size of a backlash effect is at most* $|\underline{D}|$*. Moreover,* $\widehat{\underline{D}} = \widehat{p}_{NN} - \widehat{p}_R$.

Proposition 7 shows that comparing protest probabilities in the presence and absence of observed repression yields an upper bound on the backlash effect. As a result, such comparisons tend to overstate the magnitude of backlash and thus underestimate the efficacy of repression.

5. CONCLUSION

Regimes routinely conceal acts of repression, making it difficult to reliably measure repression, its temporal trends, and spatial variation. This measurement problem undermines both policy efforts to mitigate repression, and scholarly research on state-society interactions. To better understand the scope of the problem and identify potential solutions, in this paper, we develop a model in which the state chooses not only whether to repress, but also whether to conceal that repression. The public, in turn, makes inferences—both when repression is observed and when it is not—about two dimensions: (1) the desirability of the alternatives to the status quo advocated by activists, and (2) the presence of organized activists and the window of opportunity for mobilization.

We characterize the measurement problem that arises when the state censors its own acts of repression, showing how this distortion can lead to misguided policy interventions. Strategic concealment of repression renders it directly unobservable. Taking a structural approach, we leverage equilibrium relationships to propose a method of estimating con-

cealed repression based on observable variables, including public opinion data.

We then show how the possibility of concealed repression complicates the estimation of repression's deterrence and backlash effects, leading to potentially incorrect conclusions about its efficacy. In particular, a higher estimated probability of protest following revealed repression—compared to when no repression is observed—does not imply that repression causally increases protest, so that there is backlash. When regimes sometimes conceal repression, the absence of observed repression does not mean that no repression occurred. The core difficulty lies in defining the appropriate counterfactual. We develop an informational theory of deterrence and backlash, identifying the conditions under which each effect arises and intensifies. Using our theory, we show that estimating the backlash effect based on the difference in protest probabilities with and without observed repression tends to overstate backlash and thus understate the efficacy of repression.

Several directions for future research stand out. On the empirical front, our paper highlights the value of conducting periodic public opinion surveys on activists and opposition groups in dictatorships, showing how such data can help recover unobserved repression. Given that scholars already collect public opinion data on various aspects of authoritarian regimes (see fn. 14), doing so appears both feasible and promising. Research on the popularity and perceived legitimacy of activists and the opposition not only sheds light on the dynamics of state–citizen interactions, but also enables scholars and the international community to develop more accurate measures of repression that account for both revealed and concealed forms.

On the theoretical front, it would be worthwhile to analyze the more sophisticated strategies that regimes may adopt to manipulate public opinion. For example, regimes may commit to repression strategies through institutional means—such as establishing partially independent judiciaries or shaping the personnel composition of security forces. How do such institutional arrangements affect state-citizen interactions? Do they solely

benefit the regime, or can they be Pareto improving? In particular, if regimes are concerned that repression signals the presence of organized activists, they may engage in repression of members of the general public who pose no threat, thereby muddying the public's inference about activist presence when repression is observed. This suggests a novel rationale for indiscriminate repression. These directions are left for future research.

Appendix: Proofs

Proof of Lemma 1. The public's expected utility from protesting is $r(\mu, \rho) = \mu(G)\beta_G + \mu(B)\beta_B - \rho$. Thus, $r(\mu, \rho) \geq 0 \iff \mu(G)\beta_G + \mu(B)\beta_B - \rho \geq 0 \iff \rho \leq \widetilde{\rho}(\mu)$. ∎

Remark 2. *Under Assumption 1, in any PBE that survives D1, if revealed repression is off the equilibrium path, the public's posterior following unexpected revealed repression is $\mu_R = (0, 1, 0)$. Thus, revealed repression is on the equilibrium path in any PBE that survives D1.*

Proof. Suppose that revealed repression is off-path. Then the PBE payoff of the regime of type (θ, c) is $\max\{1 - G(\widetilde{\rho}(\mu_{NN})) - c, 1 - \alpha_\theta\}$, where $1 - \alpha_G > 1 - \alpha_B$ by Assumption 1.1. Therefore, by the D1 refinement, the lowest equilibrium payoff is achieved by the regime type $(B, \overline{c})$ and the public's belief following unexpected revealed repression is $\mu_R = (0, 1, 0)$. Then, the expected payoff from repress-and-reveal for the regime of any type is $1 - G(\beta_B)$. Consequently, deviating to repress-and-reveal is strictly profitable for the type-$(B, \overline{c})$ regime, since $1 - G(\beta_B) > 1 - \alpha_G > 1 - \alpha_B$ by assumptions 1.1 and 1.5. Hence, there is no PBE where revealed repression is off-path. ∎

Proof of Proposition 1. In any equilibrium, a type-(θ, c) regime strictly prefers to repress-and-conceal if

$$1 - G(\widetilde{\rho}(\mu_{NN})) - c > \max\{1 - G(\widetilde{\rho}(\mu_R)), 1 - \alpha_\theta\},$$

which is equivalent to

$$c < \widetilde{c}_\theta := 1 - G(\widetilde{\rho}(\mu_{NN})) - \max\{1 - G(\widetilde{\rho}(\mu_R)), 1 - \alpha_\theta\}. \tag{10}$$

The regime is indifferent if $c = \widetilde{c}_\theta$ (measure-zero event), and strictly prefers not to repress-and-conceal if $c > \widetilde{c}_\theta$. Since $0 < \alpha_G < \alpha_B$ (Assumption 1.1), equation (10) implies:

$$\widetilde{c}_G \leq \widetilde{c}_B \leq 1. \tag{11}$$

Two cases are possible: $\widetilde{c}_G = \widetilde{c}_B = \widetilde{c}$ and $\widetilde{c}_G < \widetilde{c}_B$. In the first case, we have $\widetilde{c} = G(\widetilde{\rho}(\mu_R)) - G(\widetilde{\rho}(\mu_{NN}))$ and

$$1 - G(\widetilde{\rho}(\mu_R)) \geq 1 - \alpha_G > 1 - \alpha_B. \tag{12}$$

Hence, $G(\widetilde{\rho}(\mu_R)) \leq \alpha_G < 1$ and $\widetilde{c} \leq \alpha_G < 1$. Given the regime's strategy, the total probability that it chooses to repress-and-conceal a type-θ activist is $\Pr(\text{repress-and-conceal} \mid \theta) = H(\widetilde{c})$. Using the Bayes rule, we find that $\mu_{NN} = \left(\frac{\gamma H(\widetilde{c})q}{\gamma H(\widetilde{c})+1-\gamma}, \frac{\gamma H(\widetilde{c})(1-q)}{\gamma H(\widetilde{c})+1-\gamma}, \frac{1-\gamma}{\gamma H(\widetilde{c})+1-\gamma}\right)$, or, equivalently, $\mu_{NN} = (\gamma' q, \gamma'(1-q), 1-\gamma')$, where $\gamma' := \frac{H(\widetilde{c})\gamma}{H(\widetilde{c})\gamma+1-\gamma} < \gamma$.

The remainder of the proof shows that there is no equilibrium in which $\widetilde{c}_G < \widetilde{c}_B$. By contradiction, suppose that $\widetilde{c}_G < \widetilde{c}_B$. Then, from (10), $1 - \alpha_G > 1 - G(\widetilde{\rho}(\mu_R))$, so repress-and-reveal is strictly dominated if $\theta = G$. Hence, only the bad activists are publicly repressed (since repress-and-reveal must be on-path by Remark 2), the public's posterior belief upon observing revealed repression is $\mu_R = (0, 1, 0)$, and the regime's expected payoff from repress-and-reveal is $1 - G(\beta_B)$.

Next, note that if $\widetilde{c}_G < \widetilde{c}_B$, then from (11), $\widetilde{c}_G < 1$. Observe that now a regime of type (θ, c) such that $\theta = G$ and $c \in (\widetilde{c}_G, 1)$ has a profitable deviation. From (10), in any

equilibrium, this regime type either represses-and-reveals or concedes, so that its payoff is $\max\{1-\alpha_G, 1-G(\widetilde{\rho}(\mu_R))\}$, which is $1-\alpha_G$ if $\widetilde{c}_G < \widetilde{c}_B$. If this type deviates from the prescribed strategy and represses-and-reveals, then its payoff is $1-G(\beta_B)$. That deviation is profitable since $1-G(\beta_B) > 1-\alpha_G$ by Assumption 1.5, a contradiction. ∎

Proof of Proposition 2. From Proposition 1, the regime represses-and-conceals if $c < \widetilde{c}$, and represses-and-reveals or concedes otherwise. From (12) in the proof of Proposition 1, $1-G(\widetilde{\rho}(\mu_R)) \geq 1-\alpha_G > 1-\alpha_B$. We now show that, in any equilibrium, $1-G(\widetilde{\rho}(\mu_R)) = 1-\alpha_G$. If not, then $1-G(\widetilde{\rho}(\mu_R)) > 1-\alpha_G > 1-\alpha_B$ and the regime never concedes. In particular, when $c > \widetilde{c}$, the regime always represses-and-reveals. Then, $\mu_R = (q, 1-q, 0)$ and the expected payoff from repress-and-reveal is $1-G(\beta^e)$. However, by Assumption 1.2, we have $1-G(\beta^e) < 1-\alpha_G$, meaning that deviating to concession is strictly profitable for the regime of type $(G, c > \widetilde{c})$, a contradiction. Therefore, in any equilibrium, $1-G(\widetilde{\rho}(\mu_R)) = 1-\alpha_G > 1-\alpha_B$, the regime represses-and-reveals the good activist, and is indifferent between repress-and-reveal and conceding to the good activist.

Next, we characterize the probability with which the regime represses-and-reveals the good activist. Let

$$\kappa = \frac{\int \sigma(\textit{repress-and-reveal} \mid G, c)dc}{\int \sigma(\textit{repress-and-reveal} \mid B, c)dc} = \frac{\int_{\widetilde{c}}^{\overline{c}} \sigma(\textit{repress-and-reveal} \mid G, c)dc}{1-H(\widetilde{c})}$$

be the likelihood ratio of repress-and-reveal of good over bad activist, derived using the arguments we made in the previous paragraph. Then, $\mu_R = \left(\frac{\kappa q}{\kappa q+1-q}, \frac{1-q}{\kappa q+1-q}, 0\right)$ and the indifference condition of a regime with type $(G, c > \widetilde{c})$ is:

$$1-G(\widetilde{\rho}(\mu_R)) = 1-\alpha_G \iff G\left(\frac{\kappa q}{\kappa q+1-q}\beta_G + \frac{1-q}{\kappa q+1-q}\beta_B\right) = \alpha_G.$$

Therefore,

$$\kappa = \frac{1-q}{q}\frac{G^{-1}(\alpha_G)-\beta_B}{\beta_G - G^{-1}(\alpha_G)},$$

where $\kappa > 0 \iff G^{-1}(\alpha_G) > \beta_B$ (by Assumption 1.5) and $\kappa < 1 \iff \beta^e > G^{-1}(\alpha_G)$ (by Assumption 1.2).

Finally, we find the concealment cost threshold $\widetilde{c}$ using the fact that the regime with concealment cost $c = \widetilde{c}$ facing the good activist is indifferent between all three actions. In particular, type-$(G, \widetilde{c})$ regime is indifferent between repress-and-conceal and concede, meaning that $1 - G(\widetilde{\rho}(\mu_{NN})) - \widetilde{c} = 1 - \alpha_G \iff G(\widetilde{\rho}(\mu_{NN})) = \alpha_G - \widetilde{c}$. Substituting μ_{NN} from Proposition 1, we have

$$G\left(\frac{\gamma H(\widetilde{c})q}{\gamma H(\widetilde{c})+1-\gamma}\beta_G + \frac{\gamma H(\widetilde{c})(1-q)}{\gamma H(\widetilde{c})+1-\gamma}\beta_B\right) = G\left(\frac{\gamma H(\widetilde{c})}{\gamma H(\widetilde{c})+1-\gamma}\beta^e\right) = \alpha_G - \widetilde{c}.$$

This equation has a unique solution $\widetilde{c} \in (\underline{c}, \alpha_G)$ because the left-hand side $LHS(\widetilde{c})$ is continuous and increasing, while the right-hand side $RHS(\widetilde{c})$ is continuous and strictly decreasing; $LHS(\underline{c}) = G(0) = 0$, while $RHS(\underline{c}) = \alpha_G - \underline{c} > 0$ (by Assumption 1.4); $LHS(\alpha_G) = G\left(\frac{\gamma H(\alpha_G)}{\gamma H(\alpha_G)+1-\gamma}\beta^e\right) > 0$ (by Assumption 1.6), while $RHS(\alpha_G) = 0$. ∎

Proof of Proposition 3. The left-hand side in (4) is strictly increasing in $H(c) \in (0,1)$, and the right-hand side is strictly decreasing for the relevant range of $\widetilde{c}$. It follows that $\widetilde{c}_2 < \widetilde{c}_1$ and $H(\widetilde{c}_2) > H(\widetilde{c}_1)$. The results follow because Pr(revealed repression) $= (\kappa q + 1 - q)(1 - H(\tilde{c}))$ is strictly decreasing, while Pr(total repression) $= \kappa q + 1 - q + q(1-\kappa)H(\widetilde{c})$ is strictly increasing in $H(\widetilde{c})$ ∎

Proof of Proposition 5. In any equilibrium, a type-(θ, c) regime strictly prefers to repress-and-conceal if

$$1 - G(\widetilde{\rho}(\mu_{NN})) - c > \max\{1 - G(\widetilde{\rho}(\mu_R)), 1 - \alpha_\theta\},$$

which is equivalent to

$$c < \widetilde{c}_\theta := 1 - G(\widetilde{\rho}(\mu_{NN})) - \max\{1 - G(\widetilde{\rho}(\mu_R)), 1 - \alpha_\theta\}. \tag{13}$$

The regime is indifferent if $c = \widetilde{c}_\theta$ (measure-zero event), and strictly prefers not to repress-and-conceal if $c > \widetilde{c}_\theta$. Since $0 < \alpha_B < \alpha_G$ from Assumption 2.1, (13) implies:

$$\widetilde{c}_B \leq \widetilde{c}_G \leq 1. \tag{14}$$

If $\widetilde{c}_G = \widetilde{c}_B = \widetilde{c}$, then $1 - G(\widetilde{\rho}(\mu_R)) \geq 1 - \alpha_B > 1 - \alpha_G$. Thus, the type $(G, c > \widetilde{c})$ definitely represses-and-reveals, while the type $(B, c > \widetilde{c})$ represses-and-reveals or concedes. Since good activists are repressed-and-revealed more often than bad ones, we have $\mu_R = (q', 1 - q', 0)$, where $q' \geq q$, so $G(\widetilde{\rho}(\mu_R)) \geq G(\beta^e)$. Thus, $1 - \alpha_B \leq 1 - G(\widetilde{\rho}(\mu_R)) \leq 1 - G(\beta^e)$, i.e., $\alpha_B \geq G(\beta^e)$. That contradicts Assumption 2.2.

Thus, in any equilibrium, we must have $\widetilde{c}_B < \widetilde{c}_G$. If $\widetilde{c}_B < \widetilde{c}_G$, then from (13), we must have $1 - \alpha_B > 1 - G(\widetilde{\rho}(\mu_R))$, implying that the regime does not repress-and-reveal when $\theta = B$. Hence, in any PBE that satisfies the D1 refinement, the public's posterior belief upon observing revealed repression is $\mu_R = (1, 0, 0)$ and, from Lemma 1, $G(\widetilde{\rho}(\mu_R)) = G(\beta_G)$.[18] From Assumption 2.3, $\alpha_G > G(\beta_G)$, so that $1 - \alpha_G < 1 - G(\beta_G)$. Then, in equilibrium, when $c > \widetilde{c}_G$, the regime always represses-and-reveals the good activist; when $c > \widetilde{c}_B$, the regime always concedes to the bad activist. Otherwise, the regime chooses to repress-and-conceal.

[18] Reversing the payoff from concession to $\alpha_B < \alpha_G$ reverses the prediction of Remark 2 – the public's posterior belief upon observing unexpected repression must be $\mu_R = (1, 0, 0)$.

The indifference conditions are

$$\alpha_B - G\left(\frac{\gamma\left(H(\widetilde{c}_G)\ q\beta_G + H(\widetilde{c}_B)\ (1-q)\beta_B\right)}{\gamma\left(H(\widetilde{c}_G)\ q + H(\widetilde{c}_B)\ (1-q)\right) + 1 - \gamma}\right) = \widetilde{c}_B \tag{15}$$

$$G(\beta_G) - G\left(\frac{\gamma\left(H(\widetilde{c}_G)\ q\beta_G + H(\widetilde{c}_B)\ (1-q)\beta_B\right)}{\gamma\left(H(\widetilde{c}_G)\ q + H(\widetilde{c}_B)\ (1-q)\right) + 1 - \gamma}\right) = \widetilde{c}_G. \tag{16}$$

They correspond to the fixed points of the continuous function $\Gamma(\widetilde{c}_G, \widetilde{c}_B) : [-1, \alpha_B] \times [-1, G(\beta_G)] \to [-1, \alpha_B] \times [-1, G(\beta_G)]$. Thus, a solution exists. Clearly, the left-hand sides of both equations are strictly less than $\overline{c}$ by Assumptions 2.3 and 2.4. Moreover, without loss, if $\widetilde{c}_\theta \leq \underline{c}$, we set $\widetilde{c}_\theta = \underline{c}$. We show next that if $\widetilde{c}_B = \underline{c}$, then $\widetilde{c}_G > \underline{c}$.

If $\widetilde{c}_B = \underline{c}$, then (16) becomes

$$G\left(\frac{\gamma H(\widetilde{c}_G)\ q}{\gamma H(\widetilde{c}_G)\ q + 1 - \gamma}\beta_G\right) = G(\beta_G) - \widetilde{c}_G. \tag{17}$$

The left- (right-) hand side is increasing (strictly decreasing), so that the solution is unique. Moreover, $LHS(\widetilde{c}_G = \underline{c}) = 0 < G(\beta_G) - \underline{c}$, where the inequality follows from Assumption 2.5. Thus, we have $\widetilde{c}_B = \underline{c} < \widetilde{c}_G$. Further, $LHS(\widetilde{c}_G = G(\beta_G)) \geq 0$, while $RHS(\widetilde{c}_G = G(\beta_G)) = 0$ and strictly decreasing in $\widetilde{c}_G$. Thus, the solution is $\widetilde{c}_G \in (\underline{c}, G(\beta_G)]$.

We now prove additional results, which are not stated in the proposition. If $\widetilde{c}_B > \underline{c}$, then both thresholds are interior. From (15) and (16),

$$\widetilde{c}_G - \widetilde{c}_B = G(\beta_G) - \alpha_B > 0, \tag{18}$$

where the inequality follows from $G(\beta^e) < G(\beta_G)$ and Assumption 2.2. Substituting from (18) into (15) yields

$$G\left(\frac{\gamma\left(H(\widetilde{c}_B + G(\beta_G) - \alpha_B)\ q\beta_G + H(\widetilde{c}_B)\ (1-q)\beta_B\right)}{\gamma\left(H(\widetilde{c}_B + G(\beta_G) - \alpha_B)\ q + H(\widetilde{c}_B)\ (1-q)\right) + 1 - \gamma}\right) = \alpha_B - \widetilde{c}_B. \tag{19}$$

Further, if $H(x)$ is weakly log-convex, then the LHS of (19) is increasing in $\widetilde{c}_B$, while the RHS is strictly decreasing, so the solution is unique. ∎

Proof of Proposition 6. Proposition 2 describes the equilibrium under Assumption 1. From the proof of Proposition 2, $G(\widetilde{\rho}(\mu_R)) = \alpha_G$, so that $D > 0 \iff G(\gamma\beta^e) > \alpha_G$.

Proposition 5 describes the equilibrium under Assumption 2. In any equilibrium, the regime represses-and-reveals organized good activists with a strictly positive probability and never represses-and-reveals organized bad activists. Hence, there is always backlash following revealed repression: $G(\widetilde{\rho}(\mu_0)) = G(\gamma\beta^e) \leq G(\beta^e) < G(\beta_G) = G(\widetilde{\rho}(\mu_R))$.

Finally, $\widetilde{\rho}(\mu_0) = \widetilde{\rho}(\mu_R)$ happens only for a measure-zero subset of parameters. To see this, note that $\widetilde{\rho}(\mu_0) = \beta^e$. Under Assumption 1, $\widetilde{\rho}(\mu_R) = G^{-1}(\alpha_G)$. Under Assumption 2, $\widetilde{\rho}(\mu_R) = \beta_G$. ∎

Proof of Proposition 7. Under Assumption 1, from Proposition 1 and Lemma 1, $G(\widetilde{\rho}(\mu_{NN})) \leq G(\widetilde{\rho}(\mu_0))$. The result follows from (9) in the text and its surrounding explanation. ∎

References

Abrahamian, Ervand (1999), *Tortured Confessions: Prisons and Public Recantations in Modern Iran*, University of California Press. (p. 1.)

Acemoglu, Daron and James A. Robinson (2006), *Economic Origins of Dictatorship and Democracy*, Cambridge University Press. (p. 8.)

Almond, Gabriel (1956), "Comparative Political Systems", *The Journal of Politics*, 18, pp. 391-409. (p. 3.)

Amnesty International (2018), *Blood-Soaked Secrets: Why Iran's 1988 Prison Massacres Are Ongoing Crimes Against Humanity*, `https://www.amnesty.org/en/documents/mde13/9421/2018/en/`. (p. 1.)

— (2023), *Iran: Young Man at Grave Risk of Execution in Connection with Protests Amid 'Killing Spree'*, Jan. 2023, `https://www.amnesty.org/en/latest/news/2023/01/iran-young-man-at-grave-risk-of-execution-in-connection-with-protests-amid-killing-spree/`. (p. 1.)

Andirin, Veli, Yusuf Neggers, Mehdi Shadmehr, and Jesse M Shapiro (2022), *Surveillance of Repression: Theory and Implementation*, Working Paper 30167, National Bureau of Economic Research, `http://www.nber.org/papers/w30167`. (pp. 2, 5, 6, 12, 24.)

Aytaç, S. Erdem, Luis Schiumerini, and Susan Stokes (2018), "Why Do People Join Backlash Protests? Lessons from Turkey", *Journal of Conflict Resolution*, 62, 6, pp. 1205-1228. (pp. 5, 8.)

Aytaç, S. Erdem and Susan Stokes (2019), *Why Bother?: Rethinking Participation in Elections and Protests*, Cambridge University Press. (p. 8.)

Beraja, Martin, Andrew Kao, David Y. Yang, and Noam Yuchtman (2023), "AI-tocracy", *The Quarterly Journal of Economics*, 138, 3 (Mar. 2023), pp. 1349-1402. (p. 12.)

Besley, Timothy and Torsten Persson (2011), "The Logic of Political Violence", *Quarterly Journal of Economics*, 126, 3, pp. 1411-1445. (p. 8.)

Besley, Timothy and Andrea Prat (2006), "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability", *American Economic Review*, 96, 3 (June 2006), pp. 720-736. (p. 14.)

Boix, Carles (2003), *Democracy and Redistribution*, Cambridge University Press. (p. 8.)

Boix, Carles and Milan Svolik (2013), "The Foundations of Limited Authoritarian Government: Institutions and Power-Sharing in Dictatorships", *Journal of Politics*, 75, 2, pp. 300-316. (p. 8.)

Boleslavsky, Raphael, Tom Ginsburg, and Mehdi Shadmehr (2019), "Repression, Judiciary, and Protest in Autocracies", `https://ssrn.com/abstract=3419727`. (pp. 8, 13.)

Bueno de Mesquita, Ethan (2010), "Regime Change and Revolutionary Entrepreneurs", *American Political Science Review*, 104, 3, pp. 446-466. (p. 8.)

Bueno De Mesquita, Ethan and Mehdi Shadmehr (2023), "Rebel Motivations and Repression", en, *American Political Science Review*, 117, 2 (May 2023), pp. 734-750. (p. 8.)

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott (2020), "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia", *American Economic Review*, 110, 10 (Oct. 2020), pp. 2997-3029. (p. 23.)

Cantoni, Davide, David Y Yang, Noam Yuchtman, and Y Jane Zhang (2019), "Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement", *The Quarterly Journal of Economics*, 134, 2 (Jan. 2019), pp. 1021-1077. (p. 23.)

Carey, Peter D., Curtis Bell, Emily Hencken Ritter, and Scott Wolford (2022), "Oil discoveries, civil war, and preventive state repression", *Journal of Peace Research*, 59, 5, pp. 648-662. (p. 24.)

Carment, David (2010), *Country Indicators for Foreign Policy Project*, `https://carleton.ca/cifp/about-cifp/`. (p. 2.)

Casper, Brett and Scott Tyson (2014), "Popular Protest and Elite Coordination in a Coup d'etat", *Journal of Politics*, 76, 2, pp. 548-564. (p. 8.)

Chau, Tak-Huen, Mai Hassan, and Andrew T. Little (2025), "Communication, coordination, and surveillance in the shadow of repression", *American Journal of Political Science, forthcoming*. (p. 8.)

Chen, Heng and Derek D. Huo (2025), "The Scar of Political Conflict: Evidence from Tear Gas Deployments in Hong Kong", `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5180737`. (p. 3.)

Chen, Heng and Wing Suen (2017), "Aspiring for Change: A Theory of Middle Class Activism", *The Economic Journal*, 127, 603 (Aug. 2017), pp. 1318-1347. (p. 8.)

— (2021), "Radicalism in Mass Movements: Asymmetric Information and Endogenous Leadership", *American Political Science Review*, 115, 1, pp. 286-306. (p. 8.)

Chenoweth, Erica (2023), "The Role of Violence in Nonviolent Resistance", *Annual Review of Political Science*, 26, 1 (June 2023), pp. 55-77. (pp. 3, 8.)

Chenoweth, Erica and Maria Stephan (2011), *Why Civil Resistance Works: The Strategic Logic of Nonviolent Conflict*, Columbia University Press. (pp. 3, 8.)

Cingranelli, David and Mikhail Filippov (2018), "Are Human Rights Practices Improving?", *American Political Science Review*, 112, 4, pp. 1083-1089. (p. 7.)

Condra, Luke N. and Jacob N. Shapiro (2012), "Who Takes the Blame? The Strategic Effects of Collateral Damage", *American Journal of Political Science*, 56, 1, pp. 167-187. (p. 8.)

Conrad, Courtenay R., Daniel W. Hill Jr., and Will H. Moore (2018), "Torture and the Limits of Democratic Institutions", *Journal of Peace Research*, 55, 1, pp. 3-17. (p. 1.)

Conrad, Courtenay R. and Emily Hencken Ritter (2019), *Contentious Compliance: Dissent and Repression under International Human Rights Law*, Oxford University Press. (p. 6.)

Cope, Kevin L., Charles Crabtree, and Christopher J. Fariss (2020), "Patterns of disagreement in indicators of state repression", *Political Science Research and Methods*, 8, 1, pp. 178-187. (p. 7.)

Danneman, Nathan and Emily Hencken Ritter (2014), "Contagious Rebellion and Preemptive Repression", *Journal of Conflict Resolution*, 58, 2, pp. 254-279. (p. 24.)

Dassin, Joan (ed.) (1998), *Torture in Brazil: A Shocking Report on the Pervasive Use of Torture by Brazilian Military Governments, 1964–1979*, University of Texas Press. (p. 1.)

Davenport, Christian (2007), "State Repression and Political Order", *Annual Review of Political Science*, 10, 1 (June 2007), pp. 1-23. (pp. 2, 5, 8.)

— (2015), *How Social Movements Die: Repression and Demobilization of the Republic of New Africa*, Cambridge University Press. (p. 4.)

Dell, Melissa and Pablo Querubin (2018), "Nation Building through Foreign Intervention: Evidence from Discontinuities in Military Strategies", *Quarterly Journal of Economics*, 133, 2, pp. 701-764. (p. 8.)

Diermeier, Daniel, Hülya Eraslan, and Antonio Merlo (2003), "A Structural Model of Government Formation", *Econometrica*, 71, 1, pp. 27-70. (p. 5.)

Downes, Alexander (2008), *Targeting Civilians in War*, Cornell University Press. (p. 8.)
Dugan, Laura and Erica Chenoweth (2012), "Moving beyond Deterrence: The Effectiveness of Raising the Expected Utility of Abstaining from Terrorism in Israel", *American Sociological Review*, 77, 4, pp. 597-624. (pp. 7, 8.)
Dye, Ronald A. (1985), "Disclosure of Nonproprietary Information", *Journal of Accounting Research*, 23, 1, pp. 123-145. (p. 14.)
Earl, Jennifer (2011), "Political Repression: Iron Fists, Velvet Gloves, and Diffuse Control", *Annual Review of Sociology*, 37, 1, pp. 261-284. (pp. 2, 5, 8.)
Egorov, Georgy and Konstantin Sonin (2021), "Elections in Non-Democracies", *The Economic Journal*, 131, 636 (Oct. 2021), pp. 1682-1716. (pp. 8, 21.)
— (2024), "The Political Economics of Non-democracy", *Journal of Economic Literature*, 62, 2 (June 2024), pp. 594-636. (pp. 8, 21.)
Esberg, Jane (2021), "Anticipating Dissent: The Repression of Politicians in Pinochet's Chile", *The Journal of Politics*, 83, 2 (Apr. 2021), pp. 689-705. (p. 24.)
Fariss, Christopher J. (2014), "Respect for Human Rights Has Improved over Time: Modeling the Changing Standard of Accountability", *American Political Science Review*, 108, 2, pp. 297-318. (p. 7.)
— (2019), "Yes, Human Rights Practices Are Improving Over Time", *American Political Science Review*, 113, 3, pp. 868-881. (p. 7.)
Francisco, Ronald (2004), "After the Massacre: Mobilization in the Wake of Harsh Repression", *Mobilization*, 9, pp. 107-126. (pp. 5, 7.)
Gehlbach, Scott, Zhaotian Luo, Anton Shirikov, and Dmitriy Vorobyev (2025), "Is there really a dictator's dilemma? Information and repression in autocracy", *American Journal of Political Science, forthcoming*. (p. 21.)
Gibilisco, Michael (2021), "Decentralization, Repression, and Gambling for Unity", *The Journal of Politics*, 83, 4 (Oct. 2021), pp. 1353-1368. (p. 8.)
Gibilisco, Michael and Sergio Montero (2022), "Do Major-Power Interventions Encourage the Onset of Civil Conflict? A Structural Analysis", *The Journal of Politics*, 84, 2, pp. 944-959. (p. 5.)
Gieczewski, Germán and Maria Titova (2024), "Coalition-Proof Disclosure", `https://maria-titova.com/papers/CPD.pdf`. (p. 14.)
Gitmez, Arda, Pooya Molavi, and Konstantin Sonin (2025), "Propaganda and Repression in Diverse Societies", `https://pooyamolavi.com/Propaganda_and_Repression.pdf`. (p. 21.)
González, Felipe and Mounu Prem (2024), "Police Violence, Student Protests, and Educational Performance", *The Review of Economics and Statistics*, 106, 3 (May 2024), pp. 712-727. (p. 8.)
Greitens, Sheena Chestnut (2016), *Dictators and their Secret Police: Coercive Institutions and State Violence*, Cambridge University Press. (p. 12.)
Guardian (2009), *China Executes Nine Over Xinjiang Riots*, Nov. 2009, `https://www.theguardian.com/world/2009/nov/09/china-executes-nine-xinjiang-uighur`. (p. 1.)
Guriev, Sergei and Daniel Treisman (2020), "A theory of informational autocracy", *Journal of Public Economics*, 186 (June 2020), p. 104158. (pp. 8, 21.)
— (2022), *Spin Dictators: The Changing Face of Tyranny in the 21st Century*, Princeton University Press. (pp. 1, 2, 12.)

Hager, Anselm and Krzysztof Krakowski (2022), "Does State Repression Spark Protests? Evidence from Secret Police Surveillance in Communist Poland", *American Political Science Review*, 116, 2 (May 2022), pp. 564-579. (p. 3.)

Hart, Sergiu, Ilan Kremer, and Motty Perry (2017), "Evidence Games: Truth and Commitment", *American Economic Review*, 107, 3 (Mar. 2017), pp. 690-713. (pp. 9, 14.)

Hassan, Mai, Daniel Mattingly, and Elizabeth R. Nugent (2022), "Political Control", *Annual Review of Political Science*, 25, pp. 155-174. (p. 2.)

Huang, Haifeng (2015), "International Knowledge and Domestic Evaluations in a Changing Society: The Case of China", *American Political Science Review*, 109, 3, pp. 613-634. (p. 23.)

— (2018), "The Pathology of Hard Propaganda", *The Journal of Politics*, 80, 3 (July 2018), pp. 1034-1038. (p. 23.)

Human Rights Watch (2022), *Saudi Arabia: Mass Execution of 81 Men – Rampant Abuses in Criminal Justice System Make Fair Trials Highly Implausible*, Mar. 2022, `https://www.hrw.org/news/2022/03/15/saudi-arabia-mass-execution-81-men`. (p. 1.)

Kalyvas, Stathis (2006), *The Logic of Violence in Civil War*, Cambridge University Press. (p. 8.)

King, Gary, Jennifer Pan, and Margaret E. Roberts (2013), "How Censorship in China Allows Government Criticism but Silences Collective Expression", *The American Political Science Review*, 107, 2, pp. 326-343. (p. 1.)

Kocher, Matthew Adam, Thomas B. Pepinsky, and Stathis N. Kalyvas (2011), "Aerial Bombing and Counterinsurgency in the Vietnam War", *American Journal of Political Science*, 55, 2, pp. 201-218. (p. 8.)

LaFree, Gary, Laura Dugan, and Raven Korte (2009), "The Impact of British Counterterrorist Strategies on Political Violence in Northern Ireland: Comparing Deterrence and Backlash Models", *Criminology*, 47, 1, pp. 17-45. (pp. 5, 7.)

Lawrence, Adria K. (2017), "Repression and Activism among the Arab Spring's First Movers: Evidence from Morocco's February 20th Movement", *British Journal of Political Science*, 47, 3 (July 2017), pp. 699-718. (p. 8.)

Lebas, Adrienne and Lauren E. Young (2024), "Repression and Dissent in Moments of Uncertainty: Panel Data Evidence from Zimbabwe", *American Political Science Review*, 118, 2, pp. 584-601. (pp. 8, 23.)

Lebovic, James H. and Erik Voeten (2009), "The Cost of Shame: International Organizations and Foreign Aid in the Punishing of Human Rights Violators", *Journal of Peace Research*, 46, 1, pp. 79-97. (p. 2.)

Li, Yao (2018), *Playing by the Informal Rules: Why the Chinese Regime Remains Stable despite Rising Protests*, Cambridge University Press. (p. 12.)

Little, Andrew T., Joshua A. Tucker, and Tom LaGatta (2015), "Elections, Protest, and Alternation of Power", *The Journal of Politics*, 77, 4, pp. 1142-1156. (p. 8.)

Lyall, Jason (2009), "Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya", *Journal of Conflict Resolution*, 53, 3, pp. 331-362. (p. 8.)

Mansbridge, Jane (2012), "On the Importance of Getting Things Done", *PS: Political Science and Politics*, 45, pp. 1-8. (p. 3.)

— (2014), "What Is Political Science For?", *Perspectives on Politics*, 12, pp. 8-17. (p. 3.)

Martin, Brian (2007), *Justice Ignited: The Dynamics of Backfire*, Rowman & Littlefield. (pp. 5, 7.)

Mazzaro, Kyong (2023), "Anti-Media Discourse and Violence Against Journalists: Evidence from Chávez's Venezuela", *The International Journal of Press/Politics*, 28, 3, pp. 469-492. (p. 12.)

McAdam, Doug (1999), *Political Process and the Development of Black Insurgency, 1930–1970*, 2nd ed., University of Chicago Press. (p. 10.)

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal (2006), *Polarized America: The Dance of Ideology and Unequal Riches*, Updated Edition, 2016, MIT Press. (p. 5.)

Millennium Challenge Corporation (2020), *MCC Selection Indicators*, `https://web.archive.org/web/20200610141639/https://www.mcc.gov/who-we-fund/indicators`. (p. 2.)

Moore, Will H. (1998), "Repression and Dissent: Substitution, Context, and Timing", *American Journal of Political Science*, 42, 3, pp. 851-873. (p. 2.)

— (2000), "The Repression of Dissent: A Substitution Model of Government Coercion", *Journal of Conflict Resolution*, 44, 1, pp. 107-127. (p. 2.)

Nalepa, Monika and Grigore Pop-Eleches (2022), "Authoritarian Infiltration of Organizations: Causes and Consequences", *The Journal of Politics*, 84, 2 (Apr. 2022), pp. 861-873. (p. 12.)

Open Government Partnership (2019), *Open Government Partnership Global Report*, `https://www.opengovpartnership.org/wp-content/uploads/2019/09/Global-Report_Volume-1.pdf`. (p. 2.)

Opp, Karl-Dieter and Wolfgang Ruehl (1990), "Repression, Micromobilization and Political Protest", *Social Forces*, 69, 52, pp. 1-47. (pp. 3, 8, 9.)

Pan, Jennifer and Alexandra A. Siegel (2020), "How Saudi Crackdowns Fail to Silence Online Dissent", en, *American Political Science Review*, 114, 1 (Feb. 2020), pp. 109-125. (p. 9.)

Pan, Jennifer, Xu Xu, and Yiqing Xu (2025), "Disguised Repression: Targeting Opponents with Non-Political Crimes to Undermine Dissent", *The Journal of Politics, forthcoming.* (pp. 1, 12.)

Pearlman, Wendy (2018), "Moral Identity and Protest Cascades in Syria", *British Journal of Political Science*, 48, 4 (Oct. 2018), pp. 877-901. (p. 8.)

Persson, Torsten and Guido Tabellini (2009), "Democratic Capital: The Nexus of Political and Economic Change", *American Economic Journal: Macroeconomics*, 1, 2, pp. 88-126. (p. 8.)

Poe, Steven C. (1992), "Human Rights and Economic Aid under Ronald Reagan and Jimmy Carter", *American Journal of Political Science*, 36, 1, pp. 147-167. (p. 2.)

Pop-Eleches, Grigore, Graeme Robertson, and Bryn Rosenfeld (2022), "Protest Participation and Attitude Change: Evidence from Ukraine's Euromaidan Revolution", *The Journal of Politics*, 84, 2, pp. 625-638. (p. 23.)

Rasler, Karen (1996), "Concessions, Repression, and Political Protest in the Iranian Revolution", *American Sociological Review*, 61, 1, pp. 132-152. (p. 8.)

Rejali, Darius (2007), *Torture and Democracy*, Princeton University Press. (pp. 1, 12.)

Reporters Without Borders (2008), *Tiananmen Square Massacre Still a Taboo Subject in Press and Online on 19th Anniversary / 1989/6/4*, June 2008, `https://rsf.org/en/tiananmen-square-massacre-still-taboo-subject-press-and-online-19th-anniversary-1989%E5%B9%B46%E6%9C%884%E6%97%A5-%E5%AA%92%E4%BD%93%E6%B0%B8%E8%BF%9C%E7%9A%84%E7%A6%81%E5%8C%BA`. (p. 1.)

Reporters Without Borders (2023), *In 2023, RSF provided financial assistance to 460 journalists who were being threatened or had to flee abroad*, https://rsf.org/en/2023-rsf-provided-financial-assistance-460-journalists-who-were-being-threatened-or-had-flee-abroad. (p. 12.)

Ritter, Emily Hencken (2014), "Policy Disputes, Political Survival, and the Onset and Severity of State Repression." *Journal of Conflict Resolution*, 58, 1, pp. 143-168. (pp. 6, 8.)

Ritter, Emily Hencken and Courtenay R. Conrad (2016), "Preventing and Responding to Dissent: The Observational Challenges of Explaining Strategic Repression", *American Political Science Review*, 110, 1 (Feb. 2016), pp. 85-99. (p. 24.)

Roberts, Margaret E. (2018), *Censored: Distraction and Diversion Inside China's Great Firewall*, English, Princeton University Press. (p. 12.)

Rosenfeld, Bryn (2017), "Reevaluating the Middle-Class Protest Paradigm: A Case-Control Study of Democratic Protest Coalitions in Russia", *American Political Science Review*, 111, 4, pp. 637-652. (p. 23.)

— (2025), "Risk Attitudes and Political Participation Under Autocracy", https://sites.google.com/site/brynrosenfeld/. (pp. 8, 23.)

Rozenas, Arturas (2020), "A Theory of Demographically Targeted Repression", *Journal of Conflict Resolution*, 64, 7-8 (Aug. 2020), pp. 1254-1278. (p. 8.)

Rozenas, Arturas and Yuri M. Zhukov (2019), "Mass Repression and Political Loyalty: Evidence from Stalin's 'Terror by Hunger'", *American Political Science Review*, 113, 2 (May 2019), pp. 569-583. (p. 8.)

Shadmehr, Mehdi and Dan Bernhardt (2011), "Collective Action with Uncertain Payoffs: Coordination, Public Signals, and Punishment Dilemmas", *The American Political Science Review*, 105, 4, pp. 829-851. (pp. 3, 8, 10.)

— (2015), "State Censorship", *American Economic Journal: Microeconomics*, 7, 2, pp. 280-307. (p. 14.)

— (2019), "Vanguards in Revolution", *Games and Economic Behavior*, 115 (May 2019), pp. 146-166. (p. 3.)

Shadmehr, Mehdi and Raphael Boleslavsky (2022), "International Pressure, State Repression, and the Spread of Protest", *The Journal of Politics*, 84, 1 (Jan. 2022), pp. 148-165. (pp. 3, 8, 10, 13.)

Signorino, Curtis (1999), "Strategic Interaction and the Statistical Analysis of International Conflict", *American Political Science Review*, 93, 2, pp. 279-297. (p. 5.)

Siqueira, Kevin and Todd Sandler (2006), "Terrorists versus the Government: Strategic Interaction, Support, and Sponsorship", *Journal of Conflict Resolution*, 50, 6, pp. 878-898. (p. 8.)

Stoll, David (1993), *Between Two Armies in the Ixil Towns of Guatemala*, Columbia University Press. (p. 8.)

Sullivan, Christopher M. (2016), "Political Repression and the Destruction of Dissident Organizations: Evidence from the Archives of the Guatemalan National Police", *World Politics*, 68, 4 (Oct. 2016), pp. 645-676. (pp. 4, 12.)

Tarrow, Sidney (2011), *Power in Movement: Social Movements and Contentious Politics*, 3rd ed., Cambridge University Press. (pp. 3, 4, 8, 10.)

Tertytchnaya, Katerina (2023), ""This Rally is Not Authorized": Preventive Repression and Public Opinion in Electoral Autocracies", *World Politics*, 75, 3, pp. 482-522. (pp. 3, 23.)

Tertytchnaya, Katerina and Tomila Lankina (2020), "Electoral Protests and Political Attitudes under Electoral Authoritarianism", *The Journal of Politics*, 82, 1, pp. 285-299. (p. 23.)

Tilly, Charles (1978), *From Mobilization to Revolution*, Addison-Wesley. (pp. 4, 8.)

Tilly, Charles and Sidney Tarrow (2007), *Contentious Politics*, Paradigm Publishers. (p. 10.)

Toft, Monica Duffy and Yuri M. Zhukov (2015), "Islamists and Nationalists: Rebel Motivation and Counterinsurgency in Russia's North Caucasus", *American Political Science Review*, 109, 2, pp. 222-238. (p. 8.)

Verrecchia, Robert E. (2001), "Essays on disclosure", *Journal of Accounting and Economics*, 32, 1, pp. 97-180. (p. 9.)

Wasow, Omar (2020), "Agenda Seeding: How 1960s Black Protests Moved Elites, Public Opinion and Voting", *American Political Science Review*, 114, 3 (Aug. 2020), pp. 638-659. (p. 3.)

Wooldridge, Jeffrey M. (2010), *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press. (p. 6.)