Repression and the Spread of Protest*

Mehdi Shadmehr† Raphael Boleslavsky‡

Abstract

We analyze the strategic interaction between a state that must decide whether to repress a group of activists, and a bystander citizen, representing the public, who must decide whether to protest following the state's repression. The bystander is uncertain about both the nature of the activists' demands and the intentions of the state that represses them. We show that the information revealed by the state's repression can cause the spread of protest, and the potential for repression backfire depends on social norms, economic conditions, whether the reforms are minor or fundamental, and the popularity of the activists. We then characterize the subtle strategic interactions that arise when the regime can ex-ante set up institutions (e.g., independent judiciary) that limit its repression capacity. We show that commitment power may, paradoxically, increase repression, and highlight the non-monotone relationship between the activists' popularity and the likelihood of repression.

JEL Codes: D74, D83.

Keywords: Repression, Repression Backfire, Protest, Judicial Independence, Legitimate Coercion, Social Norms.

*We wish to thank Dan Bernhardt, Brandice Canes-Wrone, Odilon Câmara, Charles Cameron, Chris Cotton, Georgy Egorov, Justin Fox, Camilo Garcia-Jimeno, Adam Meirowitz, Steve Morris, Michael Peress, Andrea Prat, Carlo Prato, Soroush Ghazi, Kristopher Ramsay, Kai Steverson, Milan Svolik, Scott Tyson, Léonard Wantchékon, and participants at 2015 SPSA.

†Department of Economics, University of Miami, and Politics Department, Princeton University, 130 Corwin Hall, Princeton, NJ 08544 E-mail: shadmehr@princeton.edu

‡Department of Economics, University of Miami, 5250 University Dr., Coral Gables, FL 33146. E-mail: r.boleslavsky@bus.miami.edu

1 INTRODUCTION

On September 8, 1978, Iranian security forces fired on protesters in Tehran, killing many demonstrators. Soon after, protests continued in larger numbers, and strikes swept the country, culminating in the 1979 Iranian Revolution (Abrahamian 1982). Iran is not an isolated case. Martin (2007) provides several case studies of the spread of protest following repression, and Francisco (2004) quantifies the magnitude of such "backlash protests" in many more cases—see Earl (2011) for a review.[1] The theoretical literature has focused on the deterrence effect of repression, ignoring the empirical evidence that repression may cause, not deter, the spread of protest. This paper develops a model in which repression can cause the spread of protest by revealing information about both the nature of the activists' demands and the intentions of the state. We then investigate the role of institutions that limit the government's ability to repress legitimate dissent, showing that such institutions may, paradoxically, increase repression.

We consider contexts in which a group of activists has initiated a protest and put forth demands, and the state must decide whether to concede to their demands or repress them. We refer to the state's use of coercive force as repression, including imprisonment, killing, or other punishments.[2] Because a main responsibility of the state is to protect its citizens against harm by transgressors, the public recognizes that the use of coercive force may be legitimate.[3] However, both the activists who protest and the states that repress them claim that their actions are in the public's best interests. Given the difficulties of obtaining precise information, the public remains uncertain about the nature of the activists' demands and the intentions of the state. This often-ignored uncertainty is at the core of the interactions between the state and citizens in this paper.

We analyze the strategic interaction between a government that must decide whether to repress a group of activists, and a bystander citizen, representing the general public, who must decide whether to join the activists' protest upon observing repression. There are two types of activists, good and bad. The good activists' demands (if implemented) are beneficial to the public, while the bad activists' demands are harmful. Similarly, there are two types of governments, good and bad.

[1] Other examples abound. On March 21, 1960, South African security forces fired on demonstrators in Sharpeville, killing unarmed civilians. Subsequently, protests and strikes spread throughout the country. The Ukranian "Euromaidan" revolution began on the night of November 21, 2013, when a public protest was initiated in Kiev's Independence Square. In response to the violent dispersal of protestors on November 30, the scale of protest expanded dramatically, culminating in the ouster of President Yanukovich.

[2] This definition of repression is consistent with the notion of repression used in theoretical models (see the literature review), and with the sociology literature on repression. For example, Tilly (1978) defines repression as any action by the state that "raises the contenders cost of collective action" (p. 100).

[3] This view resonates with the Weber's idea of "the legitimate use of physical force," or "legitimate coercion" as appears in Almond (1956) and Mansbridge (2012, 2014).

Both types of government prefer to stay in power, but they differ in their preferences for reform. The good government's preferences for change are aligned with the public's: the good government prefers the good activists' reforms over the status quo and the status quo over the bad activists' reforms. However, the bad government prefers the status quo over both beneficial and harmful reforms. The government observes the activists' type. In contrast, the public does not observe the types of the activists or the government. Therefore, upon observing repression, the public cannot distinguish whether a bad government has repressed good activists—blocking beneficial reforms—or the government (of either type) has repressed bad activists—a necessary action that protects the public. This uncertainty underlies the bystander citizen's fundamental tradeoff: by supporting the government's decision to repress activists he risks blocking beneficial reforms, but by supporting activists and toppling the government he risks implementing harmful social changes.

We first analyze a setting in which the government cannot commit to a repression strategy, and must decide whether to repress after observing the activists' types. The public, upon observing repression, may join the activists and topple the government. This reduces the bad government's incentive to repress good activists, but it also reduces the good government's incentive to protect the public by repressing bad activists instead of conceding to their harmful demands. In fact, holding the bad government's strategy fixed, increases in the likelihood of repression by the good government *reduce* the public's incentives to join the activists. Because a government's best response is also decreasing, this generates the potential for multiple equilibria. In fact, we show that there can be three equilibria, which can be ranked according to their level of repression into low, intermediate, and high repression equilibrium.

The multiplicity of equilibria suggests that social norms play an important role in determining the state's response to dissent and the public's response to repression by shaping the expectations of the public and the state of each other's behavior.[4] Our results suggest that higher inequality or lower income levels can increase or decrease the level of repression, depending on the social norms that select the equilibrium. Because social norms are likely to vary across countries, this may provide a rationale for the conflicting empirical results describing the relationship between income or inequality and repression and violence. Further, we show that social norms become more empirically relevant when the public expects the bad government to be more harmful.[5]

[4]This interpretation is in line with the literature that interprets equilibrium selection as a reflection of social norms (Acemoglu and Jackson 2015; Burke and Young 2011; Kreps 1990; Postlewaite 2011; Young 2015).

[5]As the bystander's status quo payoff under the bad ruler falls, multiple equilibria arise in a larger subset of parameters. See the discussion of corollaries 1 and 2.

In the low and intermediate repression equilibria, protest can spread to bystander citizens after repression. However, this does not imply that repression has backfired because the bystander may have been willing to join the activists even absent repression. This is particularly relevant because we show that, in equilibrium, the public updates negatively about *both* the government and the activists. Thus, we adopt a narrower notion of repression backfire: we say repression backfires when the information revealed by observing repression increases the likelihood that the public protests.

We show that repression backfire occurs in equilibrium if and only if the following conditions hold. (1) Social norms select the low or intermediate repression equilibrium. When the public is apathetic about repression (as in the high repression equilibrium), repression never backfires. (2) The activists' popularity, captured by the public's prior beliefs that the activists are good, is high, but not too high. That is, the the public is sufficiently uncertain about the activists' types. When the activists are too likely to be bad, the public does not protest; When they are too likely to be good, the public would have been willing to join them even absent repression. (3) The potential gains of good reforms (upside) relative to the potential losses of bad reforms (downside) are higher under a bad government.[6] Qualitatively, this condition is easier to satisfy when economy is worse or inequality is higher. (4) The bad government's incentives to repress good activists are larger than the good government's incentive to repress bad activists. Otherwise, the good government represses the bad activists so much that, in equilibrium, the information revealed by repression does not increase the bystander's likelihood of protest. Qualitatively, this condition is easier to satisfy when the good activists demand more fundamental reforms, e.g., democratization or independence, rather than environmental reforms or combating corruption.[7]

If bad politicians could observably commit to a repression strategy (without revealing their type), they could potentially gain by manipulating the public's beliefs, so that they update less negatively about the government following repression. For example, institutions such as an independent judiciary can act as a commitment device by limiting the ability of the government to repress legitimate dissidents. This leads us to analyze what happens when bad politicians can set up institutions that enable them to commit to a repression strategy before observing the activists' type. In non-transparent authoritarian regimes, when the public observes repression, it cannot distinguish which type of politician has ordered the repression, but knows the regime's repression

[6]This condition is satisfied if, for example, when the government is already good, it is harder to make it better, and when it is already bad, it is harder to make it worse.

[7]In fact, in the low repression equilibrium—the low repression equilibrium is unique when the activists are relatively likely to be good—when the good activists demand more fundamental reforms, the bad government is more likely to *concede*, and the public is more likely to join the protest if the government uses repression.

limits. We show that the bad government's ability to commit imposes more stringent conditions for repression to backfire, but the nature of these conditions remain similar to those we discussed for the case of no commitment.

Paradoxically, setting up institutions that limit the repression of good activists can enable the bad government[8] to repress good activists *more* than it would without those institutions. The intuition hinges on the good government's response when it becomes the decision maker—in the subgame that follows the bad government choice of repression. When the good government represses the bad activists more, the bystander views repression as better news about the government and worse news about the activist, and becomes less inclined to protest. The bad government leverages this by increasing its repression of the good activists. In particular, the bad government's ability to limit its repression allows it to raise repression by "supporting" the equilibrium in which the good government represses bad activists more often, thereby reducing the likelihood of the bystander's protest. Without commitment, the bad government would increase the repression of good activists even further, thereby upsetting the equilibrium.

One may think that when the activists are more popular, the regime tends to increases the institutional constraints on repression. However, we show that the bad government's repression strategy is *non-monotone* in the activists' level of popularity as captured by the public's prior beliefs that the activists are good. When the activists are more likely to be bad, the bad government gains less from raising the repression of good activists, but the public also has less incentive to protest following repression. We show that the combined effect of these conflicting forces results in the bad government setting up institutions to limit repression when the likelihood of bad activists is low or high, but not when that likelihood is either intermediate or very high. Turning to the government's popularity (as captured by the public's prior beliefs that the government is good), we show that as the government's popularity falls, the government responds by putting stricter institutional constraints on the repression of good activists. In fact, when the public opinion is that the government is sufficiently likely to be bad, the government limits its repression capacity so much that the public never joins the protest following repression.

We next discuss the literature. Section 2 presents the model and the discussion of its assumptions, and Section 3 provides some preliminary results. We present the analyses of the equilibria with and without commitment in Sections 4 and 5. A conclusion follows. An Online Appendix contains an analysis of the bystander's commitment and the analyses of the cases omitted in the text.

[8]We use the term bad government and bad politician interchangeably.

1.1 LITERATURE REVIEW

Our paper is related to the political economy literature on protests and revolutions. This literature primarily studies coordination and information aggregation (Lohmann 1994; Shadmehr and Bernhardt 2011; Bueno de Mesquita 2014; Casper and Tyson 2014; Chen et al. 2014; Chen and Suen 2015; Tyson and Smith 2014), the role of vanguards (Majumdar and Mukand 2008; Bueno de Mesquita 2010; Shadmehr and Bernhardt 2012; Landa and Tyson 2014) and their tactics (Bueno de Mesquita 2013; Wantchékon and García-Ponce 2014), the role of media (Egorov et al. 2009; Edmond 2013; Shadmehr and Bernhardt 2015) and elections (Little 2012; Rozenas 2013; Egorov and Sonin 2015), and the synergies of ethnic mobilization (Esteban and Ray 2008, 2011). The literature has focused on the deterrence effect of repression, modeling repression as a state action that directly decreases the likelihood of revolution (Acemoglu and Robinson 2000, 2001, 2006; Boix 2003; Besley and Persson 2011; Conrad and Ritter 2013; Svolik 2013; Shadmehr 2014), or increases the cost of dissent (Lichbach 1987; Persson and Tabellini 2009; Bueno de Mesquita 2010; Fearon 2011; Shadmehr and Bernhardt 2011; Boix and Svolik 2013; Casper and Tyson 2014; Shadmehr 2015; Guriev and Treisman 2015; Ananyev et al. 2015; Egorov and Sonin 2015).[9] We abstract from the deterrence effect of repression that has been extensively studied in the literature. Instead, we focus on the informational aspect of repression. In our model, repression reveals information about both the intentions of the state and the nature of the activists' demands.

The structure of our model, *without commitment*, has surface similarities with the literature that studies the interactions between informed policy-makers and uninformed voters in democratic settings (Canes-Wrone et al. 2001; Levy 2004; Maskin and Tirole 2004; Fox 2007; Fox and Stephenson 2011). In this literature, the voters' inferences about the state of the world are irrelevant because the voters cannot change the politician's current policy, and hence, the voters' decisions only hinge on their updated beliefs about the politician's type. In contrast, in our model, the authority to make the final decision is shared between the ruler and the bystander citizen because the citizen can depose the ruler and overturn the policy by joining the activist. Therefore, updating about the activist's type enters into the bystander's calculations, complicating the strategic interactions. For example, holding the bad ruler's strategy fixed, the strategic interactions of the bystander and the good ruler feature strategic complementarity, generating the potential for multiple equilibria. More broadly, the focus of this literature is on pandering of politicians to voters' opinions (Canes-Wrone

[9] Siegel (2011) investigates how the structure of social networks determines whether and when citizens' anger and fear of repression leads to backlash protests—see also Shadmehr (2014, Appendix A). In Guriev and Treisman (2015), a dictator's repression of the elite, when observed, immediately reveals his incompetence and leads to his removal.

et al. 2001),[10] politicians' anti-herding behavior (Levy 2004), optimal institutional design (Maskin and Tirole 2004; Fox and Stephenson 2011), or government transparency (Fox 2007). In contrast, our focus is to develop an informational explanation for the spread of protest following repression, and to investigate the effects of institutions that limit repression.

A more distant literature studies the interactions between an uninformed principal and an informed expert with an uncertain bias in a cheap talk context (Sobel 1985; Morris 2001). The expert sends a cheap talk message to the principal who then decides which action to take. From a modeling perspective, our paper departs from this literature in several aspects. Unlike the cheap talk literature in which advice is free, in our model, different actions for the government are costly. Indeed, the government's cost of concession depends both on its own type and the activist's. Moreover, in contrast to the cheap talk literature in which the authority over actions rests solely with the decision maker, in our model both the ruler and the bystander share authority over the policy that is implemented. In particular, the ruler can concede to the activists, which is costly but always allows him to remain in power. However, if the ruler represses, the bystander can join the protest, topple the regime and change the policy. To our knowledge, our analysis of commitment is completely novel in the literature: we identify situations in which the bad type (government) can commit to a repression strategy without revealing its type, and characterize how this commitment power complicates the equilibria (e.g., requiring equilibrium refinements) and influences the outcome.

2 MODEL

We consider a game with two strategic players: a ruler and a bystander citizen. In addition, there is a non-strategic activist who protests, demanding that the ruler implements a set of social changes. The activist is one of two possible types: a "good" activist (type g) proposes reforms that would benefit the bystander (relative to the status quo) if implemented, while reforms proposed by a "bad" activist (type b) would hurt the bystander. The ruler is also of two possible types, "good" (G) or "bad" (B). Like the bystander, a good ruler prefers good reforms over the status quo and prefers the status quo over bad reforms. The bad ruler prefers good reforms to bad reforms but prefers the status quo to reforms of either type. Both types of ruler derive a private benefit from being in office. The ruler observes the activist's type, but the bystander does not observe either the type of the ruler or the type of the activist. Under the common prior, the ruler is bad with probability $p \in (0,1)$, the activist is bad with probability $q \in (0,1)$.

[10] See also Acemoglu et al. (2013) in which politicians choose populist policies to signal they have not been captured.

The game proceeds in multiple stages. In stage zero, nature chooses the ruler's type and the activist's type. The ruler observes whether the activist's demands are harmful or beneficial to the bystander, i.e., the ruler observes the activist's type. The ruler then decides whether to concede to the activist or repress him. If the ruler concedes to the activist, the game ends. If the ruler represses, the bystander citizen decides whether to protest.[11] When the bystander protests, the activist's reform is implemented and the current ruler is removed from office; otherwise no reform is implemented and the ruler retains power. Payoffs are realized at the end of the game.

If no reform takes place (the ruler repressed the activist and the bystander did not join the protest) the bystander's payoff depends on the ruler's type: the bystander prefers to live under a good ruler than a bad one. Thus, in the absence of reform, the bystander's payoff under the good ruler is normalized to zero, and under the bad ruler to $-\beta < 0$. If a reform does take place—either because the ruler concedes or because the ruler is deposed—the bystander's payoff depends on the activist's type. If the activist is bad, the bystander's payoff is $-\beta_b$; if the activist is good, the bystander's payoff is $\beta_g > 0$. We focus on the case in which implementing a bad reform is worse for the bystander than no reform under a bad ruler, $-\beta_b < -\beta$. That is, protesting always has a downside: the bystander's payoff is reduced if he supports the bad activist by protesting against the ruler—even when the ruler is bad. Moreover, protesting against a good ruler has more potential downside than protesting against a bad ruler.[12] When the ruler is bad, the potential downside is $(\beta_b - \beta)$; when the ruler is good, the potential downside is β_b.[13]

The ruler's payoff depends on whether or not he retains office, the terminal policy, and his type. If a ruler is deposed (no matter by whom) his payoff is normalized to zero. If the ruler maintains office without implementing any reforms, the ruler receives an office rent of 1. A good ruler prefers a good reform to the status quo, but prefers the status quo to a bad reform. Hence, if the good ruler retains office by conceding to the good activist's demands, his terminal payoff is $1 + \delta_g$; if he concedes to a bad activist, his payoff is $1 - \delta_b$, where $0 < \delta_i$ and $\delta_b < 1$. Meanwhile, a bad ruler

[11] In our model the bystander always learns that the ruler repressed the activist. The analysis could be routinely extended to incorporate censorship, whereby the bystander observes repression with some probability less than one. In essence, censorship reduces the regime's expected cost of repressing the activist. Provided that the degree of censorship is not so large that the regime always represses, our results extend with no substantive changes.

[12] This single crossing is the key feature of the payoffs structure; the specification that the bystander citizen's payoff falls to $-\beta_b$ when she supports a bad activist independent of the ruler's type serves to simplify the analysis.

[13] One interpretation is that the bystander's payoff depends on the type of policy and on the type of ruler who is in power at the end of the game. If good activists are also good rulers and bad activists are also bad rulers, then supporting a good activist generates a beneficial reform, worth β_g, and leaves a good ruler in power. The payoff of living under a good ruler is normalized to zero, and hence, the payoff of supporting a good activist is β_g. Meanwhile, supporting a bad activist generates a harmful social change, reducing the bystander's by payoff by $\beta_b - \beta$, and leaves a bad ruler in power, reducing the bystander's payoff by an β, for a net reduction of $-(\beta_b - \beta) - \beta = -\beta_b$.

prefers the status quo to either reform, but prefers good reforms to bad reforms. Hence, if the bad ruler concedes to the good activist, his payoff is $1-\alpha_g$; if he concedes to the bad activist, his payoff is $1-\alpha_b$, where $0<\alpha_g<\alpha_b<1$.

In section 5, we modify the model to investigate the effects of the bad ruler's ability to observably commit to a repression strategy. In an Online Appendix, we study the effects of the bystander citizen's commitment to a protest strategy.

2.1 DISCUSSION OF THE MODEL

We discuss some of our modeling choices before we present our analysis and results.

Information structure. We assume that the ruler is better informed than the bystander about the activist: the ruler observes the activist's type, but the bystander can only make inferences about it. This assumption is based on two observations. First, the rulers have more resources (e.g., intelligence agencies) to gather and process information about the goals and preferences of the activists and the general public. Second, bystander citizens have difficulties in learning the activists' types since rulers who use repression claim that they do so to protect their citizens against harmful dissidents. For example, in protests proceeding the April 2013 Venezuelan presidential election, officials called the protesters "the reactionary, criminal and murderous right wing that is run by Henrique Capriles" (Vyas and Gonzalez 2013).[14]

When protests broke out in February 2014 in Venezuela, Delcy Rodriguez, the minister of information stated that the protesters "are not students, they are violent gangs. They are executing a plan with the goal of a civil war in Venezuela" (Minaya 2014a). "Mr. Maduro accused what he called 'fascist leaders' financed by the U.S. of using highly trained teams to topple his socialist government from power...he charged that the demonstrators were trying 'to fill the country with violence and to create a spiral of hatred among our people.' He said his foes were hoping to generate chaos to justify a foreign military intervention. 'In Venezuela, they're applying the format of a coup d'état,' he said" (Minaya 2014b). To some citizens, these statements are dictators' cliches, but they are believable arguments to others. For example, "Danny Ojeda, 44, who works in distant Guarico state but came with other pro-government workers, said he agreed with the late president's [Chavez's] assertions that the U.S. government was behind Venezuela's troubles, a claim also made by Mr. Maduro. 'There's an economic coup against our country, but Maduro has the support

[14]Similar statements were made by Iranian officials about the supporters of Mousavi who were protesting against election fraud following the 2009 Iranian presidential election.

to overcome it,' he said" (Forero 2014). Faced with opposing claims from the government and activists, many ordinary citizens remain uncertain about whom to believe and support.

Indeed, sometimes the rulers' statements contain some truth. In June 1975, a group of seminary students gathered in Fayzyah seminary school in Qom, Iran, demonstrating against the Pahlavi regime and praising Khomeini who was in exile. The regime responded with repression, beating and arresting the protesters. The Shah argued that the protest was the result of "the unholy alliance of black reactionist[s] and stateless Reds" (Kurzman 2003, p. 289). By "black reactionists" the Shah meant religious fanatics whose goal was to establish a theocracy, and by "state-less reds" he meant Soviet-backed communists who wanted to establish a communist state. We now know that they were not communists, but a subset of those protesters did want to establish a theocratic state in Iran although they were not explicit about it then.

Why no protest following concessions? We view the bystander citizen as a follower who may join an already existing protest, but does not initiate one. When the ruler concedes to the activist's demands in an early stage, they ends their movement. This view is consistent with the robust finding in the social movements literature that sustaining protest activities take significant resources and planning that require activists (McAdam 1999; McCarthy and Zald 1973, 1977; Gamson 1975; Tilly 1978, 1996, 2004; Tarrow 1998; McAdam, Tarrow, and Tilly 2001). Thus, when these activists' demands are met, the movement ends. Spontaneous protests occasionally occur, but they typically end quickly without any policy change. Indeed, many *seemingly* spontaneous movements are based on complex networks of organizations and committed activists (Morris 1984; Diani and McAdam 2003; Khatib and Lust 2014).

Repression by the good ruler. Like most models of contentious politics, in our model, the state (regime or ruler) can either repress the activists or concede to them, e.g., redistribution or democratization in Acemoglu and Robinson's (2001, 2006) framework or Boix's (2003). That is, a protest in these models is not an obscure demonstration on a street corner that the state can simply ignore. Rather, protest refers to a movement that is strong enough to achieve its goals unless stopped by coercive actions.

With this notion of protest in mind, consider a minority of religious activists who protest against the state and attempt to implement religious laws against the preferences of the majority; or a fascist group who protests and attempts to impose its racist views on the society. A good ruler prevents this small minority from imposing their preferences on the majority by using coercive means, e.g., arresting and imprisoning the activists. This does *not* imply that the good government

prosecutes this group for their views. Rather, repression in this context means that the government uses its coercive means to prevent these groups from imposing their religious laws or supremacist views on the majority.

Non-strategic Activist. We have assumed that the activist always protests to simplify the exposition. Our results readily extend when the activist's payoffs from receiving his demands are sufficiently high. Then, the activist always decides to protests in the low and intermediate repression equilibria in which the ruler may concede or the bystander citizen may join the activist, and hence these equilibria and their characteristics remain unchanged.[15] Moreover, our modeling choice of having the activist always protest also reflects the observation that some activists protest even when they are sure to be repressed as a matter or principle or as a strategy to receive publicity and raise future recruitment.

Direct Costs of Protest and Repression. We have abstracted from direct protest and repression costs to simplify analysis. Clearly, adding known costs does not change our results qualitatively. When costs are private knowledge, equilibria are in pure strategies, however, the multiplicity of equilibria, which is inherent in the nature of the game, remains. Importantly, the same tradeoffs and strategic forces arise with private costs, while the analysis becomes significantly more cumbersome.

3 PRELIMINARY ANALYSIS

Strategies. If the ruler represses the activist, the bystander must decide whether to join the protest. The bystander strategy is a probability, $\pi \in [0,1]$, representing the probability of joining the protest. The ruler's strategy is a quadruple, $(\rho_g^G, \rho_b^G, \rho_g^B, \rho_b^B) \in [0,1]^4$, where ρ_j^i is the probability with which the type $i \in \{G, B\}$ ruler represses the type $j \in \{g, b\}$ activist.

Protest Strategy. When deciding whether to join the protest, the bystander faces a tradeoff: support the ruler and possibly prevent the implementation of beneficial changes, or support the activist and risk the implementation of bad changes. The bystander's decision depends on her (updated) belief that the activist is bad, q', and on her (updated) belief that the ruler is bad, p'. The bystander's expected payoff from protesting is $\beta_g(1-q') - \beta_b q'$: with probability q' the activist is bad, and the bystander receives $-\beta_b$; with the remaining probability $1-q'$ the activist is good, and the bystander receives β_g. If the bystander does not protest, her expected payoff is $-\beta p'$:

[15] The behavior of the ruler and the bystander in the high repression equilibrium does not change, although the activist decides never to protest in that equilibrium. Of course, none of our results are based on the existence or characteristics of the high repression equilibrium.

with probability p' the ruler is bad, and the bystander receives $-\beta$; with the remaining probability the ruler is good, and the bystander receives her payoff that is normalized to 0. Therefore, the bystander's equilibrium strategy is:

$$\pi = \begin{cases} 1 & \text{if} \quad \beta_g(1-q') - \beta_b q' > -\beta p' \\ [0,1] & \text{if} \quad \beta_g(1-q') - \beta_b q' = -\beta p' \\ 0 & \text{if} \quad \beta_g(1-q') - \beta_b q' < -\beta p', \end{cases} \tag{1}$$

where the bystander's updated beliefs p' and q' depend on the ruler's strategy in equilibrium.

Repression Strategy. If the bystander never joined the protest, a good ruler would repress a bad activist, and a bad ruler would repress either type of the activist. What complicates the ruler's decision is that the bystander may join the activist's protest following repression, in which case he is removed from office. When the ruler represses, he is deposed whenever the bystander joins the protest, which happens with probability π. Therefore, the ruler's expected payoff from repression is $1-\pi$. However, if the ruler concedes, his payoff depends on both his type and the activist's type. Because the good ruler prefers a good reform to the status quo ($\delta_g > 0$), he always concedes to the good activist, $\rho_g^G = 0$. Recall that the good ruler's payoff of conceding to a bad activist is $1-\delta_b$, and the bad ruler's payoff of conceding to a type $i \in \{g, b\}$ activist is $1-\alpha_i$. Therefore, the ruler's equilibrium strategy is

$$\rho_b^G = \begin{cases} 1 & \text{if} \quad \pi < \delta_b \\ [0,1] & \text{if} \quad \pi = \delta_b \\ 0 & \text{if} \quad \pi > \delta_b. \end{cases} \quad \rho_i^B = \begin{cases} 1 & \text{if} \quad \pi < \alpha_i \\ [0,1] & \text{if} \quad \pi = \alpha_i \\ 0 & \text{if} \quad \pi > \alpha_i. \end{cases} \tag{2}$$

The ruler's strategy weighs the benefit of repressing reforms that he does not favor with the cost of inciting protest by the bystander.

Remark. There is always an equilibrium in which no repression takes place ($\rho_j^i = 0$). In any such equilibrium, the bystander joins the protest with sufficiently high probability upon observing repression, but this information set is off-the-equilibrium path. This behavior deters the ruler from repressing, but it is supported by the bystander's off-the-equilibrium-path beliefs that if the ruler represses, then with a high probability he must be a bad ruler repressing a good activist. However, this equilibrium does not satisfy the D1 criterion for equilibrium selection (Fudenberg and Tirole 2000, p. 452).[16] To see why, note that the bad ruler dislikes conceding to the bad activist more

[16] Proof in the Appendix.

than conceding to a good one, and the bad ruler is therefore more inclined to repress the bad activist. Hence, the D1 restriction on off-the-path beliefs rules out the possibility that the bystander believes that the bad ruler may have repressed the good activist. We therefore consider equilibria in which repression takes place with a positive probability.

If the bad ruler never repressed the good activist, then repression would imply that the activist must be the bad type (because the good type would never be repressed). Consequently, the bystander would never join the protest. However, if the bystander never joins the protest, then the bad ruler would deviate by repressing the good activist. Therefore, in any equilibrium the bad ruler ruler represses the good activist with positive probability ($\rho_g^B > 0$). Because the bad ruler dislikes the bad activist more than the good activist ($\alpha_b > \alpha_g$), if the bad ruler represses the good activist with positive probability, then he always represses the bad activist ($\rho_b^B = 1$).

Lemma 1 *In equilibrium, the good ruler never represses the good activist, but the bad ruler represses the good activist with a positive probability, and always represses the bad activist:* $\rho_g^G = 0$, $\rho_g^B > 0$, $\rho_b^B = 1$. *Moreover, repression never fully reveals the activist's type.*

The lemma reveals that whenever the type of ruler and activist match, the equilibrium behavior of the ruler is in line with the bystander's ideal: $\rho_g^G = 0$ and $\rho_b^B = 1$. It also establishes that a distortion from this ideal strategy is a part of every equilibrium: the bad ruler represses the good activist with a positive probability, $\rho_g^B > 0$. Because the bad ruler represses both types of activist with positive probability, the bystander can never be certain that a repressed activist is good.

4 EQUILIBRIUM AND REPRESSION BACKFIRE

If the bystander knew the types of the ruler and activist, she would join the protest if and only if the bad ruler was repressing a good activist. However, when deciding whether to join the protest, the bystander is unsure whether the good ruler repressed the bad activist, the bad ruler repressed the bad activist, or the bad ruler repressed the good activist. Thus, she faces a tradeoff: by supporting the ruler, she risks blocking beneficial changes, but by supporting the activist, she risks implementing harmful changes. To resolve this uncertainty-induced tradeoff, the bystander uses all the available information to update her beliefs about the ruler and the activist. In particular, using Bayes' Rule:

$$p' = \Pr(\text{bad ruler}|\text{repression}) = \frac{\Pr(\text{ repression} \cap \text{bad ruler})}{\Pr(\text{ repression})}.$$

The bad ruler always represses the bad activist and represses the good activist with probability ρ_g^B. The probability of repression and a bad ruler is therefore $p(q+(1-q)\rho_g^B)$. The bad ruler is not the only one who represses; the good ruler also represses the bad activist with probability ρ_b^G. Thus, the probability of repression is $p(q+(1-q)\rho_g^B)+(1-p)q\rho_b^G$. Similar calculations for updating beliefs about the activist show:

$$p' = \frac{p[q+(1-q)\rho_g^B]}{p[q+(1-q)\rho_g^B]+(1-p)q\rho_b^G} \qquad q' = \frac{q[(1-p)\rho_b^G+p]}{q[(1-p)\rho_b^G+p]+p(1-q)\rho_g^B}. \tag{3}$$

Proposition 1 highlights the key aspects of the bystander's updating.[17]

Proposition 1 *In any equilibrium, repression causes the bystander to update negatively about <u>both</u> the ruler and the activist: $q' > q$ and $p' > p$. Moreover, holding the bad ruler's strategy fixed, when the good ruler represses the bad activist more often, the bystander updates less negatively about the ruler and more negatively about the activist: $\frac{\partial p'}{\partial \rho_b^G} < 0 < \frac{\partial q'}{\partial \rho_b^G}$. In contrast, when the bad ruler represses the good activist more often (given the good ruler's strategy), the bystander updates are the opposite: $\frac{\partial q'}{\partial \rho_g^B} < 0 < \frac{\partial p'}{\partial \rho_g^B}$.*

Proposition 1 has two implications. When the bad ruler represses the good activist more often, the bystander's incentives to protest increase. However, when the good ruler represses the bad activist more often, the bystander's incentives to protest *fall*. Of course, as the bystander protests more often, the incentives of both the good and the bad ruler fall. These underlying forces drive equilibrium behavior.[18]

Proposition 2 *Suppose $\delta_b < \alpha_g$. In equilibrium, the good ruler never represses a good activist and the bad ruler always represses a bad activist, $\rho_g^G = 0$ and $\rho_b^B = 1$. There exists an increasing curve $q_1(p) \equiv \frac{(\beta+\beta_g)p}{\beta_b+\beta_g p}$ and a constant $q_2 \equiv \frac{\beta+\beta_g}{\beta_b+\beta_g}$ with $0 < q_1(p) < q_2 < 1$, such that:*

- ***Low Repression Equilibrium:*** *When the prior likelihood that the activist is bad is low, $q < q_1(p)$, a unique equilibrium exists. The good ruler never represses the bad activist, the bad ruler represses the good activist with a positive probability less than one, and upon observing repression, the bystander protests with a positive probability less than one: $\rho_b^G = 0$, $\rho_g^B = \frac{\beta_b-\beta}{\beta_g+\beta}\frac{q}{1-q}$, and $\pi = \alpha_g$.*
- ***High Repression Equilibrium:*** *When the prior likelihood that the activist is bad is high, $q > q_2$, a unique equilibrium exists. The good ruler always represses the bad activist, $\rho_b^G = 1$,*

[17] The proof follows from (3) by simple algebraic manipulations and differentiations, and hence is omitted.
[18] Proposition 2 assumes $\delta_b < \alpha_g$. The case of $\delta_b > \alpha_g$ is presented in Lemma 3 in the Appendix B.

the bad ruler always represses the good activist, $\rho_g^B = 1$, and the bystander never protests upon observing repression, $\pi = 0$.

- ***Intermediate Repression Equilibrium:*** *When $q_1(p) \leq q \leq q_2$, the high repression and low repression equilibrium described above both exist. In addition, an equilibrium exists in which the good ruler represses the bad activist with a positive probability less than 1, the bad ruler always represses the good activist, and upon observing repression, the bystander protests with a positive probability less than 1:*

$$\rho_b^G = \frac{p}{1-p}\frac{(\beta+\beta_g)-(\beta_b+\beta_g)q}{\beta_b q},\ \rho_g^B = 1,\ \textit{and}\ \pi = \delta_b.$$

To see the intuition, consider a simplified version of the game in which the good ruler is a non-strategic player who always represses the bad activist, i.e., $\rho_b^G = 1$. The probability with which the bad ruler represses the good activist ρ_g^B is (in equilibrium) decreasing in the bystander's likelihood of protest π. Because the bystander's best response π is increasing in the bad ruler's strategy ρ_g^B, a unique equilibrium exists. Next, consider a different simplified version of the game in which the bad ruler is a non-strategic player who always represses the good activist, i.e., $\rho_g^B = 1$. The good ruler's best response ρ_b^G is decreasing in the bystander's strategy π. However, unlike the previous case, the bystander's best response π is also decreasing in the good ruler's strategy ρ_b^G: when it is more likely that repression was carried out by the good ruler against the bad activist, the bystander has less incentive to protest.[19] This structure of best responses allows for multiple equilibria because best responses can cross multiple times. In fact, it is easy to show that this simplified game has three equilibria. In one equilibrium, the bystander never protests and the good ruler always represses the bad activist. This equilibrium exists whenever $q > q_1(p)$. In another equilibrium, the bystander protests with a higher probability, $\pi = \delta_b$, and the good ruler represses the bad activist with a lower probability, $\rho_b^G = \frac{p}{1-p}\frac{(\beta+\beta_g)-(\beta_b+\beta_g)q}{\beta_b q}$. This equilibrium exists whenever $q_2 > q > q_1(p)$. In the third equilibrium, the bystander always protests and the good ruler never represses the bad activist. The first two equilibria remain even when the bad ruler acts strategically: in both of them $\pi < \alpha_g$, so it is the bad ruler's best response to always repress the good activist. However, the third equilibrium must be modified. If the bystander always protests upon observing repression, then the strategic bad ruler does not repress in equilibrium (while the non-strategic bad ruler always represses). The low repression equilibrium features the same behavior by the good ruler,

[19] Reverse the ordering of the good ruler's strategy to focus on the likelihood that the good ruler concedes to the bad activist: $1 - \rho_b^G$. Then, holding the bad ruler's strategy fixed, the best responses of the good ruler and the bystander, $1 - \rho_b^G(\pi)$ and $\pi(1 - \rho_b^G)$, are both increasing, i.e., they feature strategic complementarity.

but modifies the bad ruler's behavior to account for his strategic response. This logic shows that the multiplicity of equilibria stems from the nature the bystander's updated beliefs and the subtle interactions between the bystander and the good ruler.

The strategic considerations that generate multiple equilibria suggest that social norms play a critical role in the interactions between citizens and the state. By influencing the public's and government's expectations of the other's behavior, social norms determine which equilibrium arises. Corollary 1 establishes that social norms are more salient when the status quo under a bad ruler is worse.

Corollary 1 *When the public expects the status quo under a bad ruler to be worse, the subset of parameters in which multiple equilibria arise is larger:* $\frac{\partial[q_2 - q_1(p)]}{\partial \beta} > 0$.

The social norms that determine which equilibrium arises have important implications not only because the equilibria exhibit different levels of repression, but also because the levels of repression in different equilibria respond differently to changes in the environment. Let $R \equiv p(q\rho_b^B + (1-q)\rho_g^B) + (1-p)(q\rho_b^G + (1-q)\rho_g^G)$ be the ex-ante expected level of repression in equilibrium.

Corollary 2 *When the public expects the status quo under a bad ruler to be worse, the expected level of repression decreases in the low repression equilibrium (*$\frac{\partial R}{\partial \beta} < 0$*), but it increases in the intermediate repression equilibrium (*$\frac{\partial R}{\partial \beta} > 0$*).*

When the status quo under the bad government is worse, the public has more incentive to protest. Thus, to counter this extra incentive to revolt, in equilibrium, either the bad ruler must repress the good activist less, or the good ruler has to repress the bad activist more. In the low repression equilibrium, the former occurs, while in the intermediate repression equilibrium, the latter occurs.

To study the empirical implications of these results, consider the relationship between income and inequality on repression. In particular, let G be the a country's per capita GDP. Under a good regime, the bystander citizen receives G, but a bad regime secretly diverts d for its private consumption, leaving the bystander with $G-d$. When the bystander's utility is concave, the status quo under a bad ruler is less harmful to the bystander: $\beta \equiv u(G) - u(G-d)$ and $\frac{\partial \beta}{\partial G} < 0$. Therefore, increases in income are associated with decreases in β, which increase repression in the low repression equilibrium, but decrease repression in the intermediate repression equilibrium. Alternatively, suppose the level of income inequality is a noisy signal of β, so that higher levels of inequality are associated with

a higher β. Then, a good ruler or a good reform would reduce inequality via redistribution, raising the bystander citizen's payoff from $-\beta$ to 0 and β_g, respectively. Analogously, income per capita can be a noisy signal of β, so that lower levels of income are associated with a higher β. Then, as inequality increases or income decreases, the public, who does not know the ruler's type, believes that the ruler is more likely to be bad (p increases), and that the status quo under the bad ruler is more likely to be worse (β increases). It is easy to see that $\frac{\partial R}{\partial p} = 0$ in the low repression equilibrium, and $\frac{\partial R}{\partial p} > 0$ in the intermediate repression equilibrium. Together with Corollary 2, this implies that as income falls or inequality increases, the level of repression *falls* in the low repression equilibrium, but increases in the intermediate repression equilibrium. Therefore, without knowing which equilibrium is played, an empirical study of the effect of income or inequality on repression is challenging, especially in cross-country analyses, where it is plausible that social norms and the equilibrium change.

These results may provide an explanation for the conflicting empirical findings on the relationships between income or inequality and repression and violence. For example, although some studies have found that higher income per capita reduces the likelihood of repression (Mitchell and McCormick 1988; Henderson 1991; Poe and Tate 1994), others have not found a significant correlation (Gandhi 2008; Conrad and Moore 2010; Shadmehr and Haschke 2015; see Davenport (2007) for a review). Similarly, empirical studies of the effects of inequality on violence are inconclusive (Lichbach 1989; Miller et al. 1977; Midlarsky 1988; Muller 1985, 1986; Muller et al. 1991; Muller et al. and Midlarsky 1989; Muller and Seligson 1987; Muller and Weede 1990, 1994; Weede 1981, 1986, 1987; Wang and Dixon et al. 1993). If we posit that, within a country or a region, social norms change slowly over time and the equilibrium played between the government and the public does not change frequently, then one can infer the equilibrium by observing how repression changes with income or inequality in that country in the near past. For example, when in the preceding few years increases in income (or decreases in inequality) has increased the likelihood of repression, this implies that the country is in the low repression equilibrium. Therefore, to the extent that social norms remain unchanged in this country, one can predict the same relationship between income and repression in near future in this country.

Moreover, in the low repression equilibrium, when the good activist demands more fundamental reforms (i.e. reforms associated with higher α_g and β_g) the probability that the bad ruler concedes and the probability that the bystander joins the protest both increase. A demand for more fundamental reforms directly increases the bad ruler's incentive to repress and the bystander's incentive to support the activist. In addition, to dissuade the bad ruler from repressing the good activist

all the time, the bystander increases the likelihood of protest upon observing repression, and because the bystander protests more often, the bad ruler concedes more often to avoid being deposed. Because the low repression equilibrium is unique when $q < q_1(p)$, this analysis suggests:

Corollary 3 *In the low repression equilibrium, as the good activist demands more fundamental reforms, the bad ruler is more likely to concede, and the bystander is more likely to join the protest if the bad ruler represses.*

Proposition 2 also highlights that the bad ruler hurts the bystander in two distinct ways. When he is in power, he blocks beneficial social change. But even when he is not in power, the possibility that the ruler could be bad distorts the good ruler's behavior. Because the bystander is uncertain about the ruler's type, she sometimes protests upon observing repression, which, in turn, reduces the good ruler's incentive to repress the bad activist. Therefore, in the low and intermediate repression equilibrium, the good ruler sometimes concedes to the bad activist, which would never happen if the bystander is certain that he is the good ruler—in particular, if no bad ruler existed.

Repression Backfire. In the low and intermediate repression equilibrium, repression is sometimes followed by the bystander's protest. However, this does not necessarily imply that repression has backfired. The bystander's decision to protest may not be caused by repression because she may have been willing to join the activist's protest even without observing repression, based only on her prior beliefs. Thus, for repression to backfire, the information it conveys must be pivotal in changing the bystander's behavior, inducing her to protest when she otherwise would not.

Definition 4.1 (Repression Backfire) *We say repression backfire occurs in equilibrium when three conditions are satisfied: (1) the ruler sometimes represses the activist ($\rho_i^j > 0$ for some $i \in \{g, b\}$, $j \in \{G, B\}$), (2) the bystander sometimes protests upon observing repression ($\pi > 0$), and (3) given the prior beliefs about the ruler and activist the bystander would not want to protest ($\beta_g(1-q) - \beta_b q < -\beta p$).*

This notion of repression backfire is central in studying the causal link between repression and the stability of regimes. It allows us to separate regimes that become unstable *because* of the state's repression of dissent from regimes that are unstable for reasons other than the state's repression.[20]

[20]The reverse case in which repression conveys information that deters protest also arises in our model. For example, when $q < q_1(p) < q_b(p)$, the bystander would be willing to always join the activist's protest under the prior, while he only sometime protests following repression, i.e., $\pi < 1$ in equilibrium. This is a novel mechanism by which repression can stabilize the regime. However, the notion that repression can deter further protest has already been explored in the literature. Therefore, we focus on repression backfire, as this concept is novel in our paper.

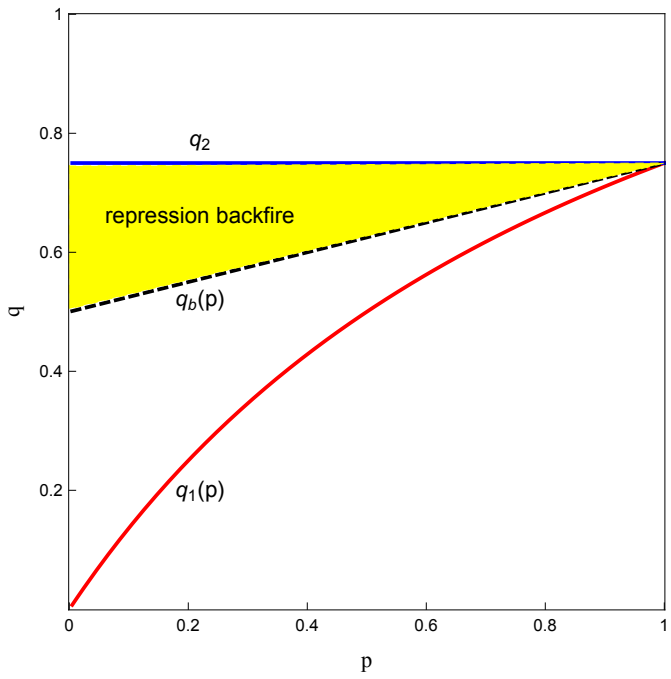

Figure 1: The high repression equilibrium exists when $q > q_1(p)$, the low repression equilibrium exists when $q < q_2$, and the intermediate repression equilibrium exists when $q \in (q_1(p), q_2)$.

Under the prior, the bystander is willing to join the activist's protest whenever the activist is relatively likely to be good ($\beta_g(1-q) - \beta_b q < -\beta p$). Thus, for repression backfire to occur, the bystander must be initially pessimistic about the activist (relative to the ruler):

$$q > q_b(p) \equiv \frac{\beta_g + \beta p}{\beta_g + \beta_b}.$$

Moreover, because there is no repression backfire without the bystander's protest, the high repression equilibrium does not feature repression backfire. The following proposition characterizes whether and when repression backfire occurs in equilibrium—see Figure 1.

Proposition 3 *Repression backfire does not occur in any equilibrium if the good ruler dislikes the bad activist more than the bad ruler dislikes the good activist ($\delta_b > \alpha_g$) or the bystander's payoff is unaffected by the ruler's type in the absence of reform ($\beta = 0$).*[21] *Otherwise ($\delta_b < \alpha_g$ and $\beta > 0$), repression backfire occurs in the low and intermediate repression equilibrium if and only if q is large, but not too large ($q_b(p) < q < q_2$).*

[21] The condition $\beta = 0$ implies that the potential upside and downside of protesting against the bad ruler and the good ruler are the same. In contrast, $\beta > 0$ implies that protesting against the bad ruler has more upside relative to downside in comparison to protesting against the good ruler.

The first part of the proposition considers repression backfire when $\delta_b < \alpha_g$, which underlies our equilibrium characterization in Proposition 2. As we discussed above, when the bystander is initially too optimistic about the activist ($q < q_b(p)$), she is willing to protest under the prior, and hence repression is not pivotal in the bystander's decision to protest. Repression backfire also does not occur when the bystander is too pessimistic about the activist ($q > q_2$), because then she never protests in equilibrium. The question then becomes whether there exists a range of prior beliefs in which the bystander is sufficiently pessimistic about the activist not to protest under the prior, but not so pessimistic that she would not protest upon observing repression. Proposition 3 shows that the answer hinges on whether the the bystander cares about the ruler's type in the absence of reform, i.e., whether $\beta = 0$ or $\beta > 0$. When $\beta = 0$, the bystander only cares about the behavior of the ruler toward the activist, not the ruler's type per se. That is, only the likelihood that the bystander is good or bad (q') enters the bystander's decision to revolt—p' is irrelevant. Because repression always makes the bystander more pessimistic about the activist ($q' > q$), if she is unwilling to protest under the prior, she will also be unwilling to protest after repression. When $\beta > 0$ the bystander cares about both the activist and ruler types, and the negative information repression reveals about both is sometimes pivotal in the bystander's decision to support the activist against the regime.

Corollary 4 *The worse the status quo under the bad ruler, the larger the subset of parameters in which repression may backfire:* $\frac{\partial[q_2 - q_b(p)]}{\partial \beta} > 0$.

When the good ruler dislikes the bad activist more than the bad ruler dislikes the good activist ($\delta_b > \alpha_g$), again, no repression backfire occurs—Lemma 3 in the Appendix B characterizes the equilibrium for this case. When $\delta_b > \alpha_g$, the good ruler always represses the bad activist in equilibrium.[22] This makes repression less of a negative signal about the ruler and more of a negative signal about the activist. This effect is sufficiently strong that whenever the bystander is not willing to protest given her prior beliefs, she will not be willing to protest after observing repression. In fact, we prove a stronger result that highlights the role of the good ruler's (lack of) repression of the bad activist in causing repression backfire.

Proposition 4 *Suppose that the good ruler is a behavioral player who represses the bad activist with probability* $\rho_b^G = r$ *and does not repress the good activist. There exists a threshold,* $R(p) < 1$*, such that for* $r > R$ *no repression backfire occurs in equilibrium.*

[22] From Lemma 3, $\rho_g^B > 0$ in any equilibrium. Moreover, when $\delta_b > \alpha_g$, if the bad ruler represses the good activist with positive probability, $\rho_g^B > 0$, then the good ruler represses the bad activist for certain (see condition (2)).

Proposition 4 implies that when the good ruler is an idealist who always concedes to the good activist and always represses the bad activist, repression does not backfire—because observing repression is never pivotal in the bystander's decision to protest. But in our model, the good ruler is not an idealist, and although his heart is in the right place (he likes to repress the bad activist), his selfish interest in remaining in office makes him concede to the bad activist to prevent the bystander's protest, which would depose him. That is, the good ruler's fear of being "mistaken" for a bad ruler and being deposed by the bystander is essential for repression backfire to occur in equilibrium.

5 BAD RULER'S COMMITMENT

Our analysis so far presumes that bad governments decide their repression strategies after observing the activists' types. However, if a bad government could observably commit to a repression strategy (without revealing its type), it could potentially gain by manipulating the public's beliefs, so that they update less negatively about the government following repression. We investigate how such a commitment alters the strategic considerations of the government and the public, and its effects on the likelihoods of repression and protest.

To illustrate the settings that we model, consider a regime in which both good politicians (doves) and bad politicians (hawks) participate in the government, but only one group has the real authority at a time. When the bad politicians are in power, they set up observable institutions (e.g., an independent judiciary) that limit how often good activists can be repressed. When power shifts to the doves behind the scenes, they decide whether to repress activists, but they cannot reform these fundamental institutions. The public recognizes that the government's ability to repress good activists is limited, but, upon observing repression, it cannot distinguish which type of politician is behind it.[23]

To investigate the effect of this commitment power, we consider an alternative timing of the game. The game begins with the bad ruler in power. The bad ruler moves first, choosing a strategy (r_b, r_g), where r_i is the probability with which he will repress the type i activist, $i \in \{b, g\}$, whenever such an activist appears. This choice is observable to all players. Next, nature moves, replacing the bad ruler with a good ruler with probability $1-p$. With the remaining probability, p, the bad ruler remains in power. The bystander knows these probabilities, but he does not observe

[23]Similar goals could also be achieved by installing an independent militia or security force, e.g., *Basij* militia in Iran, which is also known as "mobilization resistance force," and operates under the supervision of "The Organization for Mobilization of the Oppressed." The size and funding of these forces limit their capacity to repress good activists. If the good politicians grab de facto power, they can choose not to deploy these forces against the activist, but they cannot dismantle them entirely without provoking a coup.

whether the bad ruler has been replaced or remained in power.[24] Next, the activist protests. As in the original game, the activist is bad with probability q and good with probability $1-q$. Then, the ruler responds to the protest. If the ruler is bad, he represses the type i activist with the probability r_i that he had chosen earlier. If the ruler is good, he chooses the probability ρ_i^G with which he represses the type i activist. The rest is the same as our original game. If the ruler concedes to the activist, the game ends. If the ruler represses the activist, the bystander decides whether or not to join the activist's protest. If the bystander does not join the protest, the ruler remains in power. If the bystander does join the protest, the ruler is replaced and the activist's demands are met.

The extensive form of this game has a continuum of subgames, each of which follows a particular choice of (r_b, r_g) by the bad ruler. Because the bad ruler has already committed to (r_b, r_g), the good ruler and the bystander are the strategic players in these subgames. The incentives of the good ruler and the bystander are similar to the original game, except that, with commitment, they observe the bad ruler's strategy. Therefore, the bystander's best response is a mapping from the bad ruler's *observed* strategy (r_b, r_g) and the good ruler's *anticipated* strategy ρ_b^G into a protest probability.[25]

Lemma 2 *Given the strategy of the bad ruler* (r_b, r_g) *and the strategy of the good ruler* ρ_b^G*, the bystander's best response is:*

$$\pi(r_b, r_g; \rho_b^G) = \begin{cases} 1 & \text{if} \quad F(r_b, r_g) > K\rho_b^G \\ [0,1] & \text{if} \quad F(r_b, r_g) = K\rho_b^G \\ 0 & \text{if} \quad F(r_b, r_g) < K\rho_b^G \end{cases}$$

where $F(r_b, r_g) \equiv (1-q)(\beta_g + \beta)r_g - q(\beta_b - \beta)r_b$ *and* $K \equiv \frac{1-p}{p} q\beta_b$.

Function $F(r_b, r_g)$ is the net expected payoff from protesting versus not protesting against a bad ruler. With probability qr_b, the bad ruler has repressed a bad activist, and protesting reduces the bystander's payoff by $\beta_b - \beta$. With probability $(1-q)r_g$, the bad ruler has repressed a good activist, and protesting raises the bystander's payoff by $\beta_g + \beta$. When $F(r_b, r_g) > K$, the bystander has a dominant strategy to always protest in the subgame, and hence the good ruler will never repress in the unique equilibrium of the subgame. Similarly, when $F(r_b, r_g) < 0$, the bystander has a dominant strategy to never protest and the good ruler always represses the bad activist.

[24] Leadership is a complex process and power is allocated among a variety of factions with different goals and agendas. While to ordinary citizens it may appear that the same group may be leading the country, power may be shifting between different factions of the leadership behind the scenes.

[25] As in the case without commitment, $\rho_g^G = 0$ in equilibrium because the good ruler prefers good reforms to the status quo.

In contrast, when $0 < F(r_b, r_g) < K$ the bystander's best response depends on the good ruler's strategy, generating the potential for multiple equilibria.

Lemma 5 in the appendix shows that when $0 < F(r_b, r_g) < K$, the subgame has three equilibria: in one the bystander does not protest, in another the bystander always protests, and in the third the bystander protests with probability δ_b. This creates the possibility that the bystander and the good ruler switch from one equilibrium to another following each bad ruler's choice of (r_b, r_g). We impose the natural restriction that, for any (r_b, r_g) such that $0 < F(r_b, r_g) < K$, only one of the three equilibria is played in the subgame. Propositions 9 and 10 in the Online Appendix analyze the cases in which the protest probability is zero and one whenever $F(r_b, r_g) \in (0, K)$, establishing that no protest takes place in equilibrium (and no repression backfire occurs). In the text, we focus on the more interesting case in which the protest probability is δ_b whenever $F(r_b, r_g) \in (0, K)$. Therefore, depending on the bad ruler's strategy, the bystander's equilibrium protest probability takes one of the three values of 0, δ_b, or 1. As the bad ruler represses the good activist more, the likelihood that the protest spreads (weakly) increases as $F(r_b, r_g)$ moves from $F < 0$, where the bystander never protests, to $F \in (0, K)$, where she sometimes protests, to $F > K$, where she always protests. Therefore, the bad ruler has a tradeoff between repressing the good activist with a higher probability and facing a higher probability of protest by the bystander.

The bad ruler's ability to commit to a strategy creates another complication. When $F(r_b, r_g) = K$, a continuum of equilibria are possible in the subgame: any $\pi \in [0, \delta_b]$ with $\rho_b^G = 1$ can be part of the equilibrium of the subgame (see Lemma 5 in the Appendix). Without commitment, these equilibria only arise in knife-edge cases (on a set of measure zero in the parameter space). With commitment, however, the bad ruler may select a strategy with $F(r_b, r_g) = K$. Therefore, a refinement is needed to limit the set of possible equilibria. Using a similar logic to the trembling hand refinement, we show that the subgame with $\pi = \delta_b$ is uniquely selected when $F(r_b, r_g) = K$. In particular, we introduce stochastic shocks to the bad ruler's strategy, showing that as the support of the distribution of shocks vanishes, the equilibrium converges to one in which $\pi = \delta_b$ (see Proposition 7 in the Appendix). Figure 2 illustrates the four equilibrium regions that arise, and Proposition 5 formally describes the equilibrium.

Proposition 5 *In equilibrium,*

1. *If $q_2 < q$, then the strategies are identical to the high repression equilibrium (region I).*

2. *If $q_1(p) < q < q^*$, strategies are identical to the intermediate repression equilibrium (region II).*

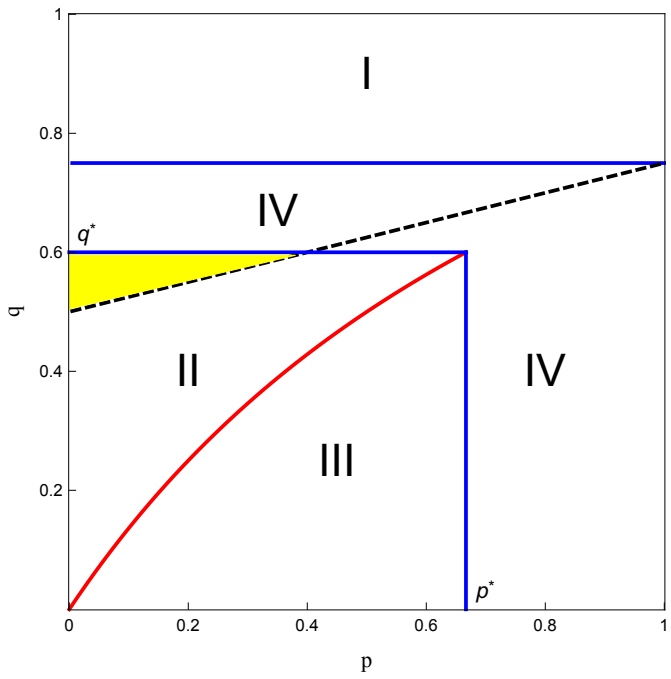

Figure 2: Equilibrium with commitment. Proposition 5 describes the equilibrium. The red curve is $q_1(p)$ and the dashed curve is $q_b(p)$. Repression backfire occurs in the yellow area.

3. *If* $q < q_1(p)$ *and* $p < p^*$, *then* $r_b = \rho_b^G = 1$, $r_g = \frac{q}{1-q}\frac{\beta_b - p\beta}{p(\beta+\beta_g)}$, *and* $\pi = \delta_b$ *(region III).*

4. *Otherwise,* $r_b = \rho_b^G = 1$, $r_g = \frac{q}{1-q}\frac{\beta_b - \beta}{\beta+\beta_g}$, *and* $\pi = 0$ *(region IV).*

Moreover, $q^* = q_2(1 - \frac{\delta_b}{\alpha_g}) < q_2$ *and* $p^* = \frac{\beta_b(\alpha_g - \delta_b)}{\beta_b\alpha_g + \delta_b\beta_g}$.

By repressing the bad activist more, the bad ruler gains directly from saving the concession costs, and indirectly from making the bystander update less negatively about the ruler and more negatively about the activist. Therefore, the bad ruler always represses the bad activist. However, when choosing how much to repress the good activist, the bad ruler faces a tradeoff between the probabilities of repression and protest: increasing r_g can increase the protest probability from zero to δ_b, or from δ_b to 1. Clearly, the bad ruler does not repress so much that the bystander always protests following repression. But he may trade off a lower probability of protest for a higher probability of repression. That is, the bad ruler's equilibrium choice boils down to choosing between *two* thresholds of repressing the good activist. Let $r_g = r_1$ be the low threshold and r_2 be the high one. The bad ruler can choose r_1 and eliminate the bystander's protest ($\pi = 0$), or he can choose r_2 and risk the bystander's protest following repression with probability δ_b.

When $q > q_2$, the likelihood of a bad activist is sufficiently high that even if the bad ruler always represses, the bystander never protests. When $q_1(p) < q < q_2$, if the bad ruler always represses,

then the bystander protests with probability δ_b. Thus, the bad ruler has two possible choices: (1) he can repress both types of activist with probability one ($r_2 = 1$) and face the bystander's protest with probability δ_b, or (2) he can limit the repression of the good activist to $r_1 = \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta} < 1$, thereby ensuring that the bystander does not protest. When the activist is relatively likely to be good ($q < q^* = q_2(1 - \frac{\delta_b}{\alpha_g})$), repressing the good activist with a higher probability is more valuable to the bad ruler, and he chooses the first option (region II). Otherwise, $q^* < q < q_2$, and he chooses the second option (in region IV).

In contrast, when $q < q_1(p)$, the likelihood of a bad activist is small enough that if the ruler were to always repress, the bystander would protest with probability one. Therefore, the bad ruler's equilibrium choices boil down to two: (1) If he represses the good activist with a smaller probability, $r_1 = \frac{q}{1-q}\frac{\beta_b-\beta}{\beta+\beta_g}$, then the bystander does not protest in equilibrium. (2) If he represses the good activist with a larger probability $r_2 = \frac{q}{1-q}\frac{\beta_b-p\beta}{p(\beta+\beta_g)}$, then the bystander protests with probability δ_b—the ruler cannot repress more often without inducing $\pi = 1$. As p increases, the bystander believes that the ruler is more likely to be bad. As a result, she becomes more inclined to join the activist's protest, limiting the ruler's ability to repress the good activist: $r_2(p)$ is decreasing in p, falling from $r_2 = 1$ on the curve $q_1(p)$ to $r_2(p) = \frac{q}{1-q}\frac{\beta_b-\beta}{\beta+\beta_g}$ at $p = 1$. Therefore, when p is large (region IV), the ruler chooses option (1) in which the bystander does not protest. When p is small (region III), $r_2(p)$ is sufficiently large that the gain from raising the repression probability to $r_2(p)$ offsets the increase in the protest probability from zero to δ_b. Hence, for small p, the rulers prefers option (2).

Corollary 5 *In equilibrium, the bystander does not protest upon repression if and only if the probability that the ruler is bad or the probability that the activist is bad is sufficiently large. There exists (p^*, q^*) such that $\pi = 0$ in equilibrium if and only if $q > q^*$ or $p > p^*$.*

The Role of Commitment. Proposition 5 shows that the bad ruler uses his ability to limit its repression of good activists in regions III and IV. To understand the effects of this commitment power, we analyze ruler's equilibrium behavior varies with q, focusing on the interesting case when $p < p^*$. Figure 3 illustrates. Increases in q has two conflicting effects: (1) it reduces the bad ruler's ex ante incentives to repress the good activist because the activist is less likely to be good, and (2) it reduces the bystander's incentives to protest following repression because it would be (i) less likely that a good activist is repressed, and (ii) more likely that the bystander would be supporting a bad activist.[26]

[26] Although (i) and (ii) seem to be the flip sides of the same coin, the bystander incurs the associated costs of (ii) *only* if she protests with a positive probability.

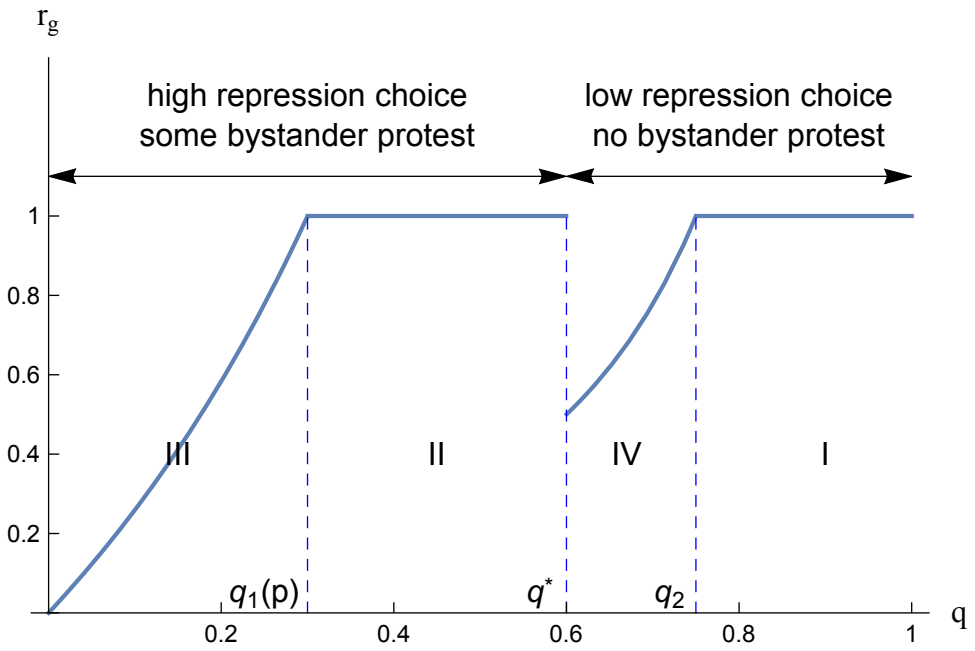

Figure 3: The bad ruler's equilibrium likelihood of repressing the good activist as a function of the prior likelihood q that the activist is bad—for a given $p < p^*$.

Recall that the bad ruler's equilibrium choices are effectively between a low likelihood r_1 and a high likelihood r_2 of repressing the good activist. When q increases, both r_1 and r_2 rise (until r_2 reaches 1), but r_2 rises faster so that $r_2 - r_1$ also increases. The reason that r_2 rises faster is that when $r_g = r_1$, the bystander does not protest ($\pi = 0$), and hence the reduction in the likelihood that her protest can lead to bad reforms, effect (ii), is irrelevant. In contrast, when $r_g = r_2$, the bystander protests with probability $\pi = \delta_d$, and hence both (i) and (ii) contribute to reduce her incentives to protest, and hence to raise r_2.

When q is small (region III), so that the good activist is relatively likely, both $(r_2 - r_1)$ and the ex ante value of repressing the good activist are sufficiently large that it is worth it for the bad ruler to risk protest, and choose the high repression threshold. As q increases, as far as $r_2 < 1$, $r_2 - r_1$ rises in region III until r_2 reaches 1 at the boundary of region II, $q = q_1(p)$. Now, because $r_2 = 1$ and cannot rise any further, increases in q reduce $r_2 - r_1$ until the bad ruler's gains of raising repression from r_1 to $r_2 = 1$ become so small that they are not worth raising the likelihood of the bystander's protest from 0 to δ_b. This happens at the boundary $q = q^*$ of region IV, where the bad ruler switches from the higher threshold to the low threshold $r_1(q)$. From this point on, increases in q keep raising $r_1(q)$ until $r_1 = 1$ at the boundary $q = q_2$ of region I. As Figure 3 shows, when $p < p^*$, both the bad ruler's likelihood of repression and whether or not he ex ante limits his

repression are non-monotone in q. The bad ruler limits its repression when q is low (region III) or high (IV), but not when it is intermediate (region II) or very high (region I).

Corollary 6 *In equilibrium, when $p < p^*$, the bad ruler's likelihood of repressing the good activist is non-monotone in the prior likelihood q that the activist is bad.*

One may think that when the bad ruler limits its repression, he must repress the good activist *less often* in order to gain by manipulating the bystander's equilibrium beliefs, so that they protest less following repression. However, this argument does not take into account the good ruler's response. If commitment helps the bad ruler to support an equilibrium in which the good ruler represses more, then the bystander's incentives to protest following repression falls, reducing the risk of repression for the bad ruler. It is this effect that allows the bad ruler, in region III, to repress *more* than what he would do absent commitment, and yet to face a lower probability of the bystander's protest.[27]

Corollary 7 *When $q < q_1(p)$ and $p < p^*$, with commitment, the bad ruler represses the good activist more often than in any equilibrium without commitment.*

Given that the bystander does not protest in equilibrium, absent commitment, the bad ruler would raise repression and always repress the good activist, thereby upsetting the equilibrium. Commitment power benefits the bad ruler by enabling him to "support" the equilibrium in which the good ruler always represses the bad activist, and leverages this by raising his repression of the good activist. Similar strategic considerations arise in region IV. There, the bad ruler represses as much as he would do in the low repression equilibrium absent commitment, but eliminates the bystander's protest. To summarize, the bad ruler exploits his commitment power in two distinct ways: in region III, he *raises* repression, and yet *lowers* the likelihood of the bystander's protest; and in region IV, he maintains the same level of repression (as in the low repression equilibrium of the game without commitment), but eliminates the risk of the bystander's protest.

Repression Backfire with Commitment. The bad ruler's commitment power enables him to eliminate the bystander's protest when the prior likelihoods that the activist or the ruler are bad are large enough. One may wonder whether the bad ruler's ability to ex ante limit its repression can also eliminate repression backfire.

[27] from Propositions 2, recall that when $q < q_1(p)$ which contains region III, the bad ruler represses with probability $\rho_g^B = \frac{\beta_b - \beta}{\beta_g + \beta} \frac{q}{1-q}$ which is *less* than his level of repression $r_g = \frac{q}{1-q} \frac{\beta_b - p\beta}{p(\beta + \beta_g)}$ when he can commit. Importantly, absent commitment, the good ruler does not repress in equilibrium; in contrast, with commitment, the good ruler always represses the bad activist.

Proposition 6 *With commitment, repression backfire occurs in equilibrium if and only if* $\delta_b < \frac{\beta}{\beta+\beta_g}\ \alpha_g$ *and* $q_b(p) < q < q^*$.

Proposition 6 shows that commitment power does not eliminate repression backfire even though it imposes more stringent conditions for repression backfire to occur relative to the case without commitment. The conditions with and without commitment are qualitatively similar. In both cases, δ_b must be smaller than α_g, β must be positive, and q must be intermediate. However, commitment power imposes a tighter bound on δ_b, requiring it to be smaller than $\frac{\beta}{\beta+\beta_g}\alpha_g$ instead of α_g, which in turn, implies $\beta > 0$. It also imposes a tighter bound on q, requiring it to be smaller than $q^* = (1 - \frac{\delta_b}{\alpha_g})q_2$ instead of q_2.

Moreover, repression backfire can occur only in a a subset of the equilibria. As in the case without commitment, this suggests that the societal culture plays a critical role in whether or not repression backfires in equilibrium. In particular, as we described in the discussion leading to Proposition 5, whenever $0 < F(r_b, r_g) < K$, three equilibria exists. In the text, we focused on the equilibrium in which the bystander protests with probability δ_b whenever $0 < F(r_b, r_g) < K$. Proposition 6 characterizes the conditions for repression backfire given this equilibrium selection. In Lemma 5 in the Appendix C, we analyze the other two equilibria of the subgame that follows the bad ruler's choice of repression strategy: one in which $\pi = 0$ and another in which $\pi = 1$ whenever $0 < F(r_b, r_g) < K$. In Proposition 9 and 10 in the Appendix C, we show that the bystander never protests and repression does not backfire if either of these two is the equilibrium of the subgame. That is, the government's and the public's expectations of each other's actions continue to play a role in determining whether repression backfires.

6 CONCLUSION

"The seed of revolution is repression," once said Woodrow Wilson. Repression can deter protest and maintain the status quo, but it can also change the people's beliefs about the Leviathan: it is informative about the nature of the beast.[28] The literature has focused on the deterrence aspect of repression, ignoring the information that repression reveals about the nature of the state. We abstracted from the well-studied deterrence effect of repression, and focused on this informational aspect. We developed a model that suggests an informational explanation for repression backfire based on the key observation that the general public has limited information about the nature

[28]Refers to the treatise by the 17th century English political philosopher Thomas Hobbes, entitled *Leviathan: Or the Matter, Forme, and Power of a Common-Wealth Ecclesiasticall and Civill.*

of the activists' demands and the intentions of the state that represses them. We then analyzed the novel strategic considerations that arise when the state can set up institutions that limit its ability to repress, showing that the state's commitment power imposes more stringent conditions for repression backfire to occur, but does not eliminate it.

7 APPENDIX

7.1 PROOFS OF THE RESULTS WITHOUT COMMITMENT

Proof of Remark 1. Let $\eta \in [0,1]$ be the probability that the bystander protests. From equation (1), any value of $\eta \in [0,1]$ is supported by some combination of beliefs. If a bad ruler faces a good activist he represses whenever $\eta < \alpha_g$, but if a bad ruler faces a bad activist he represses whenever $\eta < \alpha_b$. Because $\alpha_g < \alpha_b$ the belief that the bad ruler is repressing the good activist violates the D1 criterion. ■

Proof of Proposition 2. From Lemma 1, $\rho_g^B > 0$. Hence, equation (2) implies that $\pi \leq \alpha_g$. An equilibrium with $\pi \in (\delta_b, \alpha_g)$ is generically impossible. If $\pi \in (\delta_b, \alpha_g)$, then $\rho_b^G = 0$ and $\rho_g^B = 1$. Hence, $p' = 1$ and $q' = q$. Because $\pi \in (0,1)$, equation (1) implies that $\beta_g(1-q) - \beta_b q = -\beta$, but this is non-generic. Similarly, an equilibrium with $\pi \in (0, \delta_b)$ is generically impossible. Hence, there are only three possibilities: $\pi = \alpha_g$, $\pi = \delta_b$, and $\pi = 0$.

Suppose that $\pi = \alpha_g$. Because $\delta_b < \alpha_g$, equation (2) implies that $\rho_b^G = 0$ and $\rho_g^B \in [0,1]$, and hence (3) implies:

$$q' = \frac{q}{q + (1-q)\rho_g^B} \qquad p' = 1.$$

Because $\pi \in (0,1)$, equation (1) implies that $\beta_g(1-q') - \beta_b q' = -\beta p'$. Substituting (p', q') gives:

$$\beta_g\left(1 - \frac{q}{q + (1-q)\rho_g^B}\right) - \beta_b \frac{q}{q + (1-q)\rho_g^B} = -\beta \iff \rho_g^B = \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta},$$

and hence, $\rho_g^B \in [0,1] \Leftrightarrow q \leq \frac{\beta + \beta_g}{\beta_b + \beta_g}$. Hence, an equilibrium with $\pi = \alpha_g$, $\rho_b^G = 0$, and $\rho_g^B = \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}$ exists if and only if $q \leq q_2$.

Suppose that $\pi = \delta_b$. Because $\delta_b < \alpha_g$, equation (2) implies that $\rho_b^G \in [0,1]$ and $\rho_g^B = 1$. Hence, (3) implies:

$$q' = \frac{q[(1-p)\rho_b^G + p]}{q[(1-p)\rho_b^G + p] + p(1-q)} \qquad p' = \frac{p}{p + (1-p)q\rho_b^G}.$$

Because $\pi \in (0,1)$, equation (1) implies that $\beta_g(1-q') - \beta_b q' = -\beta p'$. Substituting (p', q') gives:

$$\beta_g(1 - \frac{q[(1-p)\rho_b^G + p]}{q[(1-p)\rho_b^G + p] + p(1-q)}) - \beta_b \frac{q[(1-p)\rho_b^G + p]}{q[(1-p)\rho_b^G + p] + p(1-q)} = -\beta \frac{p}{p + (1-p)q\rho_b^G},$$

which implies:

$$\rho_b^G = \frac{p}{1-p} \frac{(\beta + \beta_g) - (\beta_b + \beta_g)q}{\beta_b q}.$$

Hence, $\rho_g^B \in [0,1] \Leftrightarrow \frac{(\beta+\beta_g)p}{\beta_b+\beta_g p} \leq q \leq \frac{\beta+\beta_g}{\beta_b+\beta_g}$, where we recognize that $\frac{(\beta+\beta_g)p}{\beta_b+\beta_g p} < \frac{\beta+\beta_g}{\beta_b+\beta_g}$ for $p < 1$. Hence, an equilibrium with $\pi = \delta_b$, $\rho_b^G = \frac{p}{1-p}\frac{(\beta+\beta_g)-(\beta_b+\beta_g)q}{\beta_b q}$, and $\rho_g^B = 1$ exists if and only if $q_1(p) \leq q \leq q_2$.

Suppose that $\pi = 0$. Equation (2) implies that $\rho_b^G = \rho_g^B = 1$. Hence, (3) implies:

$$q' = \frac{q}{q + p(1-q)} \qquad p' = \frac{p}{p + (1-p)q}.$$

Because $\pi = 0$, equation (1) implies that $\beta_g(1-q') - \beta_b q' \leq -\beta p'$. Substituting (p', q') gives:

$$\beta_g(1 - \frac{q}{q + p(1-q)}) - \beta_b \frac{q}{q + p(1-q)} \leq -\beta \frac{p}{p + (1-p)q} \iff q \geq \frac{(\beta + \beta_g)p}{\beta_b + \beta_g p}.$$

Hence, an equilibrium with $\pi = 0$, $\rho_b^G = \rho_g^B = 1$ exists if and only if $q \geq q_1(p)$. ■

Next, we prove two lemmas that we use in the proof of Proposition 3.

Lemma 3 *Suppose $\delta_b > \alpha_g$. There is a unique equilibrium. In equilibrium, the good ruler never represses a good activist and the bad ruler always represses a bad activist, $\rho_g^G = 0$ and $\rho_b^B = 1$.*

- *When $q \leq q_1(p)$, The good ruler always represses the bad activist, the bad ruler represses the good activist with a positive probability less than one, and upon observing repression, the bystander protests with a positive probability less than one: $\rho_b^G = 1$, $\rho_g^B = \frac{\beta_b - p\beta}{(\beta_g+\beta)p}\frac{q}{1-q}$, and $\pi = \alpha_g$.*
- *When $q \geq q_1(p)$ the equilibrium is identical to the high repression equilibrium: the good ruler always represses the bad activist, $\rho_b^G = 1$, the bad ruler always represses the good activist, $\rho_g^B = 1$, and the bystander never protests, $\pi = 0$.*

Proof of Lemma 3. From Lemma 1, $\rho_g^B > 0$. Hence, equation (2) implies that $\pi \leq \alpha_g$. An equilibrium with $\pi \in (0, \alpha_g)$ is generically impossible. If $\pi \in (0, \alpha_g)$, then $\rho_b^G = 1$ and $\rho_g^B = 1$. Hence, $p' = \frac{p}{p+p(1-q)}$ and $q' = \frac{q}{q+p(1-q)}$. Because $\pi \in (0,1)$, equation (1) implies that $\beta_g(1-q') - \beta_b q' =$

$-\beta p'$, but this is non-generic. Hence, there are only two possibilities: $\pi = \alpha_g$ and $\pi = 0$.

Suppose that $\pi = \alpha_g$. Because $\delta_b > \alpha_g$, equation (2) implies that $\rho_b^G = 1$ and $\rho_g^B \in [0,1]$. Hence, (3) implies:

$$q' = \frac{q}{q + p(1-q)\rho_g^B} \qquad p' = \frac{p[q + (1-q)\rho_g^B]}{p[q + (1-q)\rho_g^B] + (1-p)q}.$$

Because $\pi \in (0,1)$, equation (1) implies that $\beta_g(1-q') - \beta_b q' = -\beta p'$. Substituting (p', q') gives:

$$\beta_g(1 - \frac{q}{q + p(1-q)\rho_g^B}) - \beta_b \frac{q}{q + p(1-q)\rho_g^B} = -\beta \frac{p[q + (1-q)\rho_g^B]}{p[q + (1-q)\rho_g^B] + (1-p)q} \iff \rho_g^B = \frac{q}{1-q}\frac{\beta_b - p\beta}{(\beta_g + \beta)p},$$

and hence, $\rho_g^B \in [0,1] \Leftrightarrow q \leq \frac{\beta + \beta_g}{\beta_b + \beta_g}$. Hence, an equilibrium with $\pi = \alpha_g$, $\rho_b^G = 1$, and $\rho_g^B = \frac{q}{1-q}\frac{\beta_b - p\beta}{(\beta_g + \beta)p}$ exists if and only if $q \leq q_1(p)$.

Suppose that $\pi = 0$. Equation (2) implies that $\rho_b^G = \rho_g^B = 1$. Hence, (3) implies:

$$q' = \frac{q}{q + p(1-q)} \qquad p' = \frac{p}{p + (1-p)q}.$$

Because $\pi = 0$, equation (1) implies that $\beta_g(1-q') - \beta_b q' \leq -\beta p'$. Substituting (p', q') gives:

$$\beta_g(1 - \frac{q}{q + p(1-q)}) - \beta_b \frac{q}{q + p(1-q)} \leq -\beta \frac{p}{p + (1-p)q} \iff q \geq \frac{(\beta + \beta_g)p}{\beta_b + \beta_g p}.$$

Hence, an equilibrium with $\pi = 0$, $\rho_b^G = \rho_g^B = 1$ exists if and only if $q \geq q_1(p)$. ■

Lemma 4 *(1) $q_b(p) > q_1(p)$, (2) if $\beta > 0$, then $q_b(p) < q_2$, and (3) if $\beta = 0$, then $q_b(p) = q_2$.*

Proof of Lemma 4. Recall that $p \in (0,1)$. To see (1), observe that $q_b(p) - q_1(p) = (1-p)\frac{\beta_g}{\beta_b + \beta_g}\frac{\beta_b - p\beta}{\beta_b + p\beta_g}$, which is bigger than 0 because $\beta_b > \beta > p\beta$. To see (2) and (3), observe that $q_2 - q_b(p) = \frac{\beta(1-p)}{\beta_b + \beta_g}$, which is bigger than 0 if and only if $\beta > 0$. For $\beta = 0$, $q_2 - q_b(p) = 0$ for all p. ■

Proof of Proposition 3. In order for repression backfire to occur in equilibrium, the bystander must be unwilling to protest given the prior belief, i.e. $q > q_b(p)$. From Lemma 4, $q_b(p) > q_1(p)$, and hence $q > q_b(p) \Rightarrow q > q_1(p)$. From Lemma 3, if $\delta_b > \alpha_g$ and $q > q_1(p)$, then $\pi = 0$ in the unique equilibrium. Hence, if $\delta_b > \alpha_g$, then no repression backfire occurs.

From Lemma 4, $\beta = 0 \Rightarrow q_b(p) = q_2$, and hence repression backfire requires $q > q_2$. From Proposition 2, if $\delta_b < \alpha_g$ and $q > q_2$, then $\pi = 0$ in the unique equilibrium. Hence, no repression

backfire occurs when $\beta = 0$ and $\delta_b < \alpha_g$.

Suppose $\delta_b < \alpha_g$ and $\beta > 0$. From Lemma 4, $\beta > 0 \Rightarrow q_1(p) < q_b(p) < q_2$. Hence, Proposition 2, implies that for $q_b(p) < q < q_2$ the intermediate and low repression equilibrium both exist. In both of these equilibria, $\pi > 0$ and the ruler sometimes represses. Hence, repression backfire occurs. ■

Proof of Proposition 4. Lemma 1 holds even if the good ruler is behavioral, and it implies that $\rho_g^B > 0$. Hence equation (2) implies that $\pi \leq \alpha_g$. As in the proof of Proposition 2, an equilibrium with $\pi \in (0, \alpha_g)$ is not generic. Hence $\pi = 0$ or $\pi = \alpha_g$. An equilibrium in which $\pi = 0$ cannot feature repression backfire. Hence, the proof shows that for r large, no equilibrium in which $\pi = \alpha_g$ and $q > q_b(p)$ exists.

Suppose that $\pi = \alpha_g$. Equation (3) implies:

$$q' = \frac{(1-p)qr + pq}{(1-p)qr + pq + p(1-q)\rho_g^B}, \qquad p' = \frac{pq + p(1-q)\rho_g^B}{(1-p)qr + pq + p(1-q)\rho_g^B}.$$

Because $\pi \in (0,1)$, equation (1) implies that $\beta_g(1-q') - \beta_b q' = -\beta p'$. Substituting (p', q') gives:

$$\beta_g(1 - \frac{(1-p)qr + pq}{(1-p)qr + pq + p(1-q)\rho_g^B}) - \beta_b \frac{(1-p)qr + pq}{(1-p)qr + pq + p(1-q)\rho_g^B} = -\beta \frac{pq + p(1-q)\rho_g^B}{(1-p)qr + pq + p(1-q)\rho_g^B},$$

and hence,

$$\rho_g^B = \frac{q}{1-q} \frac{(1-p)r\beta_b + p(\beta_b - \beta)}{p(\beta + \beta_g)}.$$

Thus, a necessary condition for repression backfire is that this $\rho_g^B \in [0,1]$. Let $R(p) \equiv 1 - \frac{\beta_g}{\beta_b} \frac{\beta_b - p\beta}{\beta_g + p\beta} = \frac{p\beta}{\beta_b} \frac{\beta_b + \beta_g}{\beta_g + p\beta}$. Clearly, $R(p) \in (0,1)$ for $p \in (0,1)$. In order for an equilibrium to exhibit repression backfire, the bystander must not be willing to protest under the prior, $q > q_b(p)$. Because ρ_g^B is increasing in r and q, if an equilibrium exhibits repression backfire and $r > R(p)$, then:

$$\rho_g^B > \frac{q_b(p)}{1 - q_b(p)} \frac{(1-p)R(p)\beta_b + p(\beta_b - \beta)}{p(\beta + \beta_g)} = 1.$$

Hence, for $r > R(p)$, no equilibrium exhibiting repression backfire exists. ■

7.2 PROOFS OF THE RESULTS WITH BAD RULER COMMITMENT

Let $B(r_b, r_g, \pi)$ be the bad ruler's payoff, if he remains in power, from (r_b, r_g) that induces protest probability π in the equilibrium of the subgame:

$$(4) \qquad B(r_b, r_g, \pi) = q\ [r_b(1-\pi) + (1-r_b)(1-\alpha_b)] + (1-q)\ [r_g(1-\pi) + (1-r_g)(1-\alpha_g)],$$

so that the bad ruler's ex ante payoff is $pB(r_b, r_g, \pi)$. We repeat Lemma 2 to ease communication.

Lemma 2 *Given the strategy of the bad ruler (r_b, r_g) and the strategy of the good ruler ρ_b^G, the bystander's best response is:*

$$\pi = \begin{cases} 1 & \text{if } F(r_b, r_g) > K\rho_b^G \\ [0,1] & \text{if } F(r_b, r_g) = K\rho_b^G \\ 0 & \text{if } F(r_b, r_g) < K\rho_b^G \end{cases}$$

where $F(r_b, r_g) \equiv (1-q)(\beta_g + \beta)r_g - q(\beta_b - \beta)r_b$ and $K \equiv \frac{1-p}{p} q\beta_b$.

Proof. Substituting from the bystander's posterior belief, equation (3), into equation (1) gives the result. ■

Lemma 5 *Fix the bad ruler's strategy (r_b, r_g). The following characterizes the equlibria of the subgame:*

1. *If $F(r_b, r_g) < 0$, then the unique equilibrium of the subgame has $\pi = 0$ and $\rho_b^G = 1$.*
2. *If $F(r_b, r_g) > K$, then the unique equilibrium of the subgame has $\pi = 1$ and $\rho_b^G = 0$.*
3. *If $0 < F(r_b, r_g) < K$, then the equilibria described in (1) and (2) both exist. In addition, there is an equilibrium of the subgame in which $\pi = \delta_b$ and $\rho_b^G = pF(r_b, r_g)/((1-p)q\beta_b)$.*
4. *If $F(r_b, r_g) = 0$, then the equilibrium described in (1) exists. In addition, a continuum of equilibria exist in which $\pi \in [\delta_b, 1]$ and $\rho_b^G = 0$.*
5. *If $F(r_b, r_g) = K$, then the equilibrium described in (2) exists. In addition, a continuum of equilibria exist in which $\pi \in [0, \delta_b]$ and $\rho_b^G = 1$.*

Proof. (1) If (r_b, r_g) is such that $F(r_b, r_g) < 0$, then $F(r_b, r_g) < K\rho_b^G$ for any $\rho_b^G \in [0,1]$. Hence, in the subgame $\pi = 0$ for any $\rho_b^G \in [0,1]$. Because $\pi = 0$, equation (2) requires that $\rho_b^G = 1$.

(2) If (r_b, r_g) is such that $F(r_b, r_g) > K$, then $F(r_b, r_g) > K\rho_b^G$ for any $\rho_b^G \in [0,1]$. Hence, in the subgame $\pi = 1$ for any $\rho_b^G \in [0,1]$. Because $\pi = 1$, equation (2) requires that $\rho_b^G = 0$.

(3) If (r_b, r_g) is such that $F(r_b, r_g) \in (0, \frac{1-p}{p}\beta_b)$, then for $\rho_b^G = 1$, $\pi = 0$ is consistent with Lemma 2,

and $\rho_b^G = 1$ is consistent with equation (2) when $\pi = 0$. For $\rho_b^G = 0$, $\pi = 1$ is consistent with Lemma 2, and $\rho_b^G = 0$ is consistent with equation (2) when $\pi = 1$. Assumption $F(r_b, r_g) \in (0, \frac{1-p}{p}\beta_b)$ implies $\rho_b^G = F(r_b, r_g)/K \in (0, 1)$. Thus, equation (2) implies that $\pi = \delta_b$. Because $\pi = \delta_b \in (0, 1)$, Lemma 2 implies that $\rho_b^G = F(r_b, r_g)/K$.

(4) If (r_b, r_g) is such that $F(r_b, r_g) = 0$, then for $\rho_b^G = 1$, $\pi = 0$ is consistent with Lemma 2, and $\rho_b^G = 1$ is consistent with equation (2) when $\pi = 0$. In addition, for $\rho_b^G = 0$, Lemma 2 implies that $\pi = [0, 1]$, and $\rho_b^G = 0$ is consistent with equation (2) if and only if $\pi \in [\delta_b, 1]$.

(5) If (r_b, r_g) is such that $F(r_b, r_g) = K$, then for $\rho_b^G = 0$, $\pi = 1$ is consistent with Lemma 2, and $\rho_b^G = 0$ is consistent with equation (2) when $\pi = 1$. In addition, for $\rho_b^G = 1$, Lemma 2 implies that $\pi = [0, 1]$, and $\rho_b^G = 1$ is consistent with equation (2) if and only if $\pi \in [0, \delta_b]$. ■

Lemma 6 *If $q > q_2$, then there is a unique equilibrium in which $r_b = r_g = \rho_b^G = 1$ and $\pi = 0$.*

Proof. If $q > q_2$, then $F(1, 1) < 0$, which implies $\pi = 0$ from Lemma 5. Thus, $\rho_b^G = 1$. The bad ruler's payoff is $B(1, 1, 0) = 1$ which is strictly larger than $B(r_b, r_g, \pi)$ for any $(r_b, r_g) \neq (1, 1)$ and $\pi \in [0, 1]$. ■

Equilibrium Selection. Next, we impose the equilibrium selection ES1, and make the observation ES2:

ES1. If $F(r_b, r_g) = K$, then the equilibrium of the subgame we have $\pi = \delta_b$ and $\rho_b^G = 1$.

ES2. If $F(r_b, r_g) = 0$, then in the equilibrium of the subgame we have $\pi = 0$ and $\rho_b^G = 1$.

ES1 is an equilibrium selection that is justified using a refinement similar to trembling hand in Proposition 7. ES2 is justified in Lemma 9.

Lemma 7 *1. If $q < q_2$, then $R_0 \equiv (1, \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta})$ is the unique strategy that maximizes the bad ruler's expected payoff among all (r_b, r_g) for which $F(r_b, r_g) \leq 0$, and the associated payoff is $B_0 \equiv 1 - \alpha_g(1 - q/q_2)$.*

2. If $q_1(p) \leq q < q_2$, then $R_1 \equiv (1, 1)$ is the unique strategy that maximizes the bad ruler's expected payoff among all (r_b, r_g) for which $0 < F(r_b, r_g) \leq K$, and the associated payoff is $B_1 \equiv 1 - \delta_b$.

3. If $q < q_1(p)$, then $R_2 \equiv (1, \frac{q}{1-q}\frac{\beta_b - p\beta}{p(\beta_g + \beta)})$ is the unique strategy that maximizes the bad ruler's

expected payoff among (r_b, r_g) *for which* $0 < F(r_b, r_g) \leq K$*, and the associated payoff is:*

$$B_2 \equiv 1 - \alpha_g \left(1 - \frac{q}{q_2}\right) + \frac{q(\beta_b(\alpha_g - \delta_b) - p(\beta_b\alpha_g + \delta_b\beta_g)}{p(\beta + \beta_g)}.$$

Proof. 1. If $F(r_b, r_g) \leq 0$ and $q < q_2$, then $\pi = 0$ in the equilibrium of the subgame. From (4), the payoff of such a strategy is:

$$B(r_b, r_g, 0) = q(r_b + (1 - r_b)(1 - \alpha_b)) + (1 - q)(r_g + (1 - r_g)(1 - \alpha_g)).$$

Thus, the ruler's problem becomes:

$$\max_{(r_b, r_g) \in [0,1]^2} B(r_b, r_g, 0) \text{ s.t. } F(r_b, r_g) \leq 0.$$

$B(r_b, r_g, 0)$ is increasing in both r_b and r_g. Because $F(r_b, r_g)$ is decreasing in r_b, we must have $r_b = 1$ at the optimum. Because $q < q_2$, $F(1, 1) > 0$, and hence $r_b = r_g = 1$ is not feasible. Because $F(r_b, r_g)$ is increasing in r_g, we must have $F(1, r_g) = 0$ at the optimum r_g. Finally, $F(1, r_g) = 0$ implies $r_g = \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}$. B_0 is derived from substituting $(r_b, r_g) = (1, \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta})$ into $B(r_b, r_g, 0)$.

If $0 < F(r_b, r_g) \leq K$, then $\pi = \delta_b$ in the equilibrium of the subgame. From (4), the bad ruler's payoff from any such strategy is:

$$B(r_b, r_g, \delta_b) = q\ [r_b(1 - \delta_b) + (1 - r_b)(1 - \alpha_b)] + (1 - q)\ [r_g(1 - \delta_b) + (1 - r_g)(1 - \alpha_g)].$$

Thus, the ruler's problem in parts 2 and 3 becomes:

$$\max_{(r_b, r_g) \in [0,1]^2} B(r_b, r_g, \delta_b) \quad s.t. \quad 0 < F(r_b, r_g) \leq K.$$

Because $\delta_b < \alpha_g < \alpha_b$, $B(r_b, r_g, \delta_b)$ is increasing in both r_b and r_g.

2. If $q_1(p) \leq q < q_2$, then $0 < F(1, 1) \leq K$, and hence, $r_b = r_g = 1$ is feasible. Because $B(r_b, r_g, \delta_b)$ is increasing in r_b and r_g, $r_b = r_g = 1$ is the bad ruler's optimal choice. B_1 is derived by substituting $(r_b, r_g) = (1, 1)$ into $B(r_b, r_g, \delta_b)$.

3. When $q < q_1(p)$, $F(1, 1) > K$, and hence the constraint $F(r_b, r_g) = K$ binds. Because F is decreasing in r_b, we must have $r_b = 1$ at the optimum. Then, the optimal r_g is derived from $F(r_b = 1, r_g) = K$. ■

Lemma 8 *In equilibrium, the payoff of selecting any* (r_b, r_g) *such that* $F(r_b, r_g) > K$ *is smaller than* B_0*.*

Proof. If $F(r_b, r_g) > K$, then $\pi = 1$, and hence $B(r_b, r_g, 1) = q(1-r_b)(1-\alpha_b)+(1-q)(1-r_g)(1-\alpha_g)$. The bad ruler can benefit by deviating to $(r_b, r_g) = (0, 0)$, so that $F(r_b, r_g) = 0$, and his expected payoff becomes $B(0, 0, 0) = q(1 - \alpha_b) + (1 - q)(1 - \alpha_g) > B(r_b, r_g, 1)$. Because $F(0, 0) = 0$, but $R_0 \neq (0, 0)$ is optimal among $F(r_b, r_g) \leq 0$ it must be that $B(0, 0, 0) < B_0$. Summarizing, If $F(r_b, r_g) > K$, then $B(r_b, r_g, 1) < B(0, 0, 0) < B_0$. ■

Proof of Proposition 5. Lemma 6 establishes part 1. Thus, we focus on $q < q_2$ in the rest of the proof.

Suppose $q_1(p) < q < q_2$. If $B_0 > B_1$, then Lemma 7 implies that R_0 dominates any strategy for which $F(r_b, r_g) \leq K$, and Lemma 8 implies that R_0 dominates any strategy for which $F(r_b, r_g) > K$. Hence, if $B_0 > B_1$, then $R_0 = (1, \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta})$ dominates all $(r_b, r_g) \in [0, 1]^2$, and it is the bad ruler's equilibrium choice. If $B_1 > B_0$, then Lemma 7 implies that R_1 dominates any strategy for which $F(r_b, r_g) \leq K$. Because $B_1 > B_0$, Lemma 8 implies that R_1 dominates any strategy for which $F(r_b, r_g) > K$. Hence, if $B_1 > B_0$ then $R_1 = (1, 1)$ dominates all $(r_b, r_g) \in [0, 1]^2$, and it is the bad ruler's equilibrium choice. Using B_0 and B_1 from Lemma 7,

$$B_1 > B_0 \text{ if and only if } q < q_2\left(1 - \frac{\delta_b}{\alpha_g}\right).$$

Thus, if $q_1(p) < q < q_2(1 - \frac{\delta_b}{\alpha_g})$, in equilibrium, $(r_b, r_g) = (1, 1)$, $\rho_b^G = \frac{F(1,1)}{K} \in (0, 1)$, and $\pi = \delta_b$. If $q_2(1 - \frac{\delta_b}{\alpha_g}) < q < q_2$, in equilibrium, $(r_b, r_g) = (1, \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta})$, $\rho_b^G = 1$, and $\pi = 0$.

Suppose $q < q_1(p)$. If $B_0 > B_2$, then Lemma 7 implies that R_0 dominates any strategy for which $F(r_b, r_g) \leq K$, and Lemma 8 implies that R_0 dominates any strategy for which $F(r_b, r_g) > K$. Hence, if $B_0 > B_2$, then $R_0 = (1, \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta})$ dominates all $(r_b, r_g) \in [0, 1]^2$, and it is the bad ruler's equilibrium choice. If $B_2 > B_0$, then Lemma 7 implies that R_2 dominates any strategy for which $F(r_b, r_g) \leq K$. Because $B_2 > B_0$, Lemma 8 implies that R_2 dominates any strategy for which $F(r_b, r_g) > K$. Hence, if $B_2 > B_0$ then $R_2 = (1, \frac{q}{1-q}\frac{\beta_b-p\beta}{p(\beta_g+\beta)})$ dominates all $(r_b, r_g) \in [0, 1]^2$, and it is the bad ruler's equilibrium choice. Using B_0 and B_2 from Lemma 7,

$$B_2 > B_0 \text{ if and only if } p < \frac{\beta_b(\alpha_g - \delta_b)}{\beta_b\alpha_g + \delta_b\beta_g}.$$

Thus, if $p < \frac{\beta_b(\alpha_g-\delta_b)}{\beta_b\alpha_g+\delta_b\beta_g}$ *and* $q < q_1(p)$, in equilibrium, $(r_b, r_g) = (1, \frac{p}{1-p}\frac{\beta_b-p\beta}{p(\beta+\beta_g)})$, $\rho_b^G = 1$, and $\pi = \delta_b$. If $q < q_1(p)$ *and* $p > \frac{\beta_b(\alpha_g-\delta_b)}{\beta_b\alpha_g+\delta_b\beta_g}$, in equilibrium, $(r_b, r_g) = (1, \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta})$, $\rho_b^G = 1$, and $\pi = 0$. ■

Proof of Proposition 6. If $q < q_1(p) < q_b(p)$ or $\pi = 0$ then the equilibrium cannot exhibit repression backfire. Hence, Proposition 5 implies that the equilibrium exhibits repression backfire

if and only if $q_1(p) < q < q_2(1 - \frac{\delta_b}{\alpha_g})$, so that $\pi = \delta_b$, and $q > q_b(p)$, so that the bystander does not protest under the prior. From Lemma 4, $q_b(p) > q_1(p)$, hence a necessary and sufficient condition for repression backfire to be possible is $q_b(p) < q < q_2(1 - \frac{\delta_b}{\alpha_g})$ for some $p \in (0, 1)$. Because $q_b(p)$ is an increasing, continuous function of p, $q_b(p) < q_2(1 - \frac{\delta_b}{\alpha_g})$ if and only if

$$q_b(0) < q_2(1-\frac{\delta_b}{\alpha_g}) \Leftrightarrow \frac{\beta_g}{\beta_g + \beta_b} < \frac{\beta + \beta_g}{\beta_b + \beta_g}(1-\frac{\delta_b}{\alpha_g}) \Leftrightarrow \beta_g < (\beta+\beta_g)(1-\frac{\delta_b}{\alpha_g}) \Leftrightarrow (\beta+\beta_g)\frac{\delta_b}{\alpha_g} < \beta \Leftrightarrow \delta_b < \frac{\beta}{\beta + \beta_g}\alpha_g.$$

■

7.3 EQUILIBRIUM SELECTION IN THE GAME WITH COMMITMENT

In this section, we present two results which justify our equilibrium selections.

ES1. From Lemma 5 point 5, when $F(r_b, r_g) = K$, a continuum of equilibria are possible in the subgame: any $\pi \in [0, \delta_b]$ can be part of the equilibrium of the subgame. In the text, we focus on the equilibrium of the subgame in which $\pi = \delta_b$. Here, we show that whenever the bad ruler's equilibrium strategy (r_b, r_g) has $F(r_b, r_g) = K$, the subgame with $\pi = \delta_b$ is uniquely selected by a simple refinement. In particular, we introduce stochastic shocks to the bad ruler's strategy, showing that as the support of the distribution of shocks vanishes, the equilibrium with the shocks converges to the one in which $\pi = \delta_b$.

Suppose that when the bad ruler commits to a strategy (r_b, r_g), the probability with which the type i activist is actually repressed is a random variable $R(r_i) \equiv \max\{\min\{r_i + \nu_i, 1\}, 0\}$, where ν_i's are iid continuous random variables with support $[-\epsilon, \epsilon]$. Let $\hat{r}_i$ be the realization of the random variable $R(r_i)$. Let function $\pi^*(\hat{r}_b, \hat{r}_g)$ represent the protest probability in the equilibrium of the subgame following $(\hat{r}_b, \hat{r}_g)$:

$$\pi^*(\hat{r}_b, \hat{r}_g) = \begin{cases} 1 & \text{if} \quad F(\hat{r}_b, \hat{r}_g) > K \\ \delta_b & \text{if} \quad 0 < F(\hat{r}_b, \hat{r}_g) < K \\ 0 & \text{if} \quad F(\hat{r}_b, \hat{r}_g) < 0 \end{cases}$$

From Lemma 5, if $F(\hat{r}_b, \hat{r}_g) > K$, then $\pi^*(\hat{r}_b, \hat{r}_g) = 1$, and if $F(\hat{r}_b, \hat{r}_g) < 0$, then $\pi^*(\hat{r}_b, \hat{r}_g) = 0$. In Proposition 5, we focus on the case in which $0 < F(\hat{r}_b, \hat{r}_g) < K$ implies $\pi = \delta_b$, one of the three options that follows from Lemma 5 (the others are considered in Propositions 9, 10). Because ν_is are independent and have no mass points, for any choice of the bad ruler (r_b, r_g), the probability that the realizations are such that $F(\hat{r}_b, \hat{r}_g) = K$ or $F(\hat{r}_b, \hat{r}_g) = 0$ is zero.

For a given ϵ, the bad ruler's problem is:

$$\max_{(r_b,r_g)} \; E[\; B(R(r_b),R(r_g),\pi^*(R(r_b),R(r_g)))\;],$$

where the expectation is over (ν_b,ν_g). Let $(r_b^*(\epsilon),r_g^*(\epsilon))$ be the maximand(s) and $B^*(\epsilon)$ be the value of the maximum. From Proposition 5, when $q<q_1(p)$ and $p<\frac{\beta_b(\alpha_g-\delta_b)}{\beta_b\alpha_g+\delta_b\beta_g}$, absent trembles, the bad ruler's equilibrium choice is $(r_b^*,r_g^*)\equiv(1,\frac{q}{1-q}\frac{\beta_b-p\beta}{p(\beta+\beta_g)})$, which delivers the bad ruler a payoff of $B^*\equiv B(r_b^*,r_g^*,\delta_b)$. This is the only instance in which the bad ruler's equilibrium choice is such that $F(r_b,r_g)=K$.

Proposition 7 *Suppose $q<q_1(p)$ and $p<\frac{\beta_b(\alpha_g-\delta_b)}{\beta_b\alpha_g+\delta_b\beta_g}$, so that the bad ruler's choice in the absence of stochastic shocks is (r_b^*,r_g^*). As the support of the distribution of the shocks shrinks to zero:*

1. *The bad ruler's payoff converges to his payoff in the absence of shocks: $\lim_{\epsilon\to0}B^*(\epsilon)=B^*$*
2. *The protest probability converges to δ_b: $\lim_{\epsilon\to0}\Pr\{\pi^*(R(r_b^*(\epsilon)),R(r_g^*(\epsilon)))=\delta_b\}=1$.*
3. *The bad ruler's strategy converges to (r_b^*,r_g^*): $\lim_{\epsilon\to0}r_i^*(\epsilon)=r_i^*$ for $i\in\{b,g\}$.*

Proof. (1) Because (r_b^*,r_g^*) is optimal for the bad ruler in the absence of trembles, $B^*>B(\hat r_b,\hat r_g,\pi^*(\hat r_b,\hat r_g))$ for all possible realizations $(\hat r_b,\hat r_g)\neq(r_b^*,r_g^*)$, and hence

$$B^*>B^*(\epsilon). \tag{5}$$

Recall that $(r_b^*,r_g^*)=(1,\frac{q}{1-q}\frac{\beta_b-p\beta}{p(\beta+\beta_g)})$, and consider an alternative strategy for the bad ruler:

$$(r_b',r_g')\equiv(r_b^*,r_g^*-\epsilon(1+\frac{q}{1-q}\frac{\beta_b-\beta}{\beta+\beta_g})),$$

so that $F(r_b'-\epsilon,r_g'+\epsilon)=K$ (see Figure 4). For a sufficiently small ϵ, the monotonicity properties of F imply:

$$0<F(1,r_g'-\epsilon)<F(R(r_b'),R(r_g'))<F(r_b'-\epsilon,r_g'+\epsilon)=K.$$

That is, if the bad ruler chooses (r_b',r_g'), then for any realization of shocks $0<F(R(r_b'),R(r_g'))<K$, and hence $\Pr\{\pi^*(R(r_b'),R(r_g'))=\delta_b\}=1$. Thus, the ruler's expected payoff from (r_b',r_g') is $E[B(R(r_b'),R(r_g'),\delta_b)]$. Let $k\equiv2+\frac{q}{1-q}\frac{\beta_b-\beta}{\beta+\beta_g}$, so that $B(r_b^*-\epsilon,r_g^*-k\epsilon,\delta_b)=B(r_b'-\epsilon,r_g'-\epsilon,\delta_b)$. Then,

$$B(r_b^*-\epsilon,r_g^*-k\epsilon,\delta_b)=B(r_b'-\epsilon,r_g'-\epsilon,\delta_b)<E[B(R_b(r_b'),R_g(r_g'),\delta_b)]\le B^*(\epsilon)<B^*.$$

The first inequality follows from monotonicity properties of B, the second inequality follows from

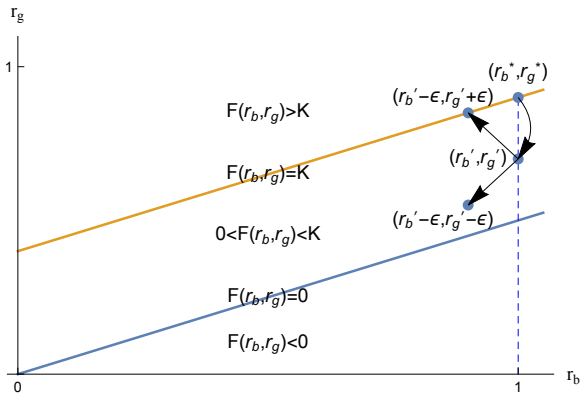

Figure 4: The nature of the deviation from (r_b^*, r_g^*) to (r_b', r_g').

optimality of $B^*(\epsilon)$, and the third is (5). From continuity of B, $\lim_{\epsilon\to 0} B(r_b^* - \epsilon, r_g^* - k\epsilon, \delta_b) = B^*$, and hence $\lim_{\epsilon\to 0} B^*(\epsilon) = B^*$.

(2) For simplicity, denote random variable $R(r_i^*(\epsilon))$ by R_i^*. When the bad ruler chooses $(r_b^*(\epsilon), r_g^*(\epsilon))$ his payoff is:

$$\begin{aligned} B^*(\epsilon) = {} & \Pr\{\pi^*(R_b^*, R_g^*) = 1\} E[B(R_b^*, R_g^*, 1) | \pi^*(R_b^*, R_g^*) = 1] \\ & + \Pr\{\pi^*(R_b^*, R_g^*) = \delta_b\} E[B(R_b^*, R_g^*, \delta_b) | \pi^*(R_b^*, R_g^*) = \delta_b] \\ & + \Pr\{\pi^*(R_b^*, R_g^*) = 0\} E[B(R_b^*, R_g^*, 0) | \pi^*(R_b^*, R_g^*) = 0]. \end{aligned} \tag{6}$$

If $\pi^*(\hat{r}_b, \hat{r}_g) = 1$, then $B(\hat{r}_b, \hat{r}_g, 1) < B(0, 0, 1)$; If $\pi^*(\hat{r}_b, \hat{r}_g) = \delta_b$, then $B(\hat{r}_b, \hat{r}_g, \delta_b) < B(r_b^*, r_b^*, \delta_b) = B^*$; And if $\pi^*(\hat{r}_b, \hat{r}_g) = 0$, then $B(\hat{r}_b, \hat{r}_g, 0) < B(1, \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}, 0) < B^*$. Substituting these into equation (6) yields:

$$B^*(\epsilon) < \Pr\{\pi^*(R_b^*, R_g^*) = 1\} B(0, 0, 1) + \Pr\{\pi^*(R_b^*, R_g^*) = \delta_b\} B^* + \Pr\{\pi^*(R_b^*, R_g^*) = 0\} B(1, \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}, 0).$$

Rearranging the right hand side yields:

$$B^*(\epsilon) < B^* - \Pr\{\pi^*(R_b^*, R_g^*) = 1\}(B^* - B(0, 0, 1)) - \Pr\{\pi^*(R_b^*, R_g^*) = 0\}(B^* - B(1, \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}, 0)).$$

Taking the limit of both sides yields:

$$\lim_{\epsilon\to 0} B(\epsilon) \le B^* - (B^* - B(0, 0, 1)) \lim_{\epsilon\to 0} \Pr\{\pi^*(R_b^*, R_g^*) = 1\} - \lim_{\epsilon\to 0} \Pr\{\pi^*(R_b^*, R_g^*) = 0\}(B^* - B(1, \frac{q}{1-q}\frac{\beta_b - \beta}{\beta_g + \beta}, 0)).$$

From part (1), $\lim_{\epsilon\to 0} B(\epsilon) = B^*$. Because $B(0,0,1) < B^*$ and $B(1, \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}) < B^*$, we must have:

$$\lim_{\epsilon\to 0} \Pr\{\pi^*(R_b^*, R_g^*) = 1\} = 0 \qquad \lim_{\epsilon\to 0} \Pr\{\pi^*(R_b^*, R_g^*) = 0\} = 0.$$

(3) From part (1), $\lim_{\epsilon\to 0} B^*(\epsilon) = B^*$. From part (2), $\lim_{\epsilon\to 0} \Pr\{\pi^*(R_b^*, R_g^*) = \delta_b\} = 1$. Hence, (6) implies that $\lim_{\epsilon\to 0} E[B(R_b^*, R_g^*, \delta_b)|\pi^*(R_b^*, R_g^*) = \delta_b] = B^*$. By continuity, $\lim_{\epsilon\to 0} r_i^*(\epsilon) = r_i^*$. ■

ES2 is justified by the following lemma.

Lemma 9 *Suppose that (1) $q_2(1-\delta_b/\alpha_g) < q < q_2$, so that $B_0 > B_1$, or (2) $q < q_1(p)$ and $p > p^*$, so that $B_0 > B_2$. Moreover, suppose ES2 is violated, so that in the subgame following the bad ruler's choice of (r_b, r_g) for which $F(r_b, r_g) = 0$ the protest probability is $\pi > 0$. Under these conditions no equilibrium exists.*

Proof. Consider the bad ruler's maximization problem over the region where $F(r_b, r_g) \leq 0$:

$$\max_{(r_b,r_g)\in[0,1]^2} B(r_b, r_g, \pi(r_b, r_g)) \quad \text{subject to} \quad F(r_b, r_g) \leq 0, \quad \text{and } \pi(r_b, r_g) = \pi \text{ if } F(r_b, r_g) = 0.$$

Lemma 7 implies that if $\pi = 0$, then the solution is R_0, generating payoff B_0. However, if $\pi > 0$, choosing R_0 does not deliver payoff B_0, because the bad ruler's payoff function is decreasing in the protest probability and $\pi > 0$ at R_0. Thus, the ruler's payoff cannot exceed B_0. Consider the choice of $(r_b', r_g') = (1, \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta} - \epsilon)$ for small ϵ. Because $F(1, \frac{q}{1-q}\frac{\beta_b-\beta}{\beta_g+\beta}) = 0$ and F is increasing in r_g, $F(r_b', r_g') < 0$. Hence, following the bad ruler's choice of (r_b', r_g'), the protest probability is zero. Therefore, the bad ruler's expected payoff of this choice is $B(r_b', r_g', 0) = B_0 - \epsilon\alpha_g(1-q)$. Therefore, as ϵ approaches zero, the bad ruler's payoff approaches B_0, but no strategy delivers payoff B_0.

From Lemma 8, B_0 is larger than the payoff of any (r_b, r_g) for which $F(r_b, r_g) > K$. Moreover, Proposition 5 implies that, under conditions (1) or (2), B_0 is larger than the payoff of any (r_b, r_g) for which $0 < F(r_b, r_g) \leq K$. We showed a sequence of strategies with $F(r_b, r_g) < 0$ delivers a payoff that approaches B_0. Hence, the above ruler's maximization problem has no solution, and no equilibrium exists. ■

8 REFERENCES

Abrahamian, Ervand. 1982. *Iran Between Two Revolutions.* Princeton, NJ: Princeton University Press.

Acemoglu, Daron, and Matthew O. Jackson. 2015. "History, Expectations, and Leadership in the Evolution of Social Norms." *Review of Economic Studies* 82: 1-34.

Acemoglu, Daron, and James A. Robinson. 2000. "Democratization or Repression?" *European Economics Review* 44: 683-93.

Acemoglu, Daron, and James A. Robinson. 2001. "A Theory of Political Transition." *American Economic Review* 91: 938-63.

Acemoglu, Daron, and James A. Robinson. 2006. *Economic Origins of Democracy and Dictatorship.* New York: Cambridge University Press.

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin. 2013. "A Political Theory of Populism." *Quarterly Journal of Economics* 128: 771-805.

Almond, Gabriel. 1956. "Comparative Political Systems." *Journal of Politics* 18: 391-409.

Ananyev, Maxim, Maria Petrova, and Galina Zudenkova. 2015. "Content and Coordination Censorship in Authoritarian Regimes." Mimeo.

Besley, Timothy, and Torsten Persson. 2011. "The Logic of Political Violence." *Quarterly Journal of Economics* 126: 1411-45.

Boix, Carles. 2003. *Democracy and Redistribution.* New York: Cambridge University Press.

Boix, Carles, and Milan Svolik. 2013. "The Foundations of Limited Authoritarian Government: Institutions and Power-Sharing in Dictatorships." *Journal of Politics* 75: 300-16.

Bueno de Mesquita, Ethan. 2010. "Regime Change and Revolutionary Entrepreneurs." *American Political Science Review* 104: 446-66.

Bueno de Mesquita, Ethan. 2013. "Rebel Tactics." *Journal of Political Economy* 121: 323-57.

Bueno de Mesquita, Ethan. 2014. "Regime Change and Equilibrium Multiplicity." Mimeo, University of Chicago.

Burke, Mary, and Peyton Young. 2011. "Social Norms." *Handbook of Social Economics, Volume 1A,* edited by Jess Benhabib, Alberto Bisin, and Matthew Jackson. Amsterdam: North-Holland.

Canes-Wrone, B., M. Herron, and Kenneth Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science* 45: 532-50.

Casper, Brett, and Scott Tyson. 2014. "Popular Protest and Elite Coordination in a Coup d'etat." *Journal of Politics* 76: 548-64.

Chen, Heng, Yang K. Lu, and Wing Suen. 2014. "The Power of Whispers: A Theory of Rumor, Communication and Revolution." *International Economic Review. Forthcoming.*

Chen, Heng, and Wing Suen. 2015. "Aspiring for Change: A Theory of Middle Class Activism." Mimeo.

Conrad, Courtney, and Emily Ritter. 2013. "Treaties, Tenure, and Torture: The Conflicting Domestic Effects of International Law." *Journal of Politics* 75: 397-409.

Conrad Courtenay, and Will Moore. 2010. "What Stops the Torture?" *American Journal of Political Science* 54: 459-76.

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

Diani, Mario, and Doug McAdam. 2003. *Social Movements and Networks: Relational Approach to Collective Action.* New York: Oxford University Press.

Earl, Jennifer. 2011. "Political Repression: Iron Fists, Velvet Gloves, and Diffuse Control." *Annual Review Sociology* 37: 261-84.

Edmond, Chris. 2013. "Information Manipulation, Coordination and Regime Change." *Review of Economic Studies* 80: 1422-58.

Egorov, Georgy, and Konstantin Sonin. 2015. "Incumbency Advantage in Non-Democracies." Mimeo.

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin. 2009. "Why Resource-poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data." *American Political Science Review* 103: 645-68.

Esteban, Joan, and Debraj Ray. 2008. "On the Salience of Ethnic Conflict." *American Economic Review* 98: 2185-2202.

Esteban, Joan, and Debraj Ray. 2011. "Linking Conflict to Inequality and Polarization." *American Economic Review* 101: 1345-74.

Fearon, James. 2011. "Self-enforcing Democracy." *Quarterly Journal of Economics* 126: 1661-1708.

Forero, Juan. 2014. "Venezuela Divided One Year After Chavez's Death; Supporters Commemorate Hugo Chavez while Opponents Protest." *Wall Street Journal (Online)*, Mar 05.

Fox, Justin. 2007. "Government Transparency and Policymaking." *Public Choice* 131: 23-44.

Fox, Justin, and Matthew Stephenson. 2011. "Judicial Review as a Response to Political Posturing." *American Political Science Review* 105: 397-414.

Francisco, Ronald. 2004. "After the Massacre: Mobilization in the Wake of Harsh Repression." *Mobilization* 9: 107-26 .

Fudenberg, Drew, and Jean Tirole. 2000. *Game Theory.* Cambridge, MA: MIT Press.

Gamson, William A. 1975. *The Strategy of Social Protest.* Homewood, IL: Dorsey.

Gandhi, Jennifer. 2008. *Political Institutions Under Dictatorship.* New York: Cambridge University Press.

Guriev, Sergei, and Daniel Treisman. 2015. "How Modern Dictators Survive: An Informational Theory of the New Authoritarianism." Mimeo.

Henderson, Conway. 1991. "Conditions Affecting the Use of Political Repression." *Journal of Conflict Resolution* 35: 120-42.

Khatib, Lina, and Ellen Lust. 2014. *Taking to the Streets: The Transformation of Arab Activism.* Baltimore, MD: Johns Hopkins University Press.

Kreps, David M. 1990. "Corporate Culture and Economic Theory," in J. E. Alt and K. A. Shepsle, eds., *Perspectives on Positive Political Economy*, Cambridge, England: Cambridge University Press, 1990.

Kurzman, Charles. 2003. "The Qum Protests and the Coming of the Iranian Revolution, 1975 and 1978." *Social Science History* 27: 287-325.

Landa, Dimitri, and Scott Tyson. 2014. "The Power of Leaders." Mimeo, NYU.

Levy, Gilat. 2004. "Anti-Herding and Strategic Consultation." *European Economic Review* 48: 503-25.

Lichbach, Mark I. 1987. "Deterrence or Escalation? The Puzzle of Aggregate Studies of Repression and Dissent." *Journal of Conflict Resolution* 31: 266-97.

Lichbach, Mark I. 1989. "A Evaluation of 'Does Economic Inequality Breed Political Conflict?' Studies." *World Politics* 41: 431-470.

Little, Andrew. 2012. "Elections, Fraud, and Election Monitoring in the Shadow of Revolution." *Quarterly Journal of Political Science* 7: 249-83.

Lohmann, Susanne. 1994. "The Dynamics of Informational Cascades: The Monday Demonstra-

tions in Leipzig, East Germany 1989-91." *World Politics* 47: 42-101.

Majumdar, Sumon, and Sharun Mukand. 2008. "The Vanguard as Catalyst: On Leadership and the Mechanics of Institutional Change." Mimeo.

Mansbridge, Jane. 2012. "On the Importance of Getting Things Done." *PS: Political Science & Politics* 45: 1-8.

Mansbridge, Jane. 2014. "What Is Political Science For?" *Perspectives on Politics* 12: 8-17.

Martin, Brian. 2007. *Justice Ignited: The Dynamics of Backfire.* Lanham, MD: Rowman & Littlefield.

Maskin, Eric, and Jean Tirole. 2004. "The Politician and the Judge: Accountability in Government." *American Economic Review* 94: 1034-54.

McAdam, Doug. 1999 *Political Process and the Development of Black Insurgency, 1930-1970.* Chicago, IL: University of Chicago Press.

McAdam, Doug, Sydney Tarrow, and Charles Tilly. 2001. *Dynamics of Contention.* New York: Cambridge University Press.

McCarthy, John D., and Mayer N. Zald. 1973. *The Trend of Social Movements in America: Professionalization and Resource Mobilization* Morristown, NJ: General Learning Corp.

McCarthy John D., and Mayer N. Zald. 1977. "Resource Mobilization and Social Movements: A Partial Theory." *American Journal of Sociology* 82: 1212-41.

Midlarsky, Manus. 1988. "Rulers and the Ruled: Patterned Inequality and the Onset of Mass Political Violence." *American Political Science Review* 82: 491-509.

Miller, Abraham, Louis Bolce, and Mark Halligan. 1977. "The J-curve Theory and the Black Urban Riots: An Empirical Test of Progressive Relative Deprivation Theory." *American Political Science Review* 71: 96482.

Minaya, Ezequiel. 2014a. "Venezuelan Opposition Leader Says He Will Risk Arrest; Leopoldo Lopez Says He Plans to March with Antigovernment Protesters Tuesday." *Wall Street Journal (Online)*, Feb 16.

Minaya, Ezequiel. 2014b. "Protests Against Venezuela's Government Escalate; Thousands Burned Tires, Cars as Security Forces Fought to Control the Streets in Caracas, Other Cities." *Wall Street Journal (Online)*, Feb 20.

Mitchell, Neil, and James McCormick. 1988. "Economic and Political Explanations of Human Rights Violations." *World Politics* 40: 476-98.

Morris, Aldon D. 1984. *The Origins of the Civil Rights Movement: Black Communities Organizing for Change.* New York: The Free Press.

Morris, Stephen. 2001. "Political Correctness." *Journal of Political Economy* 109: 231-65.

Muller, Edward. 1985. "Income Inequality, Regime Repressiveness, and Political Violence." *American Sociological Review* 50: 47-61.

Muller, Edward. 1986. "Income Inequality and Political Violence: The Effect of Influential Cases." *American Sociological Review* 51: 441-45.

Muller, Edward, Henry Dietz, and Steven Finkel. 1991. "Discontent and the Expected Utility of Rebellion: The Case of Peru." *American Political Science Review* 85: 1261-82.

Muller, Edward, and Mitchel Seligson. 1987. "Inequality and Insurgency." *American Political Science Review* 81: 425-51.

Muller, Edward, Mitchel Seligson, Hung-der Fu, and Manus Midlarsky. 1989. "Land Inequality and Political Violence." *American Political Science Review* 83: 577-96.

Muller, Edward, and Erich Weede. 1990. "Cross-National Variation in Political Violence." *Journal of Conflict Resolution* 34: 624-51.

Muller, Edward, and Erich Weede. 1994. "Theories of Rebellion: Relative Deprivation and Power Contention." *Rationality and Society* 6: 40-57.

Persson, Torsten, and Guido Tabellini. 2009. "Democratic Capital: The Nexus of Political and Economic Change." *AEJ: Macroeconomics* 1: 88-126.

Rozenas, Arturas. 2013. "Forcing Consent: Information and Power in Non-Democratic Elections." Mimeo.

Poe, Steven, and Neal Tate. 1994. "Repression of Human Rights to Personal Integrity in the 1980s: A Global Analysis." *American Political Science Review* 88: 853-72.

Postlewaite, Andrew. 2011. "Social Norms and Preferences." *Handbook of Social Economics, Volume 1A,* edited by Jess Benhabib, Alberto Bisin, and Matthew Jackson. Amsterdam: North-Holland.

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

Shadmehr, Mehdi. 2015. "Extremism in Revolutionary Movements." Mimeo.

Shadmehr, Mehdi, and Dan Bernhardt. 2011. "Collective Action with Uncertain Payoffs: Coordina-

tion, Public Signals, and Punishment Dilemmas." *American Political Science Review* 105: 829-51.

Shadmehr, Mehdi, and Dan Bernhardt. 2012. "Vanguards in Revolution." Mimeo.

Shadmehr, Mehdi, and Dan Bernhardt. 2015. "State Censorship." *American Economic Journal: Microeconomics* 7: 280-307.

Shadmehr, Mehdi, and Peter Haschke. 2015. "Youth, Revolution, and Repression." *Economic Inquiry. Forthcoming.*

Siegel, David. 2011. "When Does Repression Work? Collective Action in Social Networks." *Journal of Politics* 73: 993-1010.

Sobel, Joel. 1985. "A Theory of Credibility." *Review of Economic Studies* 52: 557-73.

Svolik, Milan. 2013. "Contracting on Violence: Moral Hazard in Authoritarian Repression and Military Intervention in Politics." *Journal of Conflict Resolution* 57: 765-94.

Tarrow, Sydney. 1998. *Power in Movement: Social Movements and Contentious Politics.* New York: Cambridge University Press.

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

Tilly, Charles. 1996. *European Revolutions: 1492-1992.* Cambridge, MA: Blackwell.

Tilly, Charles. 2004. *Social Movements: 1768-2004.* Boulder, CO: Paradigm.

Tyson, Scott, and Alastair Smith. 2014. "Two-Sided Coordination: Competition in Collective Action." Mimeo.

Vyas, Kejal and Angel Gonzalez. 2013. "Venezuela Opens Investigation of Opposition Leader." *Wall Street Journal (Online)*, Apr 25.

Wang, T.Y., William Dixon, Edward Muller, and Mitchel Seligson. 1993. "Inequality and Political Violence Revisited." *American Political Science Review* 87: 557-96.

Wantchékon, Léonard, and Omar García-Ponce. 2014. "Critical Junctures: Independence Movements and Democracy in Africa." Mimeo.

Weede, Erich. 1981. "Income Inequality, Average Income, and Domestic Violence." *Journal of Conflict Resolution* 25: 639-54.

Weede, Erich. 1986. "Income Inequality and Political Violence Reconsidered." *American Sociological Review* 51: 438-41.

Weede, Erich. 1987. "Some New Evidence on the Correlates of Political Violence: Income Inequal-

ity, Regime Repressiveness, and Economic Development." *European Sociological Review* 3: 97-108.

Wood, Gordon. 2009. *Empire of Liberty: A History of the Early Republic, 1789-1815.* New York, NY: Oxford University Press.

Young, Peyton. 2015. "The Evolution of Social Norms." *Annual Review of Economics 7: 359-87.*

Online Appendix

9 BYSTANDER COMMITMENT

The presence of multiple equilibria suggests that social norms and culture may play a critical role in how non-activist citizens respond to the state's repression of political activists. In this view, culture acts as a focal point to "select" one of the three possible equilibria. Alternatively, we consider the possibility that the bystander can ex ante commit to a strategy. Although we do not model the details of this commitment, we provide some context. For example, when p is relatively high and q is low, citizens can finance human rights organizations that do not tolerate any repression of activists. Such organizations can promote protests by providing entertainment incentives or by invoking the latent feeling of injustice among citizens, provoking protests that would not otherwise take place. In contrast, when p is relatively low and q is relatively high, citizens may deliberately avoid forming groups that could facilitate coordination and the organization of protests.

Suppose the bystander can initially commit to a protest probability π^*. Given the ruler's repression strategy, the bystander's expected payoff is

$$\begin{aligned} U_B =&(1-p)(1-q)\beta_g - (1-p)q\Big[(1-\rho_b^G) + \rho_b^G\pi^*\Big]\beta_b + \\ &p(1-q)\Big[(1-\rho_g^B)\beta_g + \rho_g^B(\pi^*\beta_g - (1-\pi^*)\beta)\Big] - pq\Big[(1-\rho_b^B)\beta_b + \rho_b^B(\pi^*\beta_b + (1-\pi^*)\beta)\Big] \end{aligned}$$

The bystander likes to choose a protest strategy that induces the ruler (good or bad) to always repress the bad activist and to always concede to the good activist. What complicates the bystander'a decision is that if he protests sufficiently rarely that induces the good ruler to always repress the bad activist, then the bad ruler will always repress the good activist. Alternatively, if the bystander protests often enough to prevent the bad ruler from repressing the good activist, then the good ruler also will not repress the bad activist. The best the bystander can do is to pick the lesser evil. When she believes that the activist is likely to be bad ($q > \bar{q}(p)$), she protests rarely enough to induce the good ruler to repress the bad activist, thereby sacrificing the good activist to the bad ruler. Otherwise ($q < \bar{q}(p)$), she protests sufficiently often that only the bad ruler represses the bad activist, thereby saving the good activist by paying the costs of concessions to the bad activist by the good ruler.

Proposition 8 *Suppose $\delta_b < \alpha_g$, and the bystander can commit to a protest strategy. There exists an increasing curve $\bar{q}(p)$ such that, in the unique equilibrium:*

- *When $q > \bar{q}(p)$, the equilibrium strategies are identical to the high repression equilibrium.*
- *When $q < \bar{q}(p)$, the bystander protests with a positive probability, the good ruler concedes to the good and bad activist, and the bad ruler concedes to the good activists, but always represses the bad activist.*

Moreover, $\bar{q}(p) > q_1(p)$ for $p \in (0,1)$, and $\bar{q}(0) < q_2 < \bar{q}(1) < 1$.

When the activist is relatively likely to be bad ($q > \bar{q}(p)$), the equilibrium behavior is the same as the behavior in the high repression equilibrium that arises without commitment. In contrast, the equilibrium behavior that arises when $q < \bar{q}(p)$ is new: with commitment power, the bystander can sometimes join the protest even when she knows that the activist is bad. This allows the bystander to induce the bad ruler to always repress the bad activist and concede to the good activist—in which case repression reveals that the activist is bad. Providing these incentives is costly for the bystander because (1) it induces the good ruler to always concede the bad activist, and (2) the bystander must sometimes follow through on her commitment to support the activist when the ruler represses, even though she is confident that the activist is bad. Both these costs increase with the likelihood that the activist is bad, making it worthwhile for the bystander to commit such a protest strategy when $q < \bar{q}(p)$.

Repression Backfire and Commitment. The bystander's ability to commit to a protest strategy fundamentally changes the forces driving repression backfire. Without commitment, the bystander protests in order to support the good activist against the bad ruler—indeed, if she knew that the activist is bad she would not protest. In contrast, with commitment, the bystander protests *knowing* (in equilibrium) that she is supporting the bad activist. Commitment allows the bystander to discipline the bad ruler, inducing him not to repress the good activist. This raises the bystander's incentive to protest. In particular, with commitment, the bystander supports the ruler under more stringent conditions: $\bar{q}(p) > q_1(p)$. Moreover, when the probability of the bad ruler is higher under the prior, the ability to discipline him is more valuable to the bystander, increasing the bystander's incentive to protest: the gap between $q_1(p)$ and $\bar{q}(p)$ is increasing in p. As a result, with commitment, repression backfires even when $q > q_2$, where no protest occurs in the unique equilibrium without commitment.

9.1 PROOFS OF PROPOSITION 8

Let $\rho \equiv (\rho_b^G, \rho_g^B, \rho_b^B)$ represent the ruler's strategy. Given a protest strategy π, let $\rho^*(\pi)$ be the ruler's best response correspondence defined by (2). In equilibrium, the bystander's choice maximizes $U_B(\pi, \rho)$, subject to the constraint $\rho \in \rho^*(\pi)$. To simplify notation we sometimes suppress the argument of ρ. We divide the proof into four lemmas.

Lemma 10 *An equilibrium exists.*

Proof. Let G be the graph of $U_B(\pi, \rho)$ subject to $\rho \in \rho^*(\pi)$, i.e., $G = \{(\pi, U_B(\pi, \rho)) : \pi \in [0,1],\ \rho \in \rho^*(\pi)\}$. From (2), G is closed. Because $\pi \in [0,1]$ and $U_B(\pi, \rho) \in [-q\beta_b, \beta_g(1-q)]$ for all $(\pi, \rho) \in [0,1]^4$, G is bounded. Hence, G is a compact set. Therefore, there exists $(\pi, \rho^*(\pi))$ that maximizes $U_B(\pi, \rho)$. ■

Lemma 11 *Suppose $\delta_b < \alpha_g$. In equilibrium, $\pi \in \{0, \alpha_g\}$.*

Proof. We prove that no other value of π can be optimal for the bystander. Suppose $\pi \in (0, \delta_b)$. Equation (2) implies that in a neighborhood of π, the ruler's best response is $\rho_b^G = \rho_g^B = \rho_b^B = 1$. Hence,

$$\frac{dU_B}{d\pi} = -\beta_b(1-p)q + p(1-q)(\beta_g + \beta) - pq(\beta_b - \beta) \gtreqless 0 \iff q \lesseqgtr q_1(p).$$

Hence, generically, $U_B(\pi, \rho^*(\pi))$ is strictly monotonic in $(0, \delta_b)$, and a marginal increase or decrease in π is strictly beneficial for the bystander. Hence, no value of $\pi \in (0, \delta_b)$ could be the bystander's equilibrium choice.

Suppose $\pi \in (\delta_b, \alpha_g)$. Equation (2) implies that in a neighborhood of π, the ruler's best response is $\rho_b^G = 0$ and $\rho_g^B = \rho_b^B = 1$. Hence,

$$\frac{dU_B}{d\pi} = p(1-q)(\beta_g + \beta) - pq(\beta_b - \beta) \gtreqless 0 \iff q \lesseqgtr q_2.$$

Hence, generically, a marginal increase or decrease in π is strictly beneficial for the bystander. Hence, no value of $\pi \in (\delta_b, \alpha_g)$ could be the bystander's equilibrium choice.

Suppose $\pi \in (\alpha_g, \alpha_b)$. Equation (2) implies that in a neighborhood of π, the ruler's best response is $\rho_b^G = \rho_g^B = 0$ and $\rho_b^B = 1$. Hence,

$$\frac{dU_B}{d\pi} = -pq(\beta_b - \beta) < 0.$$

Hence, $U_B(\pi, \rho^*(\pi))$ is strictly decreasing in (α_b, α_g), and a marginal decrease in π is strictly beneficial for the bystander. Hence, no value of $\pi \in (\alpha_g, \alpha_b)$ could be the bystander's equilibrium choice.

Suppose $\pi \in [\alpha_b, 1]$. Equation (2) implies that the ruler's best response is $\rho_b^G = \rho_g^B = 0$, and $\rho_b^B = 0$ if $\pi > \alpha_b$, and $\rho_b^B \in [0, 1]$ if $\pi = \alpha_b$. To accommodate both of these cases we explicitly write ρ_b^B in the bystander's expected payoff function:

$$U_B(\pi) = (1-p)(1-q)\beta_g - \beta_b(1-p)q + p(1-q)\beta_g - pq\Big[(1-\rho_b^B)\beta_b + \rho_b^B(\pi\beta_b + \beta(1-\pi))\Big].$$

Consider $\pi' = \alpha_b - \epsilon$. Equation (2) implies that the ruler's best response is $\rho_b^G = \rho_g^B = 0$, $\rho_b^B = 1$, and hence,

$$U_B(\pi') = (1-p)(1-q)\beta_g - \beta_b(1-p)q + p(1-q)\beta_g - pq\Big[(\pi'\beta_b + \beta(1-\pi'))\Big].$$

Hence,

$$\begin{aligned}
U_B(\pi') - U_B(\pi) &= pq\Big[(1-\rho_b^B)\beta_b + \rho_b^B(\pi\beta_b + \beta(1-\pi)) - (\pi'\beta_b + \beta(1-\pi'))\Big] \\
&= pq(\beta_b - \beta)(1 - \rho_b^B - \pi' + \rho_b^B\pi) \\
&> pq(\beta_b - \beta)(1 - \rho_b^B - \pi' + \rho_b^B\alpha_b) \\
&= pq(\beta_b - \beta)(1 - \rho_b^B - (\alpha_b - \epsilon) + \rho_b^B\alpha_b) \\
&= pq(\beta_b - \beta)[(1-\rho_b^B)(1-\alpha_b) + \epsilon],
\end{aligned}$$

which is strictly bigger than 0 for $\epsilon > 0$ and $\rho_b^B \in [0, 1]$. Hence, any protest probability $\pi \in [\alpha_b, 1]$ is dominated by π' and cannot be the bystander's equilibrium choice.

Suppose $\pi = \delta_b$. Equation (2) implies that the ruler's best response is $\rho_b^G \in [0, 1]$ and $\rho_g^B = \rho_b^B = 1$. Hence, for a particular value $\rho_b^G \in [0, 1]$, the bystander's expected payoff is:

$$U_B(\delta_b) = (1-p)(1-q)\beta_g - \beta_b(1-p)q\Big[(1-\rho_b^G) + \rho_b^G\delta_b\Big] + p(1-q)(\delta_b\beta_g - \beta(1-\delta_b)) - pq(\delta_b\beta_b + \beta(1-\delta_b)).$$

First, consider a deviation to $\pi = \delta_b - \epsilon$. Equation (2) implies that the ruler's best response is $\rho_b^G = \rho_g^B = \rho_b^B = 1$. Hence,

$$\begin{aligned}
U_B(\delta_b - \epsilon) = &(1-p)(1-q)\beta_g - \beta_b(1-p)q(\delta_b - \epsilon) + \\
&p(1-q)((\delta_b - \epsilon)\beta_g - \beta(1 - (\delta_b - \epsilon)) - pq((\delta_b - \epsilon)\beta_b + \beta(1 - (\delta_b - \epsilon))),
\end{aligned}$$

and hence,

$$\begin{aligned}U_B(\delta_b-\epsilon)-U_B(\delta_b)&=q(1-p)\beta_b(1-\delta_b)(1-\rho_b^G)+\epsilon(q(\beta_b+p\beta_g)-p(\beta+\beta_g))\\&=q(1-p)\beta_b(1-\delta_b)(1-\rho_b^G)+\epsilon(\beta_b+p\beta_g)(q-q_1(p)).\end{aligned}$$

Hence, $q>q_1(p)$ implies that $U_B(\delta_b-\epsilon)-U_B(\delta_b)>0$ for $\epsilon>0$, and for $q>q_1(p)$, $\pi=\delta_b$ cannot be the bystander's equilibrium choice. Second, consider a deviation to $\pi=\alpha_g+\epsilon$. Equation (2) implies that $\rho_b^G=\rho_g^B=0$, $\rho_b^B=1$. Hence,

$$U_B(\alpha_g+\epsilon)=(1-p)(1-q)\beta_g-(1-p)q\beta_b+p(1-q)\beta_g-pq((\alpha_g+\epsilon)\beta_b+(1-\alpha_g-\epsilon)\beta),$$

and hence,

$$\begin{aligned}U_B(\alpha_g+\epsilon)-U_B(\delta_b)=&q\beta_b(1-\delta_b)(1-p)(1-\rho_b^G)-pq(\beta_b-\beta)\epsilon+p(1-\delta_b)(\beta+\beta_g)\\&-q\Big[p(1-\alpha_g)\beta+(1-\delta_b-p(1-\alpha_g))\beta_b+p(1-\delta_b)\beta_g\Big].\end{aligned}$$

Note that because $1-\delta_b-p(1-\alpha_g)>1-\delta_b-(1-\alpha_g)=\alpha_g-\delta_b>0$, the coefficient on q is negative. Hence, for $q<q_1(p)$,

$$\begin{aligned}U_B(\alpha_g+\epsilon)-U_B(\delta_b)>&-pq(\beta_b-\beta)\epsilon+p(1-\delta_b)(\beta+\beta_g)\\&-q_1(p)\Big[p(1-\alpha_g)\beta+(1-\delta_b-p(1-\alpha_g))\beta_b+p(1-\delta_b)\beta_g\Big]\\=&p^2(1-\alpha_g)\frac{(\beta+\beta_g)(\beta_b-\beta)}{\beta_b+p\beta_g}-pq(\beta_b-\beta)\epsilon.\end{aligned}$$

Hence, $q<q_1(p)$ implies that $U_B(\alpha_g+\epsilon)-U_B(\delta_b)>0$ for small $\epsilon>0$. Hence, for $q<q_1(p)$, $\pi=\delta_b$ cannot be the bystander's equilibrium choice. ■

Lemma 12 *If the bystander selects $\pi=\alpha_g$ in equilibrium, then $\rho_g^B=0$.*

Proof. Suppose that the bystander selects $\pi=\alpha_g$. Equation (2) implies that $\rho_b^G=0$, $\rho_g^B\in[0,1]$, and $\rho_b^B=1$. Hence,

$$\begin{aligned}U_B(\alpha_g)=&(1-p)(1-q)\beta_g-(1-p)q\beta_b\\&+p(1-q)\Big[(1-\rho_g^B)\beta_g+\rho_g^B(\alpha_g\beta_g-\beta(1-\alpha_g))\Big]-pq(\alpha_g\beta_b+(1-\alpha_g)\beta).\end{aligned}$$

Consider a deviation to $\pi = \alpha_g + \epsilon$. Equation (2) implies that $\rho_b^G = \rho_g^B = 0$, and $\rho_b^B = 1$. Hence,

$$U_B(\alpha_g + \epsilon) = (1-p)(1-q)\beta_g - (1-p)q\beta_b + p(1-q)\beta_g - pq((\alpha_g + \epsilon)\beta_b + (1 - \alpha_g - \epsilon)\beta),$$

and hence,

$$U_B(\alpha_g + \epsilon) - U_B(\alpha_g) = p(1-\alpha_g)(\beta + \beta_g)(1-q)\rho_g^B - \epsilon pq(\beta_b - \beta).$$

If $\rho_g^B > 0$ then there exists small $\epsilon > 0$ such that $U_B(\alpha_g + \epsilon) - U_B(\alpha_g) > 0$, and hence, if $\rho_g^B > 0$, then $\pi = \alpha_g$ cannot be the bystander's equilibrium choice. ■

Lemma 13 *In equilibrium, $\pi = \alpha_g$ if and only if $q < \bar{q}(p)$, where*

$$\bar{q}(p) \equiv \frac{p(\beta + \beta_g)}{p(\beta + \beta_g) + (1-p)\beta_b + p\alpha_g(\beta_b - \beta)}.$$

Otherwise, $\pi = 0$ in equilibrium.

Proof. Lemma 10 established that an equilibrium exists, and Lemma 11 established that, in equilibrium, either $\pi = \alpha_g$ or $\pi = 0$. Hence, $\pi = \alpha_g$ in equilibrium if and only if $U_B(\alpha_g) \geq U_B(0)$, and $\pi = 0$ otherwise. If $\pi = \alpha_g$ in equilibrium, then from Lemma 12, $\rho_g^B = 0$, and (2) requires that $\rho_b^G = 0$ and $\rho_b^B = 1$. Hence, $U_B(\alpha_g) = (1-p)(1-q)\beta_g - (1-p)q\beta_b + p(1-q)\beta_g - pq(\alpha_g\beta_b + (1-\alpha_g)\beta)$. If $\pi = 0$, then (2) implies that $\rho_b^G = \rho_g^B = \rho_b^B = 1$. Hence, $U_B(0) = (1-p)(1-q)\beta_g - p\beta$, which implies:

$$U_B(\alpha_g) - U_B(0) = p(\beta + \beta_g) - q(p(\beta + \beta_g) + (1-p)\beta_b + p\alpha_g(\beta_b - \beta)) \geq 0 \iff q \leq \bar{q}(p).$$

Hence, $U_B(\alpha_g) \geq U_B(0)$ if and only if $q \leq \bar{q}(p)$. ■

10 OTHER CASES WITH BAD RULER COMMITMENT

Because of the multiplicity of equilibrium in the subgame that arises when $F(r_b, r_g) \in (0, K)$, the equilibrium of the full game depends on which of the three possible equilibria is anticipated to arise in the subgame. In Proposition 5 we focus on the equilibrium with $\pi = \delta_b$. In the two following propositions, we characterize the equilibrium when $\pi = 0$ and $\pi = 1$. In both cases, no repression backfire occurs.

Proposition 9 *Suppose that when $F(r_b, r_g) \in (0, K)$, the equilibrium of the subgame has $\pi = 0$. In equilibrium, the bystander never protests upon observing repression, the good ruler and the bad ruler always repress the bad activist, and (1) if $q > q_1(p)$, then the bad ruler also always represses the good activist, but (2) if $q < q_1(p)$, he represses the good activist with a positive probability less than one.*

Proof. Lemma 6 establishes part 1 for $q > q_2 > q_1(p)$. If $q_1(p) < q < q_2$, then $F(1, 1) \in (0, K)$. Hence, with $r_b = r_g = 1$, $\pi = 0$ in the equilibrium of the subgame. The monotonicity of $B(r_b, r_g, \pi)$ implies that $r_b = r_g = 1$ must be the bad ruler's equilibrium choice.

Suppose that $q < q_1(p)$. First, we show that if $F(r_b, r_g) = K$ in equilibrium, then we must have $\pi = 0$ in the equilibrium of the subgame. Suppose not, i.e., $\pi > 0$. Because $F(r_b, r_g) = K > 0$, we have $r_g > 0$, and hence r_g can be reduced. If the bad ruler slightly decrease r_g by ϵ, then $0 < F(r_b, r_g - \epsilon) < K$, hence $\pi = 0$ in the equilibrium of the subgame, and hence the bad ruler gains by such a deviation: $B(r_b, r_g - \epsilon, 0) - B(r_b, r_g, \pi) = \pi((1 - q)r_g + qr_b) - \epsilon\alpha_g(1 - q) > 0$ for sufficiently small ϵ.

In addition, any combination of (r_b, r_g) for which $F(r_b, r_g) > K$ is dominated by $r_b = r_g = 0$, and cannot be the bad ruler's equilibrium choice. If $F(r_b, r_g) > K$, then $\pi = 1$, and hence $B(r_b, r_g, 1) = q(1 - r_b)(1 - \alpha_b) + (1 - q)(1 - r_g)(1 - \alpha_g)$. The bad ruler can benefit by deviating to $(r_b, r_g) = (0, 0)$, so that $F(r_b, r_g) = 0$, and his expected payoff becomes $B(0, 0, 0) = q(1 - \alpha_b) + (1 - q)(1 - \alpha_g) > B(r_b, r_g, 1)$.

Therefore, the bad ruler's equilibrium choice solves the following maximization problem:

$$\max_{(r_b, r_g) \in [0,1]^2} B(r_b, r_g, 0) \text{ s.t. } F(r_b, r_g) \leq K.$$

B is increasing in r_b and r_g, and F is increasing in r_g and decreasing in r_b. Thus, at the optimum $r_b = 1$ and r_g satisfies $F(1, r_g) = K$. ■

Proposition 10 *Suppose that when $0 < F(r_b, r_g) \leq K$, the equilibrium of the subgame has $\pi = 1$. In equilibrium, the bystander never protests upon observing repression, the good ruler and the bad ruler always repress the bad activist, and (1) if $q > q_2$, then the bad ruler also always represses the good activist, but (2) if $q < q_2$, he represses the good activist with a positive probability less than one.*

Proof. Lemma 6 establishes part 1. We focus on $q < q_2$. There is no equilibrium in which $F(r_b, r_g) > 0$ because if $F(r_b, r_g) > 0$, then $\pi = 1$, and $B(r_b, r_g, 1) < B(0, 0, 0)$. Next, suppose that if $F(r_b, r_g) = 0$, then $\pi = 0$ in the equilibrium of the subgame. Therefore, the bad ruler's equilibrium choice becomes:

$$\max_{(r_b, r_g) \in [0,1]^2} B(r_b, r_g, 0) \text{ s.t. } F(r_b, r_g) \leq 0.$$

B is increasing in r_b and r_g, and F is increasing in r_g and decreasing in r_b. Thus, at the optimum, $r_b = 1$ and r_g satisfies $F(1, r_g) = K$.

Finally, we show that no equilibrium exists if $\pi > 0$ in the subgame that follows the bad ruler's strategy (r_b, r_g) such that $F(r_b, r_g) = 0$. Because $F(r_b, r_g) = 0$, we have either $r_b > 0$ and $r_g > 0$ or $r_b = r_g = 0$. If $r_g > 0$, then r_g can be reduced. If the bad ruler slightly decrease r_g by ϵ, then $F(r_b, r_g - \epsilon) < 0$, hence $\pi = 0$ in the equilibrium of the subgame, and hence the bad ruler gains by such a deviation. Similarly, if $r_b = r_g = 0$, then r_b can be increased. If the bad ruler slightly increases r_b by ϵ, then $F(r_b + \epsilon, r_g - \epsilon) < 0$, hence $\pi = 0$ in the equilibrium of the subgame, and hence the bad ruler gains by such a deviation. ■